JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 7 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Corrigenda and Addenda

Viewpoint

Consumer Wearables and the Integration of New Objective Measures in Oncology: Patient and Provider Perspectives

Lakshan N Fonseka[1], MS; Benjamin KP Woo[2], MD

[1]College of Osteopathic Medicine of the Pacific, Western University of Health Sciences, Pomona, CA, United States

[2]Olive View-University of California, Los Angeles Medical Center, Sylmar, CA, United States

Corresponding Author:
Lakshan N Fonseka, MS
College of Osteopathic Medicine of the Pacific
Western University of Health Sciences
615 E 3rd St
Pomona, CA, 91766
United States
Phone: 1 909 623 6116
Email: lfonseka@ucla.edu

Abstract

With one in five adults in the United States owning a smartwatch or fitness tracker, these devices are poised to impact all aspects of medicine by offering a more objective approach to replace self-reported data. Oncology has proved to be a prototypical example, and wearables offer immediate benefits to patients and oncologists with the ability to track symptoms and health metrics in real time. We aimed to review the recent literature on consumer-grade wearables and its current applications in cancer from the perspective of both the patient and the provider. The relevant studies suggested that these devices offer benefits, such as improved medication adherence and accuracy of symptom tracking over self-reported data, as well as insights that increase patient empowerment. Physical activity is consistently correlated with stronger patient outcomes, and a patient's real-time metrics were found to be capable of tracking medication side effects and toxicity. Studies have made associations between wearable data and telomere shortening, cardiovascular disease, alcohol consumption, sleep apnea, and other conditions. The objective data obtained by the wearable presents a more complete picture of an individual's health than the snapshot of a 15-minute office visit and a single set of vital signs. Real-time metrics can be translated into a digital phenotype that identifies risk factors specific to each patient, and shared risk factors across one's social network may uncover common environmental exposures detrimental to one's health. Wearable data and its upcoming integration with social media will be the foundation for the next generation of personalized medicine.

(JMIR Mhealth Uhealth 2021;9(7):e28664) doi:10.2196/28664

KEYWORDS

consumer; wearables; smartwatch; cancer; oncology; chemotherapy; apps

Introduction

In 2020, one in five adults in the United States wore a smartwatch or similar fitness tracker [1]. As its prevalence increases, its integration in medical management and research increases as well. Wearables have introduced objective measures to replace self-reported data and supply new variables that offer previously unattainable insights. With increasing options and entry points for consumer wearables, the prevalence of devices, such as the Apple Watch, Fitbit, Samsung Galaxy Watch, and others, is increasing. These consumer wearables are poised to have an impact across all aspects of medicine, and oncology has proved to be a prototypical example.

Wearable research in oncology applications has demonstrated benefits, such as improved medication adherence and accuracy of symptom tracking, over self-reported data, as well as insights that increase patient empowerment. Physical activity is consistently correlated with stronger patient outcomes, and a patient's real-time metrics were found to be capable of tracking medication side effects and toxicity. Studies have made associations between wearable data and telomere shortening, cardiovascular disease, alcohol consumption, sleep apnea, and other conditions. Herein, we review the literature on the new objective measures introduced by consumer-grade wearables and the current applications in cancer from the perspectives of both the patient and the provider.

Objective Measures

The advent of wearables brings new objective measures along with the ability to remotely monitor the data being collected in real time. Before delving into the oncological applications, we summarize the most prevalent wearable measures as an introduction to subsequent studies. The accuracy of these measures, however, remains mostly untested. Some consumers have reported perceived inaccuracy in step counts and sleep metrics in Amazon product reviews [2]. The metrics have also been shown to be affected by a slower gait [3]. Despite this potential limitation to its objectivity, consumer wearables have shown validity in tracking physical activity and are a step forward from self-reported activity data [4].

Of these new measures, physical activity is often the most commonly used and is usually calculated by consumer devices as steps or calories per day. Still, as Beauchamp et al highlights through 25 studies, there exists widespread diversity in metrics gathered. For example, physical activity may be measured as not only steps or calories per day, but also steps or calories per hour, sedentary minutes, standing time, etc. Sleep information is commonly collected, though this also is reported by differing metrics including sleep efficiency percentage, sleep hours, nighttime total wake time, time in bed, and number of awakenings [5].

Many consumer wearables promote continuous heart rate measurement and heart rate variability, with the Apple Watch expanding the concept a step further with a single-lead electrocardiogram (ECG) in certain models. As will be detailed further in this article, the focus of the single-lead ECG is atrial fibrillation detection, but it may also be helpful in QT prolongation detection.

Together, physical activity, sleep, and heart rate tracking form the core of consumer wearable measures. Other features include fall detection, handwashing, and preliminary blood oxygen saturation. Many more exciting new measures are upcoming and are mentioned in the discussion on future outlook.

Impact on Oncology Care

Patient Perspective

Apps

A search for “cancer” in any app store reveals dedicated services aimed to support cancer patients. There is an increasing number of apps that take advantage of the typical wrist location of the wearable to send essential alerts to the patient, such as timely medication reminders to increase adherence. One such app is the Medopad chemotherapy app, which provides on-wrist medication reminders, but also allows patients to submit symptoms and temperatures [6]. The LivingWith: Cancer Support app by Pfizer maintains the medication reminder system and adds the ability to track self-reported mood and pain in just a few taps.

One of the most interesting features comes from the chemoWave app, which maintains the symptom tracking and medication logging features, and combines them with wearable data to generate insights in an easily understandable manner. For example, the app may provide the insight that on days with less sleep, the patient is 43% more likely to experience a headache [7]. The app uses artificial intelligence to generate “personal insights” as in the above example, as well as “global insights” such as “on days chemoWave users like you took more than 7000 steps, they were 90% more likely to feel better” [8]. This appears to be the first consumer-level integration of wearable data with wearable-tracked activity. chemoWave allows users to generate graphs that detail symptoms, exercise, medications, water intake, and more. Patients can show these graphs to the oncologist and assisting doctors to pinpoint changes in patient activity with associated medications or procedures. The benefits include increased accuracy compared with self-reported data, such as having a patient in the office and prompting them to remember what side effects they experienced on days 3 or 4 of treatment. Oncologists have used chemoWave to view trends and better time a patient’s next cycle of chemotherapy. Communication is also streamlined, as patients are able to send doctors daily updates through the app [9].

Although not directly related to cancer, the most highly reviewed app that appears when searching for “cancer” in the Apple App Store is Motivation-Daily Quotes, with over 444,000 reviews at an average rating of 4.8/5. Reviews on the Apple App Store include comments, such as “this app helped me to get through so many tough times,” and follow a general theme of appreciating positivity from the app. Users seem to particularly enjoy the app’s reminder feature. As one user says, “whenever I need encouragement to continue living my life and keep a positive mindset, the app throws the perfect quote for the situation in the perfect moment” [10]. It may not utilize the wearable’s data collection features, but the app takes advantage of the proximity and direct interactivity of the wearable to provide consistent reminders that over time may improve mood and adherence.

Some of the aforementioned apps allow users to connect with family members and cancer support groups to further tap into the motivational aspect. Though functional, these apps are not always successful. For instance, reviews on the Pfizer app mention a “clunky interface,” and users have specifically reported choosing to create a group chat on Facebook instead. Recently, Facebook announced the development of a smartwatch intended to connect users with health apps, which is expected to release in 2022 [11]. Patients who typically receive social support over Facebook, as with the patient who reviewed the Pfizer app, may opt for this smartwatch. At a minimum, social support received through Facebook is likely to be more often seen by the patient through the smartwatch and may augment the overall perception of support to a greater degree. Though only speculation, it is likely that Facebook would incorporate sharing of tracked measures, such as step counts, which may increase the timeliness of support messages by alerting the patient’s support network when physical activity has decreased persistently. In this way, the wearable would allow a patient’s supporters to have greater involvement and interactivity in the patient’s overall health, which may contribute to increased morale for the patient and respective social network.

Geriatrics

As cancer remains a significant disease affecting the geriatric population, it is necessary to consider the relevant applications and ease of use [12,13]. Less than half of adults over the age of 65 years reported owning a smartphone in 2017, with an increase of 24% from 2013. Even so, up to two-thirds of adults over the age of 65 years say they go online [14]. Thus, there may be a subset of older adults who do not have a smartphone, but use other devices to go online [15]. In addition, Hoogland et al showed that despite feeling less confident in using technology, older adults with cancer were mostly agreeable to using health information technology to communicate with their oncology care team [16]. Yet, even if older adults were open to wearables, most consumer devices require a connected smartphone to function as intended. The lack of a smartphone therefore becomes a significant deterrent to wearable use. The Facebook smartwatch will be independent of any smartphone, perhaps offering ease of use to elderly patients without a smartphone. Another challenge identified in elderly populations is lack of internet access, with one out of three patients without any connection to home internet [14]. The Facebook smartwatch will offer its own data connection, further extending coverage of elderly patients and allowing the benefits of wearables to reach geriatric cancer patients.

Young Adults

On the other end of the age spectrum, cancer survivors aged 18 years or older tend to struggle with meeting physical activity recommendations intended to reduce cancer recurrence, mortality, and adverse chronic conditions. These guidelines consist of two strength training sessions per week with either 75 minutes of vigorous physical activity or 150 minutes of moderate activity per week [17]. Miropolsky et al tested the use of a Fitbit device with 13 young adult cancer survivors and found that the participants reported the Fitbit device helped them to reach activity goals. However, this study implemented a buddy system, in which the cancer survivors chose a friend to participate in reaching activity goals. Though participants reported that both the Fitbit device and the buddy system were significant motivators, the impact of the Fitbit device alone is unclear. Nonetheless, this study shows that consumer wearables are a feasible option that may help to increase physical activity in young adult cancer survivors [18].

Melanoma

Up to this point, the information has been relevant for all cancers, but the fifth most common cancer, melanoma, presents a unique opportunity for wearables, as the foremost modifiable risk factor is an individual's UV exposure [19,20]. Although no smartwatches or Fitbit devices can directly measure UV radiation, a new commercially available wearable called Shade contains a UV sensor and provides users with a timeframe of how long they can safely be under the sun. Notifications can be sent through the on-wrist device and companion app every time a user comes 20% closer to the UV exposure limit [21]. A randomized controlled trial involving the use of Shade in melanoma survivors is underway, with results expected to be available in 2022 [22]. Nagelhout et al investigated the feasibility of the Shade wearable and found that 73% of adults and 61% of children selected that they wore the device "all of the time they were outside" on a questionnaire following a 2-week period with Shade [23]. The largest obstacle may be the cost at US $499, but nevertheless, the first commercially available UV sensor is a milestone in consumer wearables and may be invaluable to patients with a family or personal history of melanoma to assuage concerns of sun exposure.

Provider Perspective

From the viewpoint of the provider, the focus shifts onto the role of wearables in the management of patients currently undergoing cancer treatment, as well as cancer survivors. Numerous studies have been conducted on this topic, but the most recent and relevant studies that used typical consumer-level devices or measures will be briefly summarized. The goal of this section is not to analyze the directions of future clinical trials using wearables, but rather to update clinicians on how a patient's wearable may provide clinically relevant data in the near future.

Physical Activity

Several studies have shown that physical activity tracked by wearables is a strong predictor of outcomes in cancer patients [24]. In patients undergoing chemotherapy or radiation treatment, lower step counts were found to be correlated with lower quality of life, greater hospitalization risk, decreased adherence, depression, and shorter survival [25-28]. On the other hand, higher step counts were associated with decreased postoperative complications in abdominal cancer patients and better functional status with reduced likelihood of hospitalization or death [29,30]. A patient, family member, or clinician can generate a graph of the patient's step counts before, during, and after treatment using the wearable's specific app or one of the previously discussed apps. This may aid in identifying at-risk patients and provides another tool to consider in prognosis, especially if done remotely. In addition to tracking physical activity, wearables have been shown to increase physical activity among cancer patients and survivors as well [17,31,32]. Conversely, Rahimy et al found that a Fitbit program for survivors of endometrial cancer led to a transient 22% increase in average steps at 6 months, but by the end of the study at 9 months, participants had returned to their baseline activity levels with no change in BMI. The authors interestingly found step count to be correlated with emotional well-being (P=.005), though this result may represent that subjects who are physically capable of higher step counts are more likely to be healthy, with less severe residual effects of disease and therefore an improved emotional well-being [33]. Overall, these studies suggest that while wearable fitness trackers may motivate some patients, others may require different interventions to inspire enduring lifestyle changes.

Chemotherapy Toxicity

Consumer wearable data can be further used to assess chemotherapy toxicity and side effects using step count, heart rate, and ECG data. In a pilot study, chemotherapy patients had their step counts tracked by a smartphone with a pedometer app. The study found that contacting patients if step counts decreased by 15% or more helped to identify chemotherapy toxicity [34].

While virtually all consumer wearables track step count, it is commonplace to also see wearables that can track heart rate. However, only relatively newer models of the Apple Watch are able to record a single-lead ECG. Both heart rate and ECG data may be used to track side effects in patients taking cardiotoxic chemotherapy drugs. For example, taxanes and angiogenesis inhibitors, such as thalidomide, can lead to sinus bradycardia [35]. Smartwatches can detect bradycardia, and some devices, such as the Apple Watch, allow a threshold heart rate to be set, such that the patient is notified if the heart rate falls below a specified value. Physicians can instruct patients on what heart rate to specify for the alert and thereby have access to continuous remote monitoring for bradycardia.

Atrial fibrillation is a common side effect of chemotherapy regimens that include anthracyclines, trastuzumab, busulfan, and cyclophosphamide [36]. ECG data has the potential to detect atrial fibrillation, though there is conflicting data as to how often false positives occur [36-38]. QT interval prolongation may also be tracked. Timely detection of QT prolongation can prevent conversion to torsades-de-pointes, a life-threatening ventricular tachycardia. Studies have previously approved nonconsumer wearables, such as KardiaMobile-6L, to monitor the QT interval in COVID-19 patients, as remote monitoring became a sudden essential need during the COVID-19 pandemic [39]. Regarding consumer-level wearables, the American Cancer Society recommends that patients who are starting chemotherapy drugs associated with QT prolongation should obtain a baseline 12-lead ECG, but subsequent monitoring can be "accurately and safely performed with topical devices" such as the Apple Watch [40].

Digital Phenotype

Ultimately, wearables have the potential to create a digital phenotype of each patient, which is a real-time record of patient metrics that are translated into stress levels, risk factors, and other determinants of health [41,42]. This potential was illustrated by Teo et al, where self-reported demographics, socioeconomic status, and lifestyle factors were compared with Fitbit-recorded total sleep time (the amount of time a person is asleep) and sleep efficiency (the fraction of time spent asleep over the time a person is in bed) in 482 individuals. The total sleep time was found to be associated with age, gender, habitual alcohol consumption, ethnicity, and occupation type. Interestingly, the authors also sequenced the DNA of study participants and used quantitative polymerase chain reaction to estimate telomere length, a biomarker of aging. Users with insufficient sleep were more likely to show premature telomere shortening. Subjects who had adequate sleep (7 hours or more) had telomeres that were 356 base pairs longer on average than those with insufficient sleep (less than 5 hours) (P=.02). In addition, sleep efficiency was associated with cardiovascular disease risk markers such as BMI. The authors compared the same associations with self-reported sleep data, and no significant links were found, emphasizing the overall improvement in data collection with wearables rather than self-reported information [42].

Furthermore, Tison et al used deep learning algorithms that showed promise in developing digital biomarkers for screening sleep apnea and hypertension with an accuracy of 90% and 82%, respectively [43]. These studies further support the role of wearable data in ushering in a new era of personalized medicine, especially in oncology. The massive inflow of wearable data can be integrated into a straightforward digital phenotype, enabling clinicians to assess in real time a patient's baseline risk factors and any changes that may occur as a result of disease progression or a side effect of a newly initiated chemotherapy regimen. A digital phenotype may also guide interventions, as clinicians can avoid starting a patient on medications with side effects that can deleteriously interact with a patient's risk factors (ie, a cardiotoxic drug in a patient whose digital phenotype shows significant cardiovascular risk).

Future Directions and Conclusions

The impact of wearables has seeped into all fields of research and poses unique opportunities previously unseen in oncology. The above studies indicate clear benefits to both patients and providers. Based on the presented literature, oncologists should use wearable data to increase treatment adherence and patient-physician communication, as well as monitor physical activity and ECG data to guide medical decision-making. The objective measures introduced by consumer wearables will soon be joined by others, including oxygen saturation, blood glucose, and biomarkers in sweat that may indicate cortisol levels [44,45]. Updates to existing consumer devices will include advancements to ECGs, the ability to detect smoking gestures, and smartwatch UV radiation sensors [24]. Completely new wearables will enter the consumer market, such as a headband for the detection of glioblastoma [46].

Soon to be uncovered is how the merging of social media and wearables will expand the landscape. With the imminent launch of the Facebook smartwatch, social media will be integrated with wearable data. Thus, oncologists will have the data of not only their patients, but also each patient's peers. This will allow the oncologist to individualize treatment plans to address the specific risk factors of the patient. For example, location data of a patient and members of the patient's social network may yield data on healthy food access, neighborhood crime rates, air quality, socioeconomic status, local policies on marijuana use, etc, which can serve as predictors of an individual's health in conjunction with wearable data such as physical activity and total sleep time [47]. Likewise, any regional exposures, such as air pollution and lead levels with wearable-measured disease states (ie, reduced physical activity and increased time at hospitals), may unveil carcinogenic environmental exposures [48]. The integration of social media will contribute to the development of future wearable-derived health predictors and assist oncologists in tailoring treatment to target risk factors specific to each patient.

Wearables have become increasingly prevalent in the consumer market, and a look in any direction is likely to reveal a Fitbit, a smartwatch, or another wearable on someone's wrist. From the patient's perspective, consumer wearables allow for improved medication adherence, symptom tracking and insights, communication with physicians and family members, motivational support, and increased reach to patients who were

previously inaccessible despite the advent of the smartphone. From the provider's perspective, wearables assist in achieving the benefits of continuous rather than episodic care, especially in monitoring for medication side effects and toxicity. Oncologists are able to track changes in a patient's baseline metrics outside of the office and infer if any recent treatment changes were responsible. Real-time metrics can be translated into a digital phenotype that identifies risk factors specific to each patient, and shared risk factors across one's social network may uncover common environmental exposures detrimental to one's health. The objective data obtained by the wearable, viewed in the context of one's social network, presents a more complete picture of an individual's health than the snapshot of a 15-minute office visit and a single set of vital signs. Overall, wearable data and its upcoming integration with social media will be the foundation for the next generation of personalized medicine.

Conflicts of Interest

None declared.

References

1. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2021-03-01]
2. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer Perceptions of Wearable Technology Devices: Retrospective Review and Analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
3. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Perera S, Kowalsky RJ, Brach JS. Accuracy and Acceptability of Commercial-Grade Physical Activity Monitors in Older Adults. J Aging Phys Act 2019 Apr 01;27(2):222-229 [FREE Full text] [doi: 10.1123/japa.2018-0036] [Medline: 30117355]
4. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
5. Beauchamp UL, Pappot H, Holländer-Mieritz C. The Use of Wearables in Clinical Trials During Cancer Treatment: Systematic Review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006 [FREE Full text] [doi: 10.2196/22006] [Medline: 33174852]
6. Charara S. Meet the cancer fighting Apple Watch app - the future of digital health is now. Wareable. 2015. URL: https://www.wareable.com/health-and-wellbeing/cancer-fighting-apple-watch-app-medopad-future-of-digital-health-1142 [accessed 2021-02-28]
7. chemoWave. URL: https://chemowave.com/ [accessed 2021-03-01]
8. chemoWave, a Free Health App for Cancer Patients, Launches "My Insights", a Powerful Analytics Tool for its Users. chemoWave. URL: https://chemowave.com/chemowave-free-health-app-for-cancer-patients-launches-my-insights/ [accessed 2021-03-01]
9. Colville W. How one app is helping to tighten the link between cancer patients and their doctors. CNBC. 2017 Sep 04. URL: https://www.cnbc.com/2017/09/01/the-chemowave-app-feeds-data-from-cancer-patients-to-doctors.html [accessed 2021-03-01]
10. Motivation- Daily quotes. Apple App Store. URL: https://apps.apple.com/us/app/motivation-daily-quotes/id876080126 [accessed 2021-03-01]
11. Sonnemaker T. Facebook is reportedly building a smartwatch and wants to sell it to you starting next year in exchange for your health data. Business Insider. 2021. URL: https://www.businessinsider.com/facebook-building-smartwatch-start-selling-2022-apple-health-data-privacy-2021-2 [accessed 2021-03-01]
12. Shahrokni A, Loh KP, Wood WA. Toward Modernization of Geriatric Oncology by Digital Health Technologies. Am Soc Clin Oncol Educ Book 2020 Mar;40:1-7 [FREE Full text] [doi: 10.1200/EDBK_279505] [Medline: 32243198]
13. Shu S, Woo BK. Use of technology and social media in dementia care: Current and future directions. World J Psychiatry 2021 Apr 19;11(4):109-123 [FREE Full text] [doi: 10.5498/wjp.v11.i4.109] [Medline: 33889536]
14. Anderson M, Perrin A. Technology use among seniors. Pew Research Center. 2017. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2021-02-28]
15. Lam NHT, Woo BKP. YouTube as a New Medium for Dementia Education Among Chinese Americans. Community Ment Health J 2020 Apr 22;56(3):435-439. [doi: 10.1007/s10597-019-00493-7] [Medline: 31641910]
16. Hoogland AI, Mansfield J, Lafranchise EA, Bulls HW, Johnstone PA, Jim HSL. eHealth literacy in older adults with cancer. J Geriatr Oncol 2020 Jul;11(6):1020-1022. [doi: 10.1016/j.jgo.2019.12.015] [Medline: 31917114]
17. Zhang FF, Saltzman E, Must A, Parsons SK. Do Childhood Cancer Survivors Meet the Diet and Physical Activity Guidelines? A Review of Guidelines and Literature. Int J Child Health Nutr 2012;1(1):44-58 [FREE Full text] [doi: 10.6000/1929-4247.2012.01.01.06] [Medline: 26973721]

18. Miropolsky EM, Scott Baker K, Abbey-Lambertz M, Syrjala K, Chow EJ, Ceballos R, et al. Participant Perceptions on a Fitbit and Facebook Intervention for Young Adult Cancer Survivors: A Qualitative Study. J Adolesc Young Adult Oncol 2020 Jun;9(3):410-417 [FREE Full text] [doi: 10.1089/jayao.2019.0072] [Medline: 31928489]
19. Cancer Stat Facts: Melanoma of the Skin. National Cancer Institute. URL: https://seer.cancer.gov/statfacts/html/melan.html [accessed 2021-02-28]
20. Armstrong BK, Kricker A. How much melanoma is caused by sun exposure? Melanoma Res 1993 Dec;3(6):395-401. [doi: 10.1097/00008390-199311000-00002] [Medline: 8161879]
21. Shade. 2020. URL: https://www.wearshade.com [accessed 2021-02-28]
22. Vogel RI, Nagler RH, Ahmed RL, Brown K, Luo X, Martinson BC, et al. UVR-sensor wearable device intervention to improve sun behaviors and reduce sunburns in melanoma survivors: study protocol of a parallel-group randomized controlled trial. Trials 2020 Nov 23;21(1):959 [FREE Full text] [doi: 10.1186/s13063-020-04881-3] [Medline: 33228807]
23. Nagelhout ES, Lensink R, Zhu A, Parsons BG, Jensen JD, Wu YP. The Feasibility and Acceptability of Using a Wearable UV Radiation Exposure Monitoring Device in Adults and Children: Cross-Sectional Questionnaire Study. JMIR Dermatol 2020;3(1):e15711 [FREE Full text] [doi: 10.2196/15711] [Medline: 33043275]
24. Low CA. Harnessing consumer smartphone and wearable sensors for clinical cancer research. NPJ Digit Med 2020;3:140 [FREE Full text] [doi: 10.1038/s41746-020-00351-x] [Medline: 33134557]
25. Ohri N, Halmos B, Bodner WR, Cheng H, Guha C, Kalnicki S, et al. Daily Step Counts: A New Prognostic Factor in Locally Advanced Non-Small Cell Lung Cancer? Int J Radiat Oncol Biol Phys 2019 Nov 15;105(4):745-751. [doi: 10.1016/j.ijrobp.2019.07.055] [Medline: 31398385]
26. Gupta A, Stewart T, Bhulani N, Dong Y, Rahimi Z, Crane K, et al. Feasibility of Wearable Physical Activity Monitors in Patients With Cancer. JCO Clin Cancer Inform 2018 Dec;2:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00152] [Medline: 30652571]
27. Bade BC, Brooks MC, Nietert SB, Ulmer A, Thomas DD, Nietert PJ, et al. Assessing the Correlation Between Physical Activity and Quality of Life in Advanced Lung Cancer. Integr Cancer Ther 2018 Mar;17(1):73-79 [FREE Full text] [doi: 10.1177/1534735416684016] [Medline: 28024420]
28. Low CA, Bovbjerg DH, Ahrendt S, Choudry MH, Holtzman M, Jones HL, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018 Jan 05;52(1):88-92 [FREE Full text] [doi: 10.1093/abm/kax022] [Medline: 29538623]
29. Sun V, Dumitra S, Ruel N, Lee B, Melstrom L, Melstrom K, et al. Wireless Monitoring Program of Patient-Centered Outcomes and Recovery Before and After Major Abdominal Cancer Surgery. JAMA Surg 2017 Sep 01;152(9):852-859 [FREE Full text] [doi: 10.1001/jamasurg.2017.1519] [Medline: 28593266]
30. Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med 2018;1:27 [FREE Full text] [doi: 10.1038/s41746-018-0032-6] [Medline: 31304309]
31. Coughlin SS, Caplan LS, Stone R. Use of consumer wearable devices to promote physical activity among breast, prostate, and colorectal cancer survivors: a review of health intervention studies. J Cancer Surviv 2020 Jun;14(3):386-392. [doi: 10.1007/s11764-020-00855-1] [Medline: 31933148]
32. Schaffer K, Panneerselvam N, Loh KP, Herrmann R, Kleckner IR, Dunne RF, et al. Systematic Review of Randomized Controlled Trials of Exercise Interventions Using Digital Activity Trackers in Patients With Cancer. J Natl Compr Canc Netw 2019 Jan 18;17(1):57-63 [FREE Full text] [doi: 10.6004/jnccn.2018.7082] [Medline: 30659130]
33. Rahimy E, Usoz M, von Eyben R, Fujimoto D, Watanabe D, Karam A, et al. Phase II trial evaluating efficacy of a Fitbit program for improving the health of endometrial cancer survivors. Gynecol Oncol 2021 Apr;161(1):275-281. [doi: 10.1016/j.ygyno.2021.01.033] [Medline: 33551199]
34. Soto-Perez-De-Celis E, Kim H, Rojo-Castillo MP, Sun C, Chavarri-Guerra Y, Navarrete-Reyes AP, et al. A pilot study of an accelerometer-equipped smartphone to monitor older adults with cancer receiving chemotherapy in Mexico. J Geriatr Oncol 2018 Mar;9(2):145-151. [doi: 10.1016/j.jgo.2017.09.008] [Medline: 29017891]
35. Spînu , Cismaru G, Boarescu P, Istratoaie S, Negru AG, Lazea C, et al. ECG Markers of Cardiovascular Toxicity in Adult and Pediatric Cancer Treatment. Dis Markers 2021;2021:6653971 [FREE Full text] [doi: 10.1155/2021/6653971] [Medline: 33532005]
36. Turakhia M, Perez M. Results of a Large-scale, App-based Study to Identify Atrial Fibrillation Using a Smartwatch: The Apple Heart Study. American College of Cardiology 68th Scientific Session and Exposition 2019 May;44(5):277-281 [FREE Full text] [Medline: 31080336]
37. Raja JM, Elsakr C, Roman S, Cave B, Pour-Ghaz I, Nanda A, et al. Apple Watch, Wearables, and Heart Rhythm: where do we stand? Ann Transl Med 2019 Sep;7(17):417 [FREE Full text] [doi: 10.21037/atm.2019.06.79] [Medline: 31660316]
38. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch Performance for the Detection and Quantification of Atrial Fibrillation. Circ Arrhythm Electrophysiol 2019 Jun;12(6):e006834. [doi: 10.1161/CIRCEP.118.006834] [Medline: 31113234]

39. Castelletti S, Dagradi F, Goulene K, Danza AI, Baldi E, Stramba-Badiale M, et al. A wearable remote monitoring system for the identification of subjects with a prolonged QT interval or at risk for drug-induced long QT syndrome. Int J Cardiol 2018 Sep 01;266:89-94. [doi: 10.1016/j.ijcard.2018.03.097] [Medline: 29887480]
40. Lenihan D, Carver J, Porter C, Liu JE, Dent S, Thavendiranathan P, et al. Cardio-oncology care in the era of the coronavirus disease 2019 (COVID-19) pandemic: An International Cardio-Oncology Society (ICOS) statement. CA Cancer J Clin 2020 Nov;70(6):480-504 [FREE Full text] [doi: 10.3322/caac.21635] [Medline: 32910493]
41. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
42. Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
43. Tison GH, Singh AC, Ohashi DA, Hsieh JT, Ballinger BM, Olgin JE, et al. Cardiovascular risk stratification using off-the-shelf wearables and a multi-task deep learning algorithm. In: Circulation. 2018 Presented at: American Heart Association's 2017 Scientific Sessions and Resuscitation Science Symposium; 2017; Anaheim, CA.
44. Sheibani S, Capua L, Kamaei S, Akbari SSA, Zhang J, Guerin H, et al. Extended gate field-effect-transistor for sensing cortisol stress hormone. Commun Mater 2021;2(1):10 [FREE Full text] [doi: 10.1038/s43246-020-00114-x] [Medline: 33506228]
45. Chong KP, Woo BK. Emerging wearable technology applications in gastroenterology: A review of the literature. World J Gastroenterol 2021 Mar 28;27(12):1149-1160 [FREE Full text] [doi: 10.3748/wjg.v27.i12.1149] [Medline: 33828391]
46. Samuel N, So E, Djuric U, Diamandis P. Consumer-grade electroencephalography devices as potential tools for early detection of brain tumors. BMC Med 2021 Jan 22;19(1):16 [FREE Full text] [doi: 10.1186/s12916-020-01889-z] [Medline: 33478449]
47. Bagot KS, Matthews SA, Mason M, Squeglia LM, Fowler J, Gray K, et al. Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. Dev Cogn Neurosci 2018 Aug;32:121-129 [FREE Full text] [doi: 10.1016/j.dcn.2018.03.008] [Medline: 29636283]
48. Merchant RM, Asch DA, Crutchley P, Ungar LH, Guntuku SC, Eichstaedt JC, et al. Evaluating the predictability of medical conditions from social media posts. PLoS One 2019;14(6):e0215476 [FREE Full text] [doi: 10.1371/journal.pone.0215476] [Medline: 31206534]

Abbreviations

ECG: electrocardiogram

Edited by L Buis; submitted 10.03.21; peer-reviewed by J Guo, S Shu, A Payton; comments to author 05.04.21; revised version received 12.04.21; accepted 11.06.21; published 15.07.21.

Please cite as:
Fonseka LN, Woo BKP
Consumer Wearables and the Integration of New Objective Measures in Oncology: Patient and Provider Perspectives
JMIR Mhealth Uhealth 2021;9(7):e28664
URL: https://mhealth.jmir.org/2021/7/e28664
doi:10.2196/28664
PMID:34264191

Review

Embodied Conversational Agents for Patients With Dementia: Thematic Literature Analysis

Margherita Rampioni[1], PsyD, PhD; Vera Stara[1], PhD; Elisa Felici[1], PsyD; Lorena Rossi[1], MSc; Susy Paolini[2], PsyD

[1]Model of Care and New Technologies, IRCCS INRCA - National Institute of Health and Science on Aging, Ancona, Italy

[2]Unit of Neurology, IRCCS INRCA - National Institute of Health and Science on Aging, Ancona, Italy

Corresponding Author:
Vera Stara, PhD
Model of Care and New Technologies
IRCCS INRCA - National Institute of Health and Science on Aging
via Santa Margherita, 5
Ancona, 60124
Italy
Phone: 39 0718004614
Email: v.stara@inrca.it

Abstract

Background: As the world's population rapidly ages, the number of older adults with cognitive impairment will also increase. Several studies have identified numerous complex needs of people with dementia, which assistive technologies still fail to support. Recent trends have led to an increasing focus on the use of embodied conversational agents (ECAs) as virtual entities able to interact with a person through natural and familiar verbal and nonverbal communication. The use of ECAs could improve the accessibility and acceptance of assistive technologies matching those high-level needs that are not well covered to date.

Objective: The aim of this thematic literature analysis was to map current studies in the field of designing ECAs for patients with dementia in order to identify the existing research trend and possible gaps that need to be covered in the near future. The review questions in this study were as follows: (1) what research frameworks are used to study the interaction between patients with dementia and ECAs? (2) what are the findings? and (3) what are the barriers reported in these studies?

Methods: Separate literature searches were conducted in PubMed, Web of Science, Scopus, and Embase databases by using specific umbrella phrases to target the population (patients with dementia) and the technology-based intervention (embodied conversational agent). Studies that met the inclusion criteria were appraised through the Mixed Methods Appraisal Tool and then discussed in a thematic analysis.

Results: The search process identified 115 records from the databases and study references. After duplicates (n=45) were removed, 70 papers remained for the initial screening. A total of 7 studies were finally included in the qualitative synthesis. A thematic analysis of the reviewed studies identified major themes and subthemes: the research frameworks used to gather users' perspectives on ECAs (theme 1), the insights shared by the 7 studies as well as the value of user involvement in the development phases and the challenge of matching the system functionalities with the users' needs (theme 2), and the main methodological and technical problems faced by each study team (theme 3).

Conclusions: Our thematic literature analysis shows that the field of ECAs is novel and poorly discussed in the scientific community and that more sophisticated study designs and proofs of efficacy of the approach are required. Therefore, by analyzing the main topic of the narrative review, this study underscores the challenge of synchronizing and harmonizing knowledge, efforts, and challenges in the dementia care field and its person-centered paradigm through the user-centered design approach. Enabling strict collaboration between interdisciplinary research networks, medical scientists, technology developers, patients, and their formal and informal caregivers is still a great challenge in the field of technologies for older adults.

(JMIR Mhealth Uhealth 2021;9(7):e25381) doi:10.2196/25381

KEYWORDS

dementia; patient with dementia; older adults with dementia; embodied conversational agent; virtual personal assistant; virtual agent; virtual companion; design for older adults; patients; elderly; virtual; personal assistant; cognitive; cognitive impairment

Introduction

Background

The world's population is rapidly aging and approximately 47 million people are now experiencing dementia worldwide. This number could triple by 2050 with an incremental estimated cost that will range from US $818 billion in 2015 to US $2 trillion by 2030 [1]. Dementia is characterized by the progressive deterioration of both cognitive and functional abilities that affect a person's capability to perform everyday activities [2,3]. Nowadays, the tendency to view dementia solely within a medical framework is overcome through a new personhood paradigm based on the identification of the numerous complex needs of patients with dementia as described by Maslow [4] and Kitwood [5]. There is consensus that individuals living independently [6] or in long-term care [7-9] are able to express needs [10] and preferences [11] consistently, even in the advanced stage of dementia [12].

Since the challenges of responding to the growing number of people with dementia and their complex needs are substantial for governments [13], the field of assistive technologies is receiving more and more interest. In aged care, the term "assistive technology" refers to any device, product, or equipment that helps people to perform a task they would otherwise be unable to do or that facilitates older adults' activities of daily living [14]. In dementia care, assistive technologies (ie, technologies for daily living, meaningful and pleasurable activities and health care) can compensate for cognitive impairment, improve quality of life, favor autonomy, enable people to remain in their homes for longer, and reduce care costs [15-19]. Actually, a broad spectrum of technologies supports community-dwelling persons with dementia. These technologies mostly address basic physiological and safety needs, whereas little attention is devoted to higher-level needs such as self-esteem, quality of life, recreational activities or contrast behavioral issues, for example, aggression and mood changes [6,20-22]. Recently, several studies [23-27] proposed the use of screen-based entities designed to stimulate human face-to-face conversation skills called as embodied conversational agents (ECAs) or personal virtual assistants [28,29]. Such virtual entities are able to interact with a person through verbal and nonverbal communication. There is a significant and growing list of use cases for ECAs targeting older adults with or without cognitive impairment or their caregivers [15]: virtual coaches [30], virtual companions [31-33], personal virtual assistants [26], virtual butlers [34,35], and training tools to help formal and informal caregivers [36].

The use of ECAs could improve the accessibility and acceptance of computer-based assistive technologies when compared to graphical user and voice interfaces, especially for older adults with cognitive impairment [37-39], thus matching those high-level needs that are not satisfactorily covered by assistive technologies. The extent to which this specific innovation may be able to support people affected by dementia along the progressive nature of the disease represents a great challenge for the entire scientific community. Unfortunately, it seems that research in that specific direction is still poor and little is known on how patients with dementia interact with ECAs and how this interaction should be designed and managed by the system [15]. Therefore, to bridge this gap, this paper discusses the implications derived from a thematic literature review of the available studies focusing on personal virtual assistants in favor of patients with dementia.

Aim of This Study

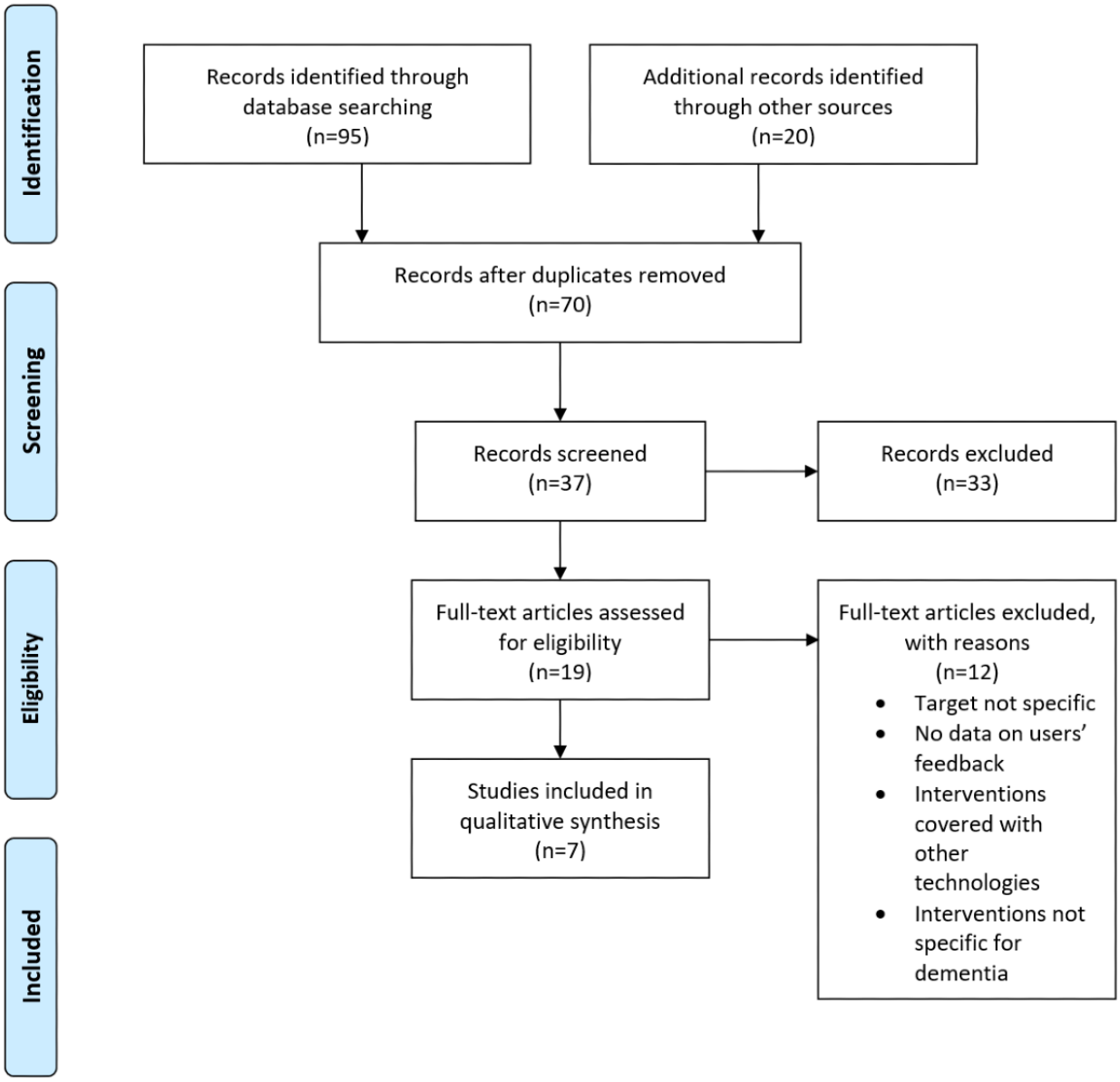
The aim of this thematic literature analysis was to map current studies in the field of designing ECAs for patients with dementia in order to identify the existing research trends and possible gaps that need to be covered in the near future. The review questions were (1) what research frameworks are used to study the interaction between patients with dementia and ECAs? (2) what are the findings? and (3) what are the barriers reported in these studies?

Methods

Design of This Study

Separate literature searches were conducted in PubMed, Web of Science, Scopus, and Embase databases by using the following umbrella phrases to target the population and the specific technology-based intervention: ("patient with dementia" OR "people with dementia" OR "person with dementia") AND ("virtual agent" OR "personal virtual assistant" OR "virtual companion" OR "embodied conversational agent"). Inclusion criteria were published papers written in English with the aim of studying the use of ECAs (1) among older adults (≥65 years) with dementia living at home, in long-term care, or nursing homes and their formal and informal caregivers, (2) for coping in patients with dementia without any restriction in terms of service applications (ie, cognitive games, reminders, medicine intake, calendar, etc), (3) for presenting empirical findings about interactions between users and ECAs, and (4) in randomized controlled trials (qualitative, quantitative, and the mixed methods approach were included). There was no restriction on publication dates, and the searches were finalized in July 2020. Papers were excluded if reviews, theoretical or technical studies, and contributions were not original research papers that met the inclusion criteria or were not written in English. According to predefined criteria, the screening phase was based on the analysis of titles and then abstracts. Later, full texts of those titles/abstracts of screened publications were reviewed independently by the first and the corresponding author in August 2020. Another researcher was involved in reaching consensus in cases of disagreement. Studies that met the inclusion criteria were included, and the results of the searches were summarized. Then, we performed a manual thematic analysis of the findings. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis [40] flowchart in the retrieval and selection process (Figure 1, Multimedia Appendix 1).

Figure 1. Flow diagram of the studies included in the thematic review as well as the main reasons for rejection.



Quality Appraisal

Three authors independently appraised the final papers for their methodological quality by using the Mixed Methods Appraisal Tool version 2018 [41]. The Mixed Methods Appraisal Tool assesses the quality of qualitative, quantitative, and mixed methods studies and can be used to appraise the quality of different study designs. Precisely, it focuses on the methodological criteria and includes 5 core quality criteria for qualitative, randomized controlled, nonrandomized, quantitative descriptive, and mixed methods. Owing to these advantages, it was chosen over other tools prior to starting the narrative review. The results of each appraisal were compared, and any disagreement was solved through the intervention of the last author and discussion among the authors. According to the scoring system proposed by Pluye et al [42], a quantitative appraisal score was calculated by assessing the presence/absence of criteria (yes/no). The quality score was calculated as a percentage by using the formula: (number of "yes" responses divided by the number of appropriate criteria) × 100.

Thematic Analysis

A thematic analysis was conducted to identify themes based on the 6 phases described by Braun and Clarke [43]. Themes were coded based on our specific review questions by following these steps: first, authors read the data from each study; and second, the corresponding author carried out a systematic manual coding of the features that led to initial codes before searching for themes in the third step. Themes were then reviewed for correlation with codes and the identification of subthemes during the fourth phase. After defining the themes in phase five, findings were evaluated for relevance to the research questions. Finally, the authors discussed the analysis process and reached consensus on the labelling.

Results

Reviewed Studies

The search process identified 95 records from the databases and an additional 20 by manually searching those studies' references.

After duplicates (n=45) were removed, 70 papers remained for initial screening by title. This resulted in 37 potentially eligible abstracts. The abstracts retained were analyzed by authors according to the research questions in order to obtain the final list of full-text papers to be reviewed. After the analysis of the abstracts, 33 of them were excluded as they did not fit the settled criteria of the target population and the specific technology-based intervention. A second screening step was performed for those full-text papers that matched all the criteria (n=19). From this process, 12 papers were excluded because (1) studies recruited both patients with dementia and older adults in good health status (n=3), (2) no empirical feedback from users was reported (n=3), (3) the intervention was performed using both ECAs and other technologies (n=2), and (4) the intervention was not specific for patients with dementia (n=4). Finally, a total of 7 studies were included.

A summary of the studies and their findings are presented in Table 1. The 7 studies were conducted in Spain [24], Japan [27], France [15,44], Canada [45], Italy, and Luxembourg [46,47] and published between 2008 and 2019. They were heterogeneous in terms of objectives, population, contexts, and methodologies. Four of the 7 studies describe the development of 2 specific agents: Louise [15,44] and Anne [46,47].

The different agents are shown in Figure 2. Carrasco et al [24] built a prototype that included a tool for real-time streaming of a realistic avatar, previously programmed by a caregiver, to a television. This technique was used to simulate a true virtual assistant on the television screen. The avatar has a realistic voice and the lips are in synchronization with its speech to ensure that its facial movements appear natural. Yasuda et al [27] developed an agent conversation system shown on a computer screen in the form of an animated face resembling a 5-year-old grandchild. This system can detect the end of the speech sound of a subject's reply to a question and begins asking the next question. When the subject speaks, the agent reacts by automatically generating nods, mouth movements, and acknowledgments. Wargnier et al [15,44] proposed a prototype of a semiautomatic system that allows the animation of a female cartoon-like character called Louise and speech synthesis from text. The character, displayed on either a computer screen or a television set, is in an idle pose and moves its lips while speaking. This system includes attention monitoring and interaction management based on a predefined script and keyboard inputs of a (presumably hidden) operator. Konig et al [45] developed an emotionally intelligent cognitive assistant in the form of a humanoid female character that shows up on a screen in an inactive pose. de Jong et al [46] and Stara et al [47] shared the know-how gained in the design and adaptation of the personal virtual assistant Anne, which is a friendly, female human-looking avatar talking and interacting with users on a screen. Older adults can communicate with the ECA through voice and touch. This system is able to learn autonomously from its users and gets to know their personal preferences and needs.

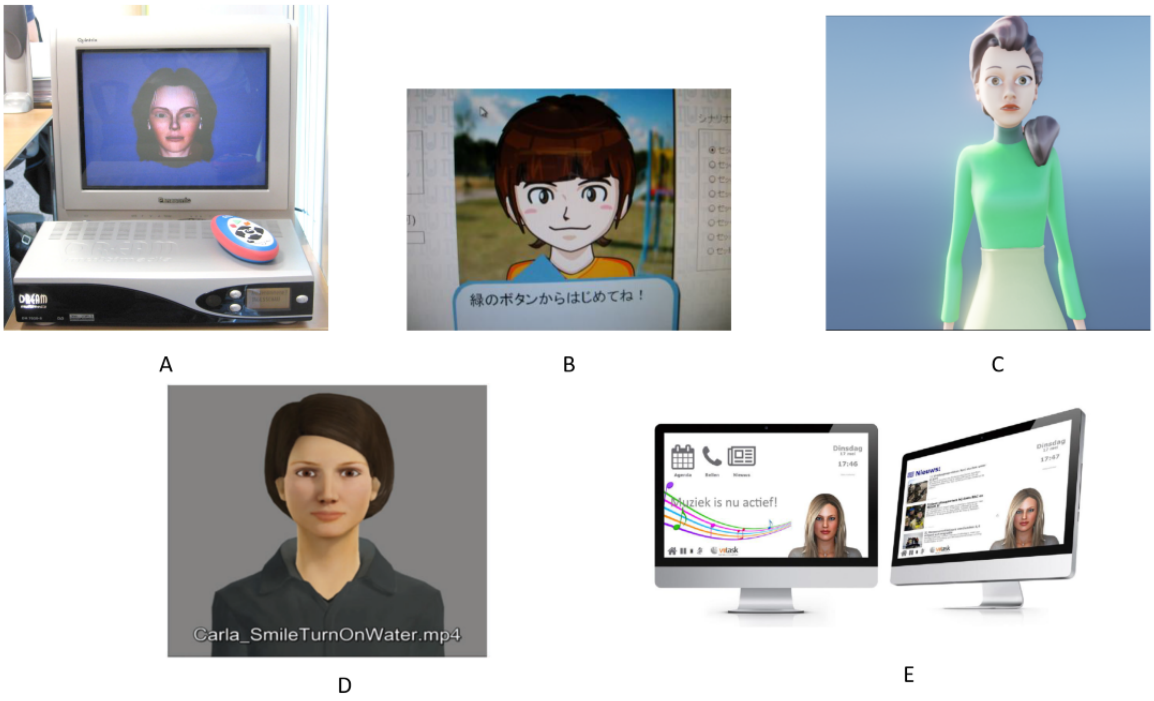
Table 1. Summary of the reviewed studies.

Studies	Purpose	Type of system	Method for data collection	Sample	Country, test setting	Findings
Carrasco et al [24]	To validate a functional prototype that gives a measure on how natural the interaction between avatars and people with Alzheimer disease is. To increase the acceptability of the system by target users	Avatar displayed on a standard television set	Yes/No questions and one-to-one observation	21 persons had Alzheimer disease, with a Global Deterioration Scale [48] measure ranging from 3 to 5 (from mild to moderate)	Spain, day care center	All users engaged naturally with the avatar, understood the information conveyed by the avatar, and answered successfully by means of the television remote control
Yasuda et al [27]	To investigate the effectiveness of a conversation system based on an animated face of a child	A computer screen that shows an animated face of a child agent	Qualitative interviews	8 older adults (2 males and 6 females) had mild Alzheimer disease, with a Mini-Mental State Examination [49] mean score of 22.2. The average age was 78.5 years.	Japan, hospital	All users conversed with the conversational agent system and enjoyed the conversation.
Wargnier et al [44]	To collect design guidelines to develop a semiautomated ECA[a] prototype	A semiautomated cartoon like ECA prototype that runs on a standard personal computer with Microsoft Windows	Semiautomated Wizard of Oz, video, observation, open interview, questionnaire	14 specialists (4 males and 10 females) in assistive technologies for older adults or care professionals (medical doctors and neuropsychologists, mostly)	France, hospital	All participants interacted naturally with the ECA. Most displayed high levels of attention. Globally, the feedbacks turned out to be rather positive.
Konig et al [45]	To identify affective identities in patients with dementia for the design of cognitive assistive technologies	An intelligent cognitive assistant in the form of a humanoid female character shown on a screen	Qualitative interview	12 older adult care home residents (5 males and 7 females) with Alzheimer disease who showed cognitive and functional impairment to an extent that it affected their autonomy in performing certain complex activities of daily living and 9 associated caregivers (2 males and 7 females). The average age of the residents was 84.5 years.	Canada, University and Research Institute for Aging	Definition of user requirements for the design
Wargnier et al [15]	To conduct a usability study to refine and validate the Louise ECA	A semiautomated cartoon like ECA prototype that runs on a standard personal computer under Microsoft Windows	Realistic assistive scenarios and semistructured interview	14 participants (3 males and 11 females) with mild cognitive impairment (9/14) or Alzheimer disease (5/14), whose Mini-Mental State Examination [49] scores ranged from 8 to 30 (mean 23.8, SD 4.9). The average age was 78.8 years.	France, hospital and University	Most of the participants were able to interact with the ECA, succeeded in completing the proposed tasks, and enjoyed the design
de Jong et al [46]	Report the first iteration of a comprehensive user-centered development process of virtual agents for patients with dementia and their caregivers	A personal assistant called Anne that works on a Surface Pro tablet under the Microsoft Windows 10 operating system	Focus group	16 caregivers: 10 in Luxembourg (6 qualified nursing assistants and 4 informal carers) and 6 in Italy (3 care professionals and 3 informal caregivers)	Luxembourg, Italy, hospital, day care center	Definition of user requirements for the design

Studies	Purpose	Type of system	Method for data collection	Sample	Country, test setting	Findings
Stara et al [47]	How patients xperience a personal virtual assistant in the stage of moderate dementia; how a personal virtual assistant can be modified to the requirements of people in the stage of moderate dementia	A personal assistant called Anne that works on a Surface Pro tablet under the Microsoft Windows 10 operating system	Protected environment test scenarios with observation of the interactions between patients and the personal virtual assistant and interview to formal caregivers	5 female patients with moderate dementia and 2 formal caregivers in Italy; 1 female patient with dementia and 2 formal caregivers in Luxembourg	Italy, Luxembourg, hospital, day care center	Definition of user requirements for the design

[a]ECA: embodied conversational agent.

Figure 2. Embodied conversational agents described in the reported studies. A) Carrasco et al [24]; B) Yasuda et al [27]; C) Wargnier et al [15,44]; D) Konig et al [45]; E) de Jong et al [46] and Stara et al [47].



Quality Appraisal of the Selected Studies

The design of the 7 research studies was assessed by using screening questions and the 5 criteria of the Mixed Methods Appraisal Tool [41] for qualitative and mixed methods studies reported in Textbox 1: (1) appropriateness of objective/research question, (2) adequacy of the qualitative approach/method, (3) adequate gathering of findings from data, (4) sufficient interpretation of results from data, and (5) coherence between qualitative data sources, collection, analysis, and interpretation.

Textbox 1. Mixed Methods Appraisal Tool version 2018 criteria to appraise the study design.

Methodological quality criteria

Screening questions (for all types)

S1. Are there clear research questions?

S2. Do the collected data allow to address the research questions?

Qualitative

Q1. Is the qualitative approach appropriate to answer the research question?

Q2. Are the qualitative data collection methods adequate to address the research question?

Q3. Are the findings adequately derived from the data?

Q4. Is the interpretation of results sufficiently substantiated by data?

Q5. Is there coherence between qualitative data sources, collection, analysis, and interpretation?

Mixed methods

M1. Is there an adequate rationale for using a mixed methods design to address the research question?

M2. Are the different components of the study effectively integrated to answer the research question?

M3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?

M4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?

M5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?

All the studies used appropriate research design, taking into account the research questions and problems related to the use of a specific innovation technology among persons with dementia. The prevalent data collection methods were open or semistructured interviews [15,27,45,47], observations [24,44,47], a questionnaire [24,44], and a focus group [46]. The findings and interpretations of the results were coherent with data sources, collection, analysis, and discussion in all the studies. According to Pluye et al [42], the score for each study was calculated and then synthetized using 3 different categories: low score, <35%; medium score, 36%-70%; and high score, 71%-100%. All studies met the 5 quality criteria; therefore, the score calculation synthesized a high score of methodological quality results (Table 2).

Table 2. Quality scores of the selected studies.

Studies	Screening question score	Qualitative studies score	Mixed methods studies score	Total score	Appropriate criteria (n)	Quantity score (%)	Score category
Carrasco et al [24]	2	N/A[a]	5	7	7	100	High
Yasuda et al [27]	2	5	N/A	7	7	100	High
Wargnier et al [44]	2	N/A	5	7	7	100	High
Konig et al [45]	2	5	N/A	7	7	100	High
Wargnier et al [15]	2	N/A	5	7	7	100	High
de Jong et al [46]	2	5	N/A	7	7	100	High
Stara et al [47]	2	5	N/A	7	7	100	High

[a]N/A: not applicable.

Thematic Analysis

Following the analysis of reviewed studies, 3 major themes and subthemes within each theme were identified: (1) research frameworks, (2) efficacy of ECAs, (3) limitations of the studies and problems faced.

Theme 1: Research Frameworks

All the studies dealt with 2 research questions: (1) could virtual agent be a technology that patients with dementia can really use? and (2) which are the design features that can facilitate or hinder this usage? Methodologically, both qualitative and quantitative designs were applied to answer these questions focusing on meanings and understanding of experiences of people with dementia and their carers. In all cases, participants took part in the researches in participatory sessions or observations to avoid the discomfort of being the subjects of an experimental study. Overall, the 7 studies came to the general assumption that the use of ECAs is suitable for people with dementia. This common evidence based its foundation on the use of voice as an interaction modality between the systems. The voice as input/output modality is the natural and familiar way to engage people with dementia in such researches. Therefore, from 2008 to 2019, even though the readiness level of the virtual agents was changed considerably, especially in

the human-looking appearance, none of the studies shared skepticism or disadvantages in the use of ECAs by the enrolled patients. Additionally, the use of common screen devices as the presentation platform on the television [24], computer [15,27,44,45], or tablet [47] concur with the positive engagement of users. A notable consideration that emphasizes this outcome is the rigorous recruitment strategy followed by each team: as reported in Table 1, enrolled participants were previously diagnosed and scored on the Global Deterioration Scale [48] and the Mini-Mental State Examination [49]. Moreover, the experimental settings were always under the responsibility of researchers or formal caregivers in all the studies. However, the objectives of the 7 studies were different from the research question about which feature design could improve or impede the use of virtual characters. This is due, in part, to the specific purpose of each system under development that varied from natural interaction [24,27,45] and their specificities [15,44] to support independent living [46,47]. Features such as interaction paradigm and prompting style are seen as the main components that could be personalized and used for matching the needs and capabilities of users, thereby improving the user experience. Probably because the ECA field is still green, none of the studies pursued the general purpose to achieve general features design as the guidelines for future researches in the field.

Theme 2: Efficacy of ECAs

Findings of the Selected Studies

As previously mentioned, all the studies reported positive feedback on the use of agents by users: the majority of patients with dementia naturally interacted and responded to the virtual character fulfilling the assigned tasks [15,24,27,45]. The artificial assistant caught the attention of people, and in some cases, it was seen as a companion by older participants as well as formal and informal caregivers [46,47]. On the contrary, the preference of having a "real person helping me rather than a machine" [45] clearly showed the importance of a human contact. Moreover, participants with the most severe memory impairment frequently forgot how to interact with the system or talk to the ECA and it caused them some frustration or they seemed to be intimidated by the character [15]. These results are in line with the findings of the study conducted by Stara et al [47], who found that patients with moderate dementia spoke freely to the ECA without using the touch control. Persons with dementia generally had little experience with mobile technology before they became affected by dementia; therefore, patients seemed to have trouble using a touchscreen and navigating the different applications of the ECA. For that reason, formal caregivers involved in this research suggested a supervised usage of the virtual agent in a controlled environment such as daily centers. Formal caregivers were strongly convinced that their patients with moderate dementia could not use Anne independently at home. For them, ECAs could be useful tools for their daily working activities but only if controlled by professional staff. Nonetheless, positive results in terms of efficacy are reported in studies from the perspective of formal caregivers [44,46,47]. In particular, according to de Jong et al [46], virtual agents could help and support people with forgetful problems who are living independently in their homes.

Value of a User-Driven Approach

User-centered design as "a philosophy based on the needs and interests of the user, with an emphasis on making products usable and understandable" [50] is a common methodology involving end users during all iterations of the design process [51]. Four of the reviewed studies dealt with the development of the ECA through this user-driven approach: Louise [15,44] and Anne [46,47]. Louise was developed by adopting a co-design living lab approach involving older adults with dementia and stakeholders such as care professionals. The ECA called Anne [46,47] was developed using the umbrella framework of the ISO standard user-centered design [51] and providing a human-centered perspective through the active involvement of patients and their formal and informal caregivers. For us, this underlies key strengths to overcome the main barriers in applying technology for older adults in general and, in particular, for people who experience dementia. Therefore, Louise and Anne could be considered as good examples for the next generations of ECAs that will benefit from user-centered design by using input from users, patients, formal and informal caregivers as well as clinicians and other stakeholders. As reported by de Jong et al [46] and Konig et al [45], examining feedback from the point of view of family caregivers or other carers (ie, nurses, health care professionals, or care workers who work with persons with dementia) could build more awareness on how to develop effective technologies. This benefit can create better design, thus enhancing usability, user experience, acceptance, and potential market success. Nonetheless, how users come to accept and use the given technology was mainly explored by Wargnier et al [15] and Stara et al [47] who evaluated Louise's and Anne's usability and acceptance through direct observations. In addition, Carrasco et al [24] observed users during their interactions with the system and noted their observations in usability reports. None of the other studies reported in this narrative review mentioned information on the user-driven approach despite all the studies being clearly oriented to define the requirements definitions of the ECAs as a starting point of the development process. These requirements were analyzed to map functionalities and preferences for the next iterations. The user and technical requirements definitions are indeed the first actions of the user-centered design process [51].

Matching Needs With Technology

People with dementia have many changing needs during the progression of their disease, varying from memory support to almost all aspects of daily functioning. In these studies, specific ECA functions matched the physiological, comfort, and attachment needs. ECAs described in 3 of the included studies [24,27,45] sought mainly to overcome the memory problems of persons with dementia, guide them in their daily activities, and meet their need for communication and social interaction. Memory problems are certainly a basic component of dementia, but communication, emotional, and behavioral problems represent significant issues too. The evolution of this disease can easily lead these people to refuse or forget to drink, eat, and take medication and to feel more alone, apathetic, and isolated. In these cases, the ECA could play a supporting role not only in helping patients to perform their daily activities more

independently but also in stimulating dialogues and encouraging their participation in the conversation and, consequently, in maintaining social relationships. For example, in the study of Yasuda et al [27], a man with early onset dementia said, "In this system, I can talk freely without any hesitation or anxiety." This comment means that conversations with normal people were stressful to him owing to the difficulty of answering questions that he cannot reply to. As dementia severity progresses, depression and boredom increase as well as the sense of powerlessness, lack of control, and social withdrawal. For this reason, 4 of the reviewed studies [15,44,46,47] designed ECAs to meet multiple and more complex user needs. More precisely, in the studies on Louise [15,44] and Anne [46,47], the development of multi-purpose tools is described. The virtual agent Anne [46,47] offers some features such as reminder (personal and medication agenda), communication (video calls), information (news), and entertainment (games and music) that support users in all aspects of daily life. In particular, these features engage persons with dementia in various activities and help them pass the time in a more meaningful way and improve their quality of life. Louise [15,44] proposes others features to patients such as guiding through a task, cognitive stimulation exercises, and attention management. In this latter case, the ECA aimed to compensate for users' attentional disorders by performing autonomous prompting (ie, calling the user to regain attention in case of distractions). In summary, these multi-purpose agents are able to engage persons with dementia both emotionally and motivationally and stimulate their attentional skills, thereby making them feel part of a larger project and therefore less apathetic and isolated. This engagement could increase the physical, cognitive, and emotional well-being of patients but this possibility is only postulated in the mapped studies.

Theme 3: Limitations of These Studies and the Problems Faced

Common limitations and problems are reported in the selected studies. The challenge of small samples is clearly mentioned even if they underwent the preliminary evaluation of the systems [15,27,47]. This limitation is strictly correlated to the willingness to involve the entire spectrum of dementia severity [15,46,47] in order to adapt the virtual agents to the changing needs of older patients through various stages of this illness. A problem that arose concerns the importance of personalized solutions [47]. For example, a specific personalized prompting style should be programmed reaffirming the user in his or her overall persona and situational identity [45]. This means that each user has different preferences that may influence adherence to the system. For instance, people who want the ECA to address them in a very polite manner may reject the ECA considering it disrespectful if it calls them by their first name and uses a personal pronoun [15]. In addition, with regard to pleasantness, in the study of Wargnier et al [44], all users expressed a positive or very positive opinion and more than half (7/13) said they would like to be able to personalize the character's appearance. Moreover, the ECA's prompting styles is another feature that needs to be adapted since patients with neutral valence and weaker identity profiles prefer to be less in control and, therefore, did not mind a dominating prompting style. On the other hand, users with identities that were positive and powerful reported they prefer to be in control over what happens to them, thus preferring more subtle prompts [45].

Discussion

Principal Findings

This review surveyed the literature on the usage of ECAs by patients with dementia with the aim to identify the current research trends and possible gaps to cover in the future. Three main questions piloted this study: (1) what research frameworks are used to study the interaction between persons with dementia and ECAs? (2) what are the findings? and (3) what are the barriers? Only 7 papers were returned from the search. Overall, the main findings of this narrative review demonstrate that research on ECAs as an innovative way to cope with dementia is little covered in the state of the art even though interesting topics emerged from the mapped studies: the research design used to perform such studies (theme 1), major findings, the value of a user-driven approach, the importance of a well-balanced matching between the system functionalities and the users' needs (theme 2), as well as the reported problems faced by each study team (theme 3). Therefore, this section will discuss the implications raised through the lens of each theme reported in this study.

By examining the research frameworks of studies included in this review, it clearly emerges that the use of ECAs deserves (1) a more sophisticated study design and (2) proof of the efficacy of the approach, as in any other technology designed for people with dementia. The key to managing both demands is to directly involve older adults with dementia in the design of services dedicated to them. The early engagement of users from the outset and across all stages of the development cycle is relevant for people affected by dementia, since they progressively lose the ability to generalize between past and present experiences or to modify cognitive representations. For this reason, familiarity with the technology to be used is to be firmly considered and needs to be carefully planned. Indeed, while healthy adults are easily able to manage routine changes such as introducing a new device into their home environment, for older adults with dementia, these novelties can become extremely distressing and disorientating. Moreover, different limitations can overburden users: limitations in knowledge and understanding of the technology and limitations in communication between the user and the technology. As discussed in themes 1 and 2, the use of a common television or computer screen and the possibility of using verbal response as an input/output mode enables a more natural way of interaction. This is a valuable benefit for people with mild and moderate dementia. According to Kaplan and Kaplan [52], familiarity is the relationship between an individual and something that this individual has had considerable experience with. The experience leads to the development of an internal model on how one expects something to work. Some of the included studies [15,46,47] revealed that less familiarity (a low experience with new technologies) and greater dementia severity will almost certainly lead to greater difficulty in accessing technology and inevitable intervention of the caregiver who will have to spend

additional time to teach and support persons with dementia in using a new tool. This is in line with the state of the art in this field [53-57]. In fact, familiarity helps in encouraging older adults to learn and understand how to interact with new technologies by using their existing knowledge. Moreover, considering that perceived difficulties become more pronounced as dementia severity increases, the ability of a system to adapt to the changing needs and capabilities of users will determine the successful implementation of this system in everyday use.

The concept of familiarity is not the only principle to follow in the field of designing ECAs for persons with dementia. In the last decade, we conceived the important shift to a model of care centered on the person, which broke the traditional disease-focused approach. Thanks to this new paradigm, care and support are seen as ways to prevent functional decline, frailty, and disability [58] and to create a multifunctional status (ie, intrinsic capacity) to follow up over time [59,60]. Therefore, when approaching technological solutions that enable older people to remain independent at home, it is decisive to embrace the same paradigm: a model of design that follows the same path of the model of care, giving value to the person's functions and needs [61-65]. Just as it is important to disseminate a model of care centered on the patients and their needs [66-68], it would be desirable to have a model of design that allows the participation of end users to propose more customized and consequently, more effective solutions. The benefits of a personalized design will spill over to end users who will achieve a higher level of well-being because they will see their needs met and caregivers whose burden will lighten. These concepts also emerged from some of the included studies [15,45-47] discussed in the theme 2. According to de Jong et al [46], persons with dementia cannot be treated as a homogeneous group and not in the same way. There is no "one right way" to take care of them, and one tool that will "fit all" cannot be created. Wargnier et al [15] also recommended a planning of technological strategies that consider the interpersonal variability of dementia and its evolution in time for each person. Moreover, Konig et al [45] underscored the importance of understanding past and current identities of persons with impaired cognitive abilities in technology-based efforts to provide individualized care and to suggest participatory design so that personalized solutions can be provided and the quality of life can be enhanced.

This new paradigm emphasizes the power of self-determination over decisions that affect the individual's body and mind. Therefore, the individual dignity and autonomy, which are the primary values and the fundamental rights of every human being, are restored. In this new vision, patients actively participate in clinical decisions outside the old schema of only being a sick person who needs to be treated. Nowadays, well-being is the goal of dementia care that offers individualized interventions and considers the person as a whole, thus considering individuals' medical, cognitive, psychological, environmental, cultural, and social needs [69]. Such individualized intervention can be co-designed with the direct involvement of patients. Co-designing creates a common knowledge base among designers, patients, and other stakeholders on the quality of life, pains, and gains and on how to support the remaining capabilities of persons with dementia [70]. The value of co-design is well recognized in all the studies reported in this narrative review. These considerations suggest that the field of new technologies such as ECAs needs to synchronize and harmonize knowledge, efforts, and challenges in the dementia care field and the new person-centered paradigm. For example, the respect of the principle of familiarity can promote the major involvement of people with dementia in the design of artifacts from the initial stage of the development process. Moreover, this engagement can also provide a sense of continuity for them, facilitating long-term use and acceptance of assistive devices along the disease's progression. At the same time, usage continuity could open up the possibility of recruiting bigger samples of patients for enrollment in high-quality scientific research frameworks, thereby overcoming the limitations reported in the theme 3.

Another instance of harmonization is to focus on personhood and needs by clarifying what functionality and attributes are important in the new products for target users, what motivates them to use a product, what factors would hinder a positive user experience with a proposed product, and to conceptualize how parts of their lives could be improved by technology. Across the 3 themes analyzed in this review, the user experience of ECAs could be improved firstly by responding to the changing needs of people with dementia. This matching will enable technologies that better support the quality of life of people with dementia. This is particularly highlighted in the studies with more advanced ECAs such as Louise [15,44] and Anne [46,47]. Moreover, as predominantly reported in the 7 research studies, technologies should be able to adapt to the reserved skills of people with dementia without discouraging people with dementia from engagement; therefore, the natural modality of interaction by voice and the use of common screen devices are significant features to enable positive user experience. Additionally, as argued in themes 1 and 3, features such as interaction paradigm and prompting style are seen as the main components that can be personalized and used for matching the needs and capabilities of users, thereby improving the user experience. To date, the predominant use of technological solutions for safety and security [22] needs to be overcome by embracing a new paradigm that offers innovation supporting higher-level needs such as belonging, self-esteem, identity, and self-actualization [6]. The use of ECAs could be the future response to these higher-level needs and the management of everyday life across the disease's journey. In any case, this approach seeks coordination between multidisciplinary teams composed of research elements, technology developers, health care communities, formal and informal caregivers, and primary users [6] as the core of the user-driven approach [51]. As reported in this narrative review, the significant valence of the user-centered design as well as the iterative measurements of the usability and acceptance rate are still milestones to achieve during the research and devolvement process of technologies for people with dementia.

Comparison With Prior Works and Limitations

To the best of our knowledge, no other narrative reviews are reported in the literature regarding the research frameworks used to study the interaction between persons with dementia

and ECAs and between the mapped outcomes and barriers. Despite this positive aspect, there are some limitations to this review. Data sources were drawn from only 4 databases (ie, Scopus, Web of Science, PubMed, and Embase) and accessed only during a specific period of time (July 2020). The choice of using specific phrases to target the population ("patient with dementia" OR "people with dementia" OR "person with dementia") and the specific technology-based intervention ("virtual agent" OR "personal virtual assistant" OR "virtual companion" OR "embodied conversational agent") could have omitted some results from the search. It is possible that other literary sources were available in other unselected databases. However, well-known and broad-spectrum databases were used in this review. Moreover, we collected a relatively small sample of studies and excluded non-English language studies. Therefore, even if the 7 studies included in this paper were homogeneous in terms of their qualitative research design and their meeting our inclusion criteria, this may have created some biases. In addition, some authors of this review are co-authors in 2 of the reported studies [46,47]. Despite these limitations, our study offers several research directions, which may take the existing debates to the next level.

Future of ECAs

This review mapped the actual use of ECAs in the research field of dementia. The readiness level of this specific technology-based intervention grew across the years, shifting from to be initially displayed on a standard television set [24] and computer screen [15,27,44,45] to mobile standalone solutions [46,47]. This led to important achievements in the visual representation as well as in the conversational abilities of the ECAs that technologically could be seen as the foundation for advanced applications in the near future as nontherapeutic interventions to assist individuals with dementia. However, despite these promising improvements over 12 years, it remains difficult to prove that ECAs are effective to mimic interpersonal communication when interacting with users and safe to use in the care practice. Technological advances in the embodiment, content, communication modality, and strategy are not indeed the only axes of improvement since there is still to discover how preferences regarding the appearance, animation, and personalized features can influence user acceptance and efficacy of the intervention. The scarcity of evaluation and implementation phase studies underlined the necessity for further research with larger sample sizes, suitable control groups, and clinical populations but the emerging interest on the field is a gaining advancement [71,72]. Anyway, possible steps forward for the use of such systems in health care delivery can be seen in integrated platform services. For example, telemedicine, ambient intelligence, and machine learning systems can be improved through conversational agents especially in the area of health counselling, coaching, psychotherapy, and self-monitoring. Additionally, interactions between virtual agents and advanced robotics is a new design challenge [73] that could embody ECAs in social robots enforcing the attention, facial expressions, and tone of voice of future human-like robots.

Conclusions

This narrative review summarizes the current research on ECAs for patients with dementia. Technologically, these artificial characters are very interesting and the mapped studies shared promising results in terms of engagement of patients. Unfortunately, until now, it has been difficult to prove that ECAs are effective and more efforts need to be spent to achieve to this evidence. Therefore, our thematic analysis reported on 3 main themes, namely, the research frameworks used to gather users' perspectives on ECAs (theme 1), the valuable insights shared by the 7 studies as well as the value of user involvement in the development phases and the challenge of matching the system functionalities with users' needs (theme 2), and the main methodological and technical problems faced by each study team (theme 3). It emerged that this specific field of research is novel and poorly discussed in the scientific community, but possible steps forward for the use of such systems in health care delivery are predictable. Moreover, analyzing the main metaphors across the studies, our work underscored the challenge to synchronize and harmonize knowledge, efforts, and challenges within the dementia care field and its person-centered paradigm. This can be effectively possible by adopting the well-known but still little used user-centered design [51] approach, which standardizes the compelling multidisciplinary vision of research and development of innovative technologies. The challenge is therefore to enable strict collaboration between interdisciplinary research networks, medical scientists, technology developers, patients, and their formal and informal caregivers.

Acknowledgments

This study is cofunded by the EU Active and Assisted Living Program (reference AAL-call-2016-102) and partially supported by Ricerca Corrente funding from the Italian Ministry of Health.

Authors' Contributions

VS and MR conceptualized and designed the study and drafted the manuscript. LR, EF, and SP analyzed the data and reviewed the manuscript. All authors provided intellectual contributions and critical feedback and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA checklist.
[DOCX File , 22 KB - mhealth_v9i7e25381_app1.docx]

References

1. The global dementia observatory reference guide. World Health Organization. URL: https://apps.who.int/iris/handle/10665/272669 [accessed 2021-07-15]
2. Arrighi HM, Gélinas I, McLaughlin TP, Buchanan J, Gauthier S. Longitudinal changes in functional disability in Alzheimer's disease patients. Int Psychogeriatr 2013 Jun;25(6):929-937. [doi: 10.1017/S1041610212002360] [Medline: 23406898]
3. Alzheimer's Association. Alzheimer's disease facts and figures. Alzheimer's & Dementia 2020 Mar 10;16(3):391-460. [doi: 10.1002/alz.12068]
4. Maslow A. Motivation and Personality. 2nd edition. New York: Pearson; 1970.
5. Kitwood T. The experience of dementia. Aging & Mental Health 1997 Feb;1(1):13-22. [doi: 10.1080/13607869757344]
6. Koo BM, Vizer LM. Examining Mobile Technologies to Support Older Adults With Dementia Through the Lens of Personhood and Human Needs: Scoping Review. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e15122 [FREE Full text] [doi: 10.2196/15122] [Medline: 31710305]
7. Cadieux M, Garcia LJ, Patrick J. Needs of people with dementia in long-term care: a systematic review. Am J Alzheimers Dis Other Demen 2013 Dec;28(8):723-733 [FREE Full text] [doi: 10.1177/1533317513500840] [Medline: 24005852]
8. Milte R, Shulver W, Killington M, Bradley C, Ratcliffe J, Crotty M. Quality in residential care from the perspective of people living with dementia: The importance of personhood. Arch Gerontol Geriatr 2016;63:9-17 [FREE Full text] [doi: 10.1016/j.archger.2015.11.007] [Medline: 26791166]
9. Barker S, Board M. Dementia Care in Nursing. London: SAGE Publications; 2012.
10. von Kutzleben M, Schmid W, Halek M, Holle B, Bartholomeyczik S. Community-dwelling persons with dementia: what do they need? What do they demand? What do they do? A systematic review on the subjective experiences of persons with dementia. Aging Ment Health 2012;16(3):378-390. [doi: 10.1080/13607863.2011.614594] [Medline: 22250961]
11. Miller LM, Whitlatch CJ, Lyons KS. Shared decision-making in dementia: A review of patient and family carer involvement. Dementia (London) 2016 Sep;15(5):1141-1157. [doi: 10.1177/1471301214555542] [Medline: 25370075]
12. Whitlatch C, Menne H. Don't forget about me! Decision making by people with dementia. Generations. 2009. URL: https://www.ingentaconnect.com/content/asag/gen/2009/00000033/00000001/art00011 [accessed 2021-07-15]
13. Dementia: a public health priority. World Health Organization. URL: https://apps.who.int/iris/handle/10665/75263 [accessed 2021-07-15]
14. Vichitvanichphong S, Talaei-Khoei A, Kerr D, Ghapanchi A. Adoption of Assistive Technologies for Aged Care: A Realist Review of Recent Studies. 2014 Presented at: Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS); 2014; Hawaii p. 6-9.
15. Wargnier P, Benveniste S, Jouvelot P, Rigaud A. Usability assessment of interaction management support in LOUISE, an ECA-based user interface for elders with cognitive impairment. TAD 2018 Nov 26;30(3):105-126. [doi: 10.3233/tad-180189]
16. Malmgren Fänge A, Schmidt SM, Nilsson MH, Carlsson G, Liwander A, Dahlgren Bergström C, TECH@HOME Research Group. The TECH@HOME study, a technological intervention to reduce caregiver burden for informal caregivers of people with dementia: study protocol for a randomized controlled trial. Trials 2017 Feb 09;18(1):63 [FREE Full text] [doi: 10.1186/s13063-017-1796-8] [Medline: 28183323]
17. Ploeg J, McAiney C, Duggleby W, Chambers T, Lam A, Peacock S, et al. A Web-Based Intervention to Help Caregivers of Older Adults With Dementia and Multiple Chronic Conditions: Qualitative Study. JMIR Aging 2018 Apr 23;1(1):e2 [FREE Full text] [doi: 10.2196/aging.8475] [Medline: 31518230]
18. Cahill S, Begley E, Faulkner J, Hagen I. "It gives me a sense of independence" – Findings from Ireland on the use and usefulness of assistive technology for people with dementia. TAD 2007 Apr 13;19(2-3):133-142. [doi: 10.3233/tad-2007-192-310]
19. Carrillo MC, Dishman E, Plowman T. Everyday technologies for Alzheimer's disease care: Research findings, directions, and challenges. Alzheimers Dement 2009 Nov;5(6):479-488. [doi: 10.1016/j.jalz.2009.09.003] [Medline: 19896587]
20. Lorenz K, Freddolino PP, Comas-Herrera A, Knapp M, Damant J. Technology-based tools and services for people with dementia and carers: Mapping technology onto the dementia care pathway. Dementia (London) 2019 Feb;18(2):725-741. [doi: 10.1177/1471301217691617] [Medline: 28178858]
21. Evans J, Brown M, Coughlan T, Lawson G, Craven M. A systematic review of dementia focused assistive technology. 2015 Presented at: Proceedings of the 17th International Conference on Human-Computer Interaction (HCI); August 2-7; Los Angeles, CA, USA p. 406-417.
22. Meiland F, Innes A, Mountain G, Robinson L, van der Roest H, García-Casal JA, et al. Technologies to Support Community-Dwelling Persons With Dementia: A Position Paper on Issues Regarding Development, Usability, Effectiveness and Cost-Effectiveness, Deployment, and Ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1 [FREE Full text] [doi: 10.2196/rehab.6376] [Medline: 28582262]

23. Morandell M, Hochgatterer A, Fagel S, Wassertheurer S. Avatars in assistive homes for the elderly: a user-friendly way of interaction? 2008 Presented at: Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering (HCI&UE) of the Austrian Computer Society; November 20-21; Graz, Austria p. 391-402.
24. Carrasco E, Epelde G, Moreno A, Ortiz A, Garcia I, Buiza C. Natural interaction between avatars and persons with Alzheimer disease. 2008 Presented at: Proceedings of the 11th International Conference on Computers Helping People with Special Needs (ICCHP); July 9-11; Linz, Austria p. 38-45.
25. Huang H, Matsushita H, Kawagoe K, Sakai Y, Nonaka Y, Nakano Y. Toward a memory assistant companion for the individuals with mild memory impairment. 2012 Presented at: Proceedings of the 11th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC); January 2012; Kyoto, Japan p. 22-24.
26. Yaghoubzadeh R, Kramer M, Pitsch K, Kopp S. Virtual agents as daily assistants for elderly or cognitively impaired people. 2013 Presented at: Proceedings of the 13th International Workshop on Intelligent Virtual Agents (IVA); August 29-31; Edinburgh, UK p. 79-91.
27. Yasuda K, Aoe J, Fuketa M. Development of an agent system for conversing with individuals with dementia. 2013 Presented at: Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence; June 4-7; Toyama, Japan.
28. Tsiourti C, Quintas J, Ben-Moussa M, Hanke S, Nijdam N, Konstantas D. The CaMeLi framework? A multimodal virtual companion for older adults. 2016 Presented at: Proceedings of SAI Intelligent Systems Conference (IntelliSys); September 21-22; London, UK p. 196-217.
29. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
30. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
31. Shibusawa S, Huang H, Hayashi Y, Kawagoe K. Toward a virtual companion for the elderly: exploring the behaviors that potentially achieve rapport in human communication. In: Cross-Cultural Design: Cultural Differences in Everyday Life. Berlin/Heidelberg: Springer; 2013:147-156.
32. Vardoulakis L, Ring L, Barry B, Sidner C, Bickmore T. Designing relational agents as long term social companions for older adults. 2012 Presented at: Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA); September 12-14; Santa Cruz, CA, USA p. 289-302.
33. Ring L, Barry B, Totzke K, Bickmore T. Addressing loneliness and isolation in older adults: proactive affective agents provide better support. 2013 Presented at: Proceedings of the 5th Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII); September 2-5; Geneva, Switzerland p. 61-66.
34. Zancanaro M, Gabrielli S, Jameson A, Leonardi C, Not E, Pianesi F. Virtual helper or virtual card player? Contrasting responses of older users. In: Your Virtual Butler: The Making Of. Berlin/Heidelbergpringer: Springer-Verlag; 2013.
35. Fagel S, Hilbert A, Mor M, Mayer C. The virtual counselor automated character animation for ambient assisted living. 2013 Presented at: Proceedings of the 6th International Conference on Advances in Computer-Human Interactions (ACHI); February 24; Nice, France p. 184-187.
36. Maiden N, D'Souza S, Jones S, Müller L, Pannese L, Pitts K, et al. Computing technologies for reflective, creative care of people with dementia. Commun. ACM 2013 Nov;56(11):60-67. [doi: 10.1145/2500495]
37. Chittaro L, Ieronutti L, Rigutti S. Supporting presentation techniques based on virtual humans in educational virtual worlds. 2005 Presented at: Proceedings of the International Conference on Cyberworlds (CW); November; Singapore p. 245-252.
38. Ortiz A, del Puy Carretero M, Oyarzun D, Yanguas J, Buiza C, Gonzalez M. Elderly users in ambient intelligence: does an avatar improve the interaction? 2006 Presented at: Proceedings of the 9th ERCIM Workshop on User Interfaces for All; September 27-28; Königswinter, Germany p. 99-114.
39. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
40. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther 2009 Sep;89(9):873-880. [Medline: 19723669]
41. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M. Mixed Methods Appraisal Tool (MMAT) version 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria%20manual_2018%202008%202001_ENG.pdf [accessed 2021-07-15]
42. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Wargnier P, Malaisé A, Jacquemot J. Towards attention monitoring of older adults with cognitive impairment during interaction with an Embodied Conversational Agent. 2015 Presented at: Proceedings of the 3rd IEEE International Workshop on Virtual and Augmented Assistive Technology (VAAT); March 23; Arles, France p. 23-28.

45. König A, Francis LE, Joshi J, Robillard JM, Hoey J. Qualitative study of affective identities in dementia patients for the design of cognitive assistive technologies. J Rehabil Assist Technol Eng 2017 Jan;4:2055668316685038 [FREE Full text] [doi: 10.1177/2055668316685038] [Medline: 31186921]
46. de Jong M, Stara V, Von Döllen V, Bolliger D, Heerink M, Evers V. Users requirements in the design of a virtual agent for patients with dementia and their caregivers. 2018 Presented at: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good (Goodtechs); November 28-30; Bologna, Italy.
47. Stara V, de Jong M, Felici E, Bolliger D, Birrer E, von Döllen V. The design adaptation of the Virtual Assistant Anne for moderate dementia patients and their formal caregivers in protected environment tests. 2019 Presented at: Proceedings of the AHFE 2019 International Conference on Human Factors Ergonomics in Healthcare Medical Devices; July 24-28; Washington DC, USA p. 270-279.
48. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982 Sep;139(9):1136-1139. [doi: 10.1176/ajp.139.9.1136] [Medline: 7114305]
49. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
50. Norman D. The Design of Everyday Things. Cambridge, MA: MIT Press; 1998.
51. ISO 9241-210:2010 Ergonomics of human-system interaction-Part 210: Human-centered design for interactive systems. ISO. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2021-07-12]
52. Kaplan S, Kaplan R. Cognition and Environment: Functioning in an Uncertain World. New York: Praeger; 1982.
53. Boger J, Craig T, Mihailidis A. Examining the impact of familiarity on faucet usability for older adults with dementia. BMC Geriatr 2013 Jun 20;13:63 [FREE Full text] [doi: 10.1186/1471-2318-13-63] [Medline: 23786533]
54. Burmeister OK. The development of assistive dementia technology that accounts for the values of those affected by its use. Ethics Inf Technol 2016 May 25;18(3):185-198. [doi: 10.1007/s10676-016-9404-2]
55. Holthe T, Halvorsrud L, Karterud D, Hoel K, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
56. Teipel S, Babiloni C, Hoey J, Kaye J, Kirste T, Burmeister OK. Information and communication technology solutions for outdoor navigation in dementia. Alzheimers Dement 2016 Jun;12(6):695-707 [FREE Full text] [doi: 10.1016/j.jalz.2015.11.003] [Medline: 26776761]
57. Zhang H, Shen Z, Lin J, Chen Y, Miao Y. Familiarity design in exercise games for elderly. Int J Inf Technol. 2016. URL: https://www.ntulily.org/wp-content/uploads/journal/Familiarity_Design_in_Exercise_Games_for_Elderly_accepted.pdf [accessed 2021-07-15]
58. Murray CJ, Lopez AD. Measuring the global burden of disease. N Engl J Med 2013 Aug 01;369(5):448-457. [doi: 10.1056/NEJMra1201534] [Medline: 23902484]
59. de Carvalho IA, Martin F, Cesari M, Summi Y, Thiyagarajan J, Beard J. Operationalising the concept of intrinsic capacity in clinical settings. In: Background papers of the WHO clinical consortium on healthy ageing. Geneva, Switzerland: World Health Organization; 2017 Presented at: WHO clinical consortium on healthy ageing; November 21-22; Geneva, Switzerland URL: https://www.who.int/ageing/health-systems/clinical-consortium/CCHA2017-backgroundpaper-1.pdf
60. WHO clinical consortium on healthy ageing: topic focus: frailty and intrinsic capacity. Report of consortium meeting; 2016 Dec 1-2 in Geneva, Switzerland. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/272437 [accessed 2021-07-15]
61. World report on ageing and health. World Health Organization. URL: https://apps.who.int/iris/rest/bitstreams/835463/retrieve [accessed 2021-07-15]
62. Low L, Yap M, Brodaty H. A systematic review of different models of home and community care services for older persons. BMC Health Serv Res 2011 May 09;11:93 [FREE Full text] [doi: 10.1186/1472-6963-11-93] [Medline: 21549010]
63. Eklund K, Wilhelmson K. Outcomes of coordinated and integrated interventions targeting frail elderly people: a systematic review of randomised controlled trials. Health Soc Care Community 2009 Sep;17(5):447-458. [doi: 10.1111/j.1365-2524.2009.00844.x] [Medline: 19245421]
64. Woo J. Frailty, Successful Aging, Resilience, and Intrinsic Capacity: a Cross-disciplinary Discourse of the Aging Process. Curr Geri Rep 2019 Apr 25;8(2):67-71. [doi: 10.1007/s13670-019-0276-2]
65. Irving P. Self-empowerment in later life as a response to ageism. Generations 2015;39:72-77 [FREE Full text] [doi: 10.14361/9783839444221-toc]
66. Chenoweth L, Stein-Parbury J, Lapkin S, Wang A, Liu Z, Williams A. Effects of person-centered care at the organisational-level for people with dementia. A systematic review. PLoS One 2019;14(2):e0212686 [FREE Full text] [doi: 10.1371/journal.pone.0212686] [Medline: 30794651]
67. Fazio S, Pace D, Flinner J, Kallmyer B. The Fundamentals of Person-Centered Care for Individuals With Dementia. Gerontologist 2018 Jan 18;58(suppl_1):S10-S19. [doi: 10.1093/geront/gnx122] [Medline: 29361064]
68. Kim SK, Park M. Effectiveness of person-centered care on people with dementia: a systematic review and meta-analysis. Clin Interv Aging 2017;12:381-397 [FREE Full text] [doi: 10.2147/CIA.S117637] [Medline: 28255234]

69. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
70. Wang G, Marradi C, Albayrak A, van der Cammen TJM. Co-designing with people with dementia: A scoping review of involving people with dementia in design research. Maturitas 2019 Sep;127:55-63 [FREE Full text] [doi: 10.1016/j.maturitas.2019.06.003] [Medline: 31351521]
71. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
72. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
73. Luria M, Reig S, Tan X, Steinfeld A, Forlizzi J, Zimmerman J. Re-embodiment and co-embodiment exploration of social presence for robots and conversational agents. 2019 Presented at: DIS 2019 International Conference on Designing Interactive Systems; June 23-28; San Diego, USA p. e-644.

Abbreviations

ECA: embodied conversational agent

Edited by L Buis; submitted 30.10.20; peer-reviewed by K Moore, V von Wyl; comments to author 07.12.20; revised version received 27.01.21; accepted 10.03.21; published 16.07.21.

Please cite as:
Rampioni M, Stara V, Felici E, Rossi L, Paolini S
Embodied Conversational Agents for Patients With Dementia: Thematic Literature Analysis
JMIR Mhealth Uhealth 2021;9(7):e25381
URL: https://mhealth.jmir.org/2021/7/e25381
doi:10.2196/25381
PMID:34269686

Review

The Value of Mobile Health in Improving Breastfeeding Outcomes Among Perinatal or Postpartum Women: Systematic Review and Meta-analysis of Randomized Controlled Trials

Jiafen Qian[1,2*], BS; Tingting Wu[1,2*], BS; Meina Lv[1,2], BS; Zongwei Fang[1,2], BS; Mingrong Chen[1,2], BS; Zhiwei Zeng[1,2], BS; Shaojun Jiang[1,2], BS; Wenjun Chen[1,2], BS; Jinhua Zhang[1,2], PhD

[1]Department of Pharmacy, Fujian Medical University Union Hospital, Fuzhou, China

[2]College of Pharmacy, Fujian Medical University, Fuzhou, China

[*]these authors contributed equally

Corresponding Author:
Jinhua Zhang, PhD
Department of Pharmacy
Fujian Medical University Union Hospital
#29 Xinquan Road
Fuzhou, 350001
China
Phone: 86 591 83357896 83
Email: pollyzhang2006@126.com

Abstract

Background: Breastfeeding is essential for maintaining the health of mothers and babies. Breastfeeding can reduce the infection rate and mortality in newborns, and can reduce the chances of overweight and obesity in children and adolescents. For mothers, a longer duration of breastfeeding can reduce the risk of breast cancer, ovarian cancer, and type 2 diabetes. Although breastfeeding has many benefits, the global breastfeeding rate is low. With the progress of time, the popularity of mobile devices has increased rapidly, and interventions based on mobile health (mHealth) may have the potential to facilitate the improvement of the breastfeeding status.

Objective: The main objective of this study was to analyze the existing evidence to determine whether mHealth-based interventions can improve the status of breastfeeding.

Methods: We systematically searched multiple electronic databases (PubMed, Web of Science, The Cochrane Library, Embase, CNKI, WanFang, and Vip) to identify eligible studies published from 1966 to October 29, 2020. Included studies were randomized controlled trials (RCTs) studying the influence of mHealth on breastfeeding. The Cochrane Collaboration Risk of Bias tool was used to examine the risk of publication bias. RevMan 5.3 was used to analyze the data.

Results: A total of 15 RCTs with a total sample size of 4366 participates met the inclusion criteria. Compared with usual care, interventions based on mHealth significantly increased the postpartum exclusive breastfeeding rate (odds ratio [OR] 3.18, 95% CI 2.20-4.59; P<.001), enhanced breastfeeding self-efficacy (mean difference [MD] 8.15, 95% CI 3.79-12.51; P=.002; I^2=88%), reduced health problems in infants (OR 0.62, 95% CI 0.43-0.90; P=.01; I^2=0%), and improved participants' attitudes toward breastfeeding compared with usual care (MD 3.94, 95% CI 1.95-5.92; P<.001; I^2=0%). There was no significant difference in the initiation of breastfeeding within an hour of birth between the intervention group and the usual care group (OR 1.26, 95% CI 0.55-2.90; P=.59). In addition, subgroup analysis was carried out according to different subjects and publication times. The results showed that the breastfeeding rate was not limited by the types of subjects. The breastfeeding rate based on mHealth at 1 month and 2 months after delivery did not change over the time of publication (2009 to 2020), and the breastfeeding rate based on mHealth at 3 months and 6 months after delivery gradually increased with time (2009 to 2020).

Conclusions: Interventions based on mHealth can significantly improve the rate of postpartum exclusive breastfeeding, breastfeeding efficacy, and participants' attitudes toward breastfeeding, and reduce health problems in infants. Therefore, encouraging women to join the mHealth team is feasible, and breastfeeding-related information can be provided through simple measures, such as text messages, phone calls, and the internet, to improve the health of postpartum women and their babies.

(JMIR Mhealth Uhealth 2021;9(7):e26098) doi:10.2196/26098

KEYWORDS

mHealth; breastfeeding; randomized controlled trial; meta-analysis

Introduction

Proper feeding is a prerequisite for the healthy growth of babies. The World Health Organization (WHO) recommends starting exclusive breastfeeding within an hour of birth and continuing it for at least 6 months after delivery. However, maintaining breastfeeding to 2 years or longer can be beneficial to the health of both infants and mothers. For babies, early initiation and exclusive breastfeeding within 6 months can reduce the infection rate and mortality in newborns, and continuous breastfeeding for 2 years or longer can reduce the chances of overweight and obesity in children and adolescents. For mothers, a longer duration of breastfeeding can reduce the risk of breast cancer, ovarian cancer, and type 2 diabetes [1].

Although breastfeeding has many benefits, the rate of exclusive breastfeeding within 6 months in low-income countries and middle-income countries is only 37%, and in high-income countries, the duration of exclusive breastfeeding is shorter than that in low-income and middle-income countries [2]. Victora et al reported that a total of 63% of infants younger than 6 months were not breastfed, and the weighted prevalence for 6 months of exclusive breastfeeding was 20.8% [2]. Moreover, the exclusive rate was found to be only 17% in Chinese urban areas [3].

Many factors have been identified as having an impact on breastfeeding outcomes, and a key to solving the problem of the low breastfeeding rate is to improve awareness among pregnant women and mothers, as well as perform regular follow-ups [4]. Face-to-face interventions require high levels of cooperation in postpartum women, and it is easy for women to be lost to follow-up. One proposed solution is mobile health (mHealth), which could provide medical assistance with the help of electronic mobile devices. Compared to face-to-face medical assistance, mHealth is cheaper and can have improved compliance [5]. Thus, mHealth is being applied in an increasing number of fields [5-7]. A new mother's mood may change from extreme joy to tension and anxiety, which may stimulate her to use electronic mobile devices to search for breastfeeding knowledge. These therefore provide the best entry point for mHealth [8]. Information can be provided by professional medical staff or trained volunteers with breastfeeding experience [8]. Since volunteers are more likely to resonate with primiparous mothers, they may be more suitable to help primiparous women with low income or with basic or no education.

Previous research into the effectiveness of mHealth-based interventions for promoting breastfeeding have been inconclusive. Therefore, the purpose of this study was to integrate the best evidence to clarify whether these interventions can improve the current breastfeeding status.

Methods

Search Strategy

A systematic search of databases (PubMed, Embase, The Cochrane Library, and Web of Science) was conducted to identify eligible studies published from 1966 to October 29, 2020. The retrieval strategy of the PubMed database was as follows: ((“breastfeeding” OR “exclusive breastfeeding”) AND (“Mobile Applications” OR “Telemedicine” OR “Text Messaging” OR “Cell Phone” OR “Smartphone” OR “mHealth” OR “eHealth” OR “Mobile” OR “Portable Software Application” OR “Tele*” OR “e-Health” OR “m-Health” OR “?phone*” OR “Text*” OR “Short Message” OR “SMS” OR “App” OR “Apps” OR “App-based” OR “Electronic” OR “Message*” OR “Web” OR “Web-based” OR “Internet*” OR “Digital*”) AND (“randomized controlled trial” OR “controlled clinical trial” OR “randomized” OR “placebo” OR “clinical trials as topic” OR “randomly” OR “trial”) NOT (“animals”) NOT (“humans” AND “animals)). The detailed search strategy for each database is presented in Multimedia Appendix 1. To ensure that the search was comprehensive, we also searched the reference lists of the studies yielded by the original search. This study was performed in accordance with the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [9].

Study Inclusion Criteria

We included all studies that met the following requirements: (1) research subjects were pregnant or postpartum women; (2) the intervention group included studies that involved mHealth interventions, such as phone calls, text messages, and interactive computer systems, and the control group received usual care; (3) the study was a randomized controlled trial (RCT); (4) the definition of breastfeeding conformed to the WHO definition; and (5) the study mentioned the calculation of sample size and reported enough data to calculate the effect size.

Study Exclusion Criteria

Studies were excluded from the meta-analysis if (1) both the intervention and control groups accepted mHealth treatment; (2) the data could not be obtained, or the extracted data could be combined with other data; and (3) the study was not published in English.

Literature Screening and Data Extraction

Literature screening first involved reading the title and abstract to determine if the study met the inclusion criteria and then reading the full text before finally determining whether it should be included. The main data extracted were (1) the name of the first author and the date of publication; (2) research characteristics, such as the mean sample age, interventions, and sample size; and (3) outcomes, including exclusive breastfeeding rate, breastfeeding self-efficacy, health problems of infants, rate of initiation of breastfeeding within an hour of birth, and

maternal attitude to breastfeeding. Data extraction was performed independently by two reviewers. Any discrepancies were resolved by discussion or by a third investigator.

Study Quality Assessment

The bias of RCTs included in the systematic review was assessed using the Cochrane tool [10]. The following indicators of internal validity specific to the methodology of an RCT were collected: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other bias. An additional researcher was asked to conduct an evaluation to help resolve disputes that arose during the evaluation process.

Statistical Analysis

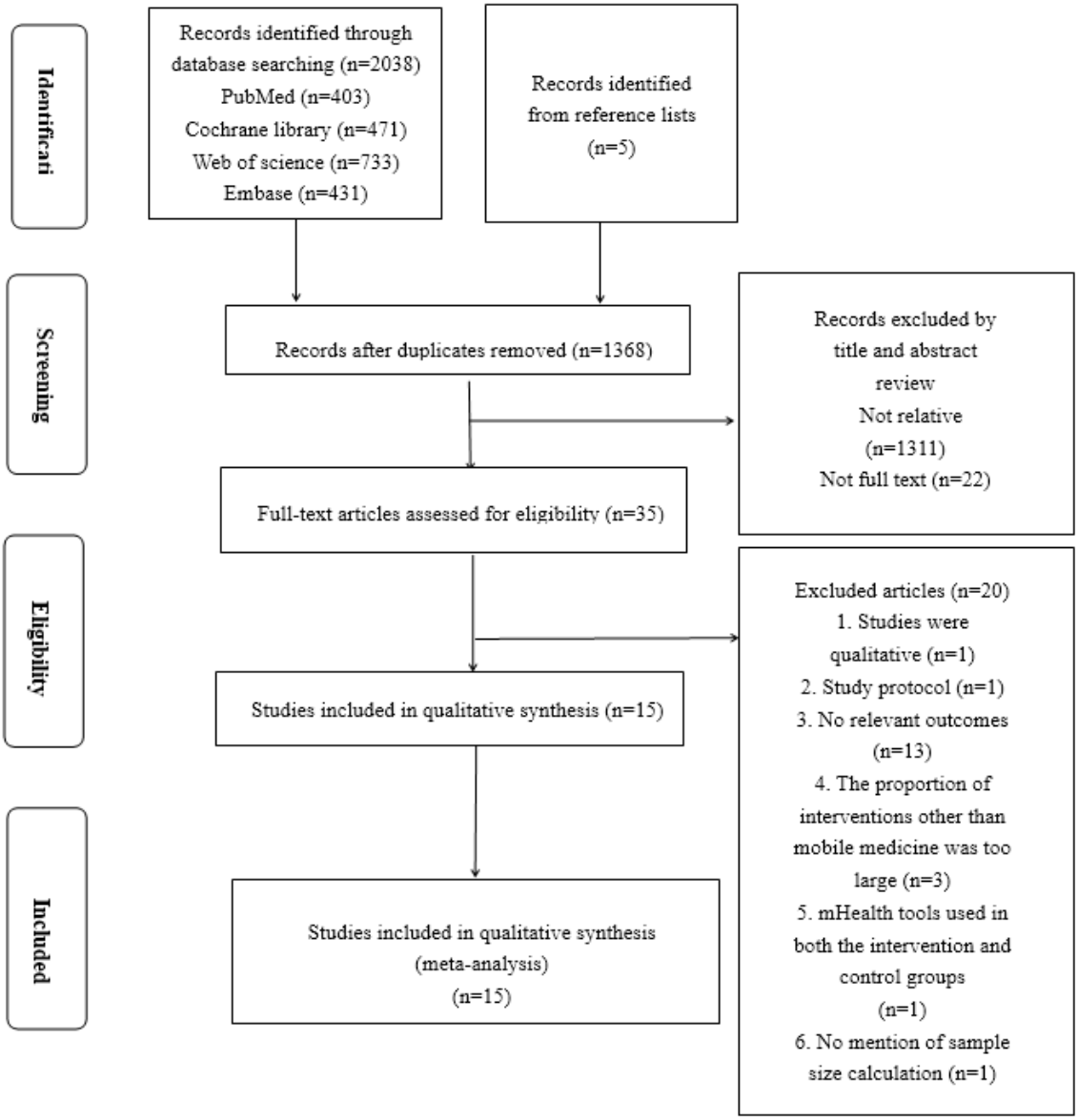
The meta-analysis was performed using RevMan 5.3. The odds ratios (ORs), mean differences (MDs), 95% CIs, and *P* values were calculated. Statistical significance was considered at a *P* value <.05. The heterogeneity among the included studies was analyzed by the chi-square test, and the test level was α=.10. If *P* was ≥.1 and I^2 was ≤50%, a fixed effects model was used for the meta-analysis. If *P* was <.1 and I^2 was >50%, a random effects model was used for the meta-analysis. We also used subgroup analysis to detect the source of heterogeneity and carried out a sensitivity analysis using the method of one-by-one exclusion.

Results

Search Results

A total of 1368 papers were found, and further screening yielded 35 papers for the full-text search. Of these, 20 papers were excluded owing to irrelevant content, failure to meet the inclusion criteria, and qualitative results. The screening process is shown in Figure 1.

Figure 1. Screening flowchart.



Study Characteristics

A total of 15 RCTs were included in this study, and the basic characteristics of the included studies are shown in Table 1 [4-8,11-20]. The intervention measures included were divided into (1) telephone support (one article mentioned this intervention), (2) SMS text messaging (one article), (3) internet intervention (seven articles), and (4) telephone, SMS text messaging, and other interventions (six articles). The subjects of 11 studies were pregnant women, and the subjects of four studies were postpartum women. The age of the subjects ranged from 16 to 49 years, and the follow-up duration ranged from 24 hours to 6 months.

Table 1. Characteristics of the clinical trials included in this study.

First author and year	Mode of intervention	Location	Type of participant	Intervention subjects, n	Control subjects, n	Outcomes
Sari, 2020 [12]	Web-based program	Turkey	Pregnant women	35	36	1. Infant prevalence
Wen, 2020 [13]	Telephone support + SMS support	Australia	Pregnant women	770	385	1. Exclusive breastfeeding rate
Uscher-Pines, 2019 [11]	Video call	North Central Pennsylvania	Postpartum women	94	93	1. Exclusive breastfeeding rate
Puharic, 2019 [5]	Telephone support + booklet	Split Dalmatia County	Pregnant women	232	123	1. Exclusive breastfeeding rate 2. Infant prevalence 3. II FAS[a]
Cavalcanti, 2019 [4]	Online social network	Northeast Brazil	Postpartum women	123	128	1. Exclusive breastfeeding rate
Patel, 2018 [20]	Telephone support + SMS support + standard management	Rural India	Women in the third trimester	519	518	1. Exclusive breastfeeding rate 2. Infant prevalence
Araban, 2018 [6]	SMS support + courses + standard management	Iran	Pregnant women	56	54	1. Exclusive breastfeeding rate 2. BSES[b], BSES-SF[c]
Ahmed, 2017 [17]	Breastfeeding monitoring system	Midwestern Hospital	Postpartum women	49	57	1. Exclusive breastfeeding rate 2. Infant prevalence
Efrat, 2016 [8]	Telephone support + standard management	Spain	Pregnant women	111	109	1. Exclusive breastfeeding rate
Flax, 2014 [14]	Telephone support + courses	Nigeria	Pregnant women	196	194	1. Exclusive breastfeeding rate
Bonuck, 2014 [7]	E-prompt	Bronx	First or second trimester of a singleton pregnancy	236	77	1. Exclusive breastfeeding rate
Scott, 2013 [16]	Web-based program	United States	Pregnant women	49	50	1. II FAS
Tahir, 2012 [18]	Telephone support + standard management	Malaysia	Postpartum women	179	178	1. Exclusive breastfeeding rate
Simonetti, 2011 [19]	Telephone support	Italy	Postpartum women	55	59	1. Exclusive breastfeeding rate
Pate, 2009 [15]	Web-based program	United States	Pregnant women	23	23	1. BSES, BSES-SF

[a] II FAS: Infant Feeding Attitude Scale (17-item 5-point scale).

[b]BSES: Breastfeeding Self-Efficacy Scale, a mother's confidence in her ability to breastfeed.

[c]BSES-SF: Breastfeeding Self-Efficacy Scale-Short Form, a measurement of exclusive breastfeeding self-efficacy (14-item 5-point scale).

Risk of Bias

The bias of RCTs included in the systematic review was assessed using the Cochrane tool, and the results are shown in Figures 2 and 3. Of the 15 papers included, seven papers did not mention the generation and allocation of random sequences, nine papers did not mention whether the participants and intervention providers were blinded, and five papers did not mention whether the evaluator was blinded. All studies reported the outcome indicators mentioned in the research protocol or method.

Figure 2. Bias risk assessment chart.

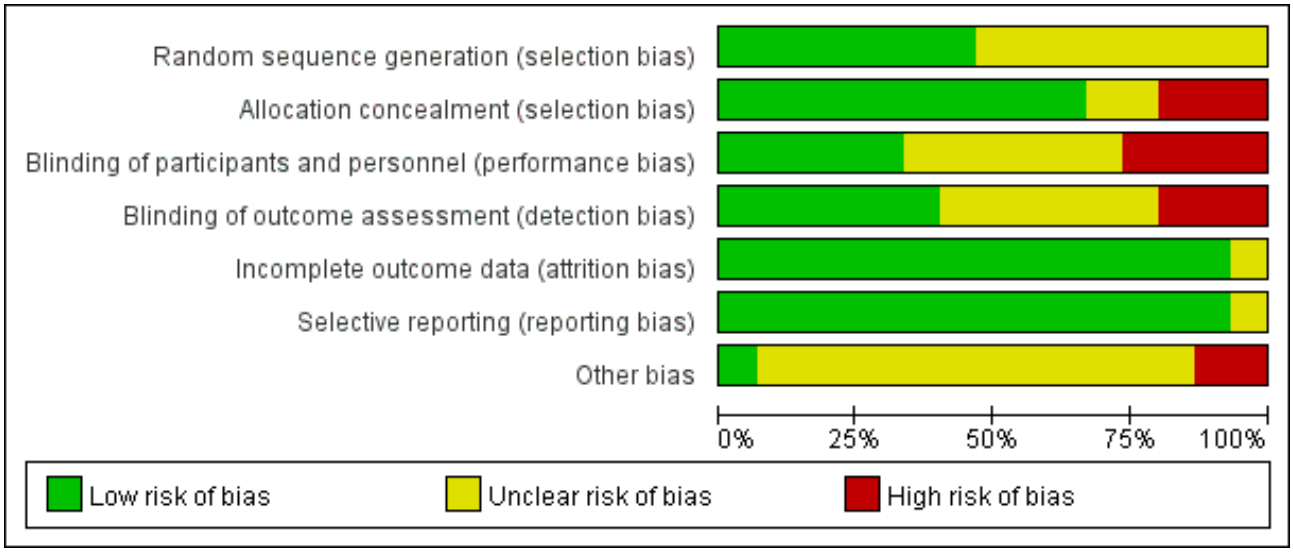
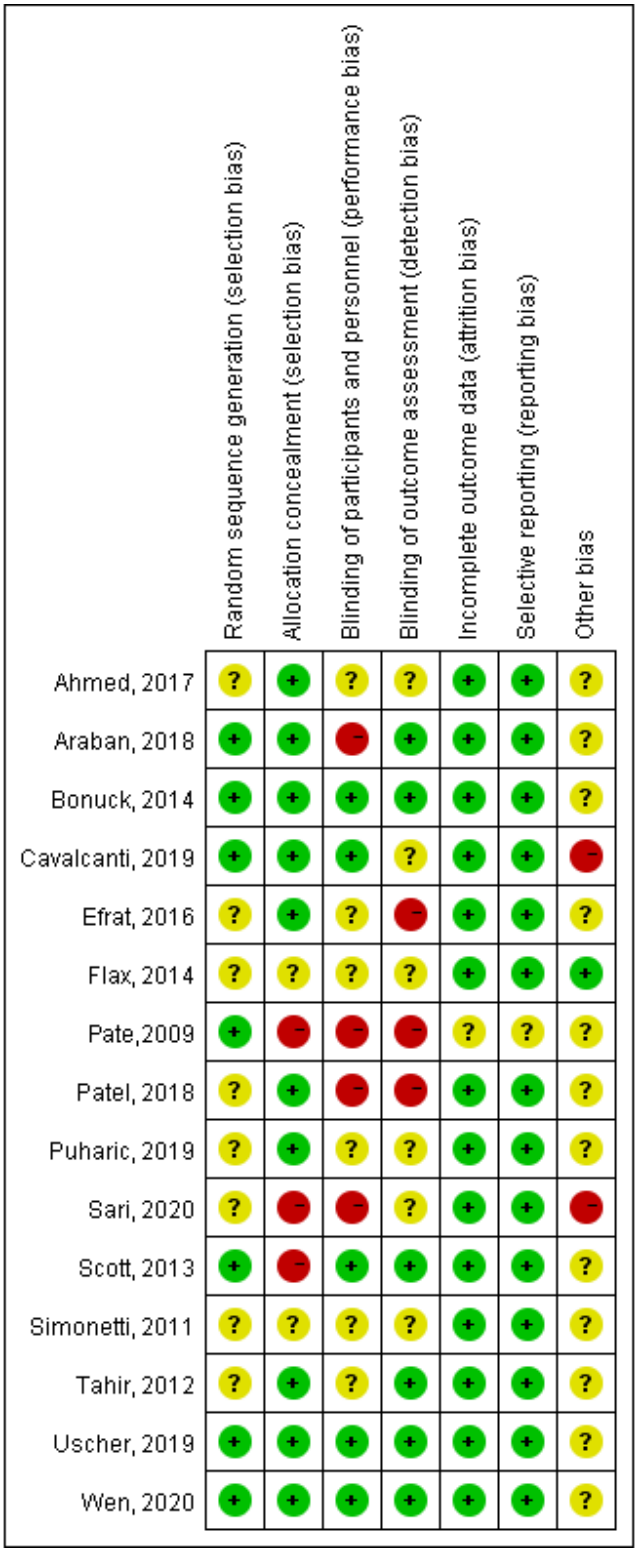


Figure 3. Summary of risk of bias.

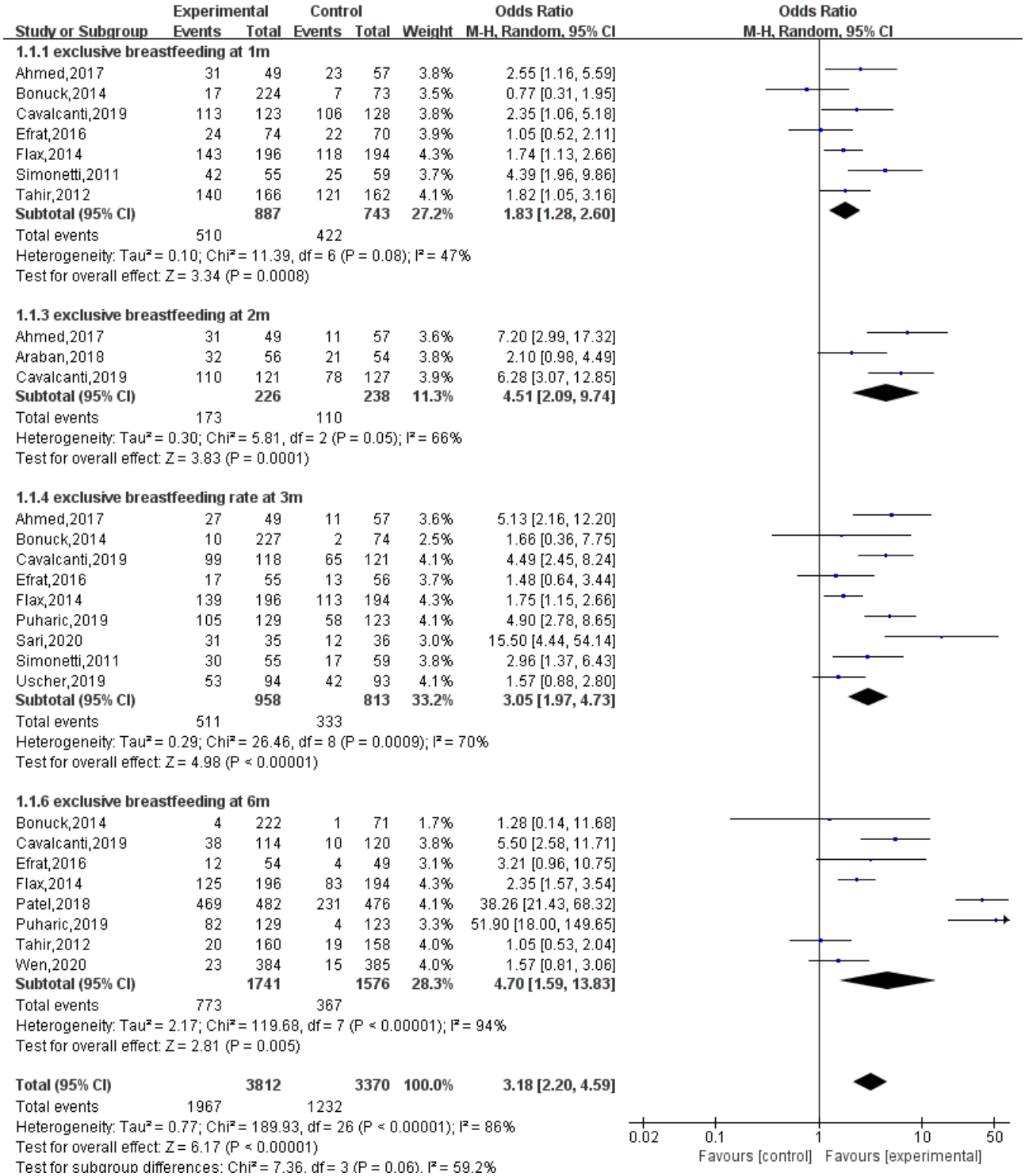


Meta-analysis Results

Exclusive Breastfeeding Rate at 1 Month After Delivery

A total of seven studies [4,7,8,14,17-19] reported exclusive breastfeeding rates at 1 month after delivery. The results of random effect model analysis showed that mHealth-based interventions significantly improved the rate of exclusive breastfeeding compared with usual care (OR 1.83, 95% CI 1.28-2.06; P<.001; I^2=74%). The sensitivity analysis showed that the results were stable. The forest plot is shown in Figure 4.

Figure 4. Forest plot of exclusive breastfeeding rates.



Exclusive Breastfeeding Rate at 2 Months After Delivery

A total of three studies [4,6,17] reported exclusive breastfeeding rates at 2 months after delivery. The results of random effect model analysis showed that mHealth-based interventions significantly improved the rate compared with usual care (OR 4.51, 95% CI 2.09-9.74; P<.001; I^2=66%). Although there was heterogeneity among the studies, the sensitivity analysis showed that the results were stable. We conducted a subgroup analysis to find heterogeneity from intervention measures, sample size, publication year, types of subjects, etc. The source of heterogeneity was not found, and we inferred that heterogeneity may have been the result of a combination of multiple factors. The forest plot is shown in Figure 4.

Exclusive Breastfeeding Rate at 3 Months After Delivery

A total of nine studies [4,5,7,8,11,12,14,17,19] reported exclusive breastfeeding rates at 3 months after delivery. The results of random effect model analysis showed that mHealth-based interventions significantly improved the rate (OR 3.05, 95% CI 1.97-4.73; P<.001; I^2=70%). Although the studies showed heterogeneity, the sensitivity analysis showed

that the results were stable. We conducted a subgroup analysis to find heterogeneity from intervention measures, sample size, publication year, types of subjects, and average number of interventions, etc. The source of heterogeneity was not found. The forest plot is shown in Figure 4.

Exclusive Breastfeeding Rate at 6 Months After Delivery

A total of eight studies [4,5,7,8,13,14,18,20] reported exclusive breastfeeding rates at 6 months after delivery. The results of random effect model analysis showed that mHealth-based interventions significantly improved the rate compared with usual care (OR 4.70, 95% CI 1.59-13.83; *P*=.005; I^2=94%). Although there was heterogeneity among the studies, the sensitivity analysis showed that the results were stable. We conducted a subgroup analysis to find heterogeneity from intervention measures, sample size, publication year, types of subjects, average number of interventions, literature quality, etc. The source of heterogeneity was not found, and we inferred that excessive heterogeneity may have been the result of a combination of multiple factors. The forest plot is shown in Figure 4.

Subgroup Analysis

In order to explore whether a different starting time of the intervention has an effect on the rate of exclusive breastfeeding, a subgroup analysis was carried out according to different types of subjects. The results of the study showed that there was no significant difference between the pregnancy group and the postpartum group for the increase in the rate of exclusive breastfeeding at 1, 2, 3, and 6 months after delivery, indicating that the time to start the intervention had no effect on the increase in the breastfeeding rate. The forest plots are shown in Figures S1-S4 in Multimedia Appendix 2.

We also conducted a subgroup analysis of the publication year. We found that the publication time of the study did not influence the breastfeeding rate at 1 and 2 months after delivery, and the reason may be that people generally think exclusive breastfeeding in the short term after delivery is very important. Therefore, it does not show a significant time effect. However, with extension of the follow-up, the publication time of the study had an impact on the breastfeeding rate. The possible reason is that with the extension of time, people stop exclusive breastfeeding due to lack of corresponding knowledge. However, with the comprehensive popularization of mobile devices in recent years, people's perceptions have changed in all directions. They are paying more attention to breastfeeding, and there are increasing number of ways to obtain breastfeeding knowledge. Thus, the breastfeeding rate at 3 months after delivery has gradually increased with time. The forest plots are shown in Figures S5-S8 in Multimedia Appendix 2.

Breastfeeding Self-Efficacy

A total of three [5,6,15] studies reported on breastfeeding efficacy. The results of random effects model analysis showed that mHealth-based interventions significantly improved breastfeeding efficacy compared with usual care (MD 8.15, 95% CI 3.79-12.51; *P*<.001; I^2=88%). Although there was heterogeneity among the studies, sensitivity analysis showed that the results were stable. Through subgroup analysis of the data, the heterogeneity between the studies was significantly reduced, indicating that the intervention measures may have been the main source of the heterogeneity. The forest plots are shown in Figures 5 and 6.

Figure 5. Forest plot of breastfeeding self-efficacy.

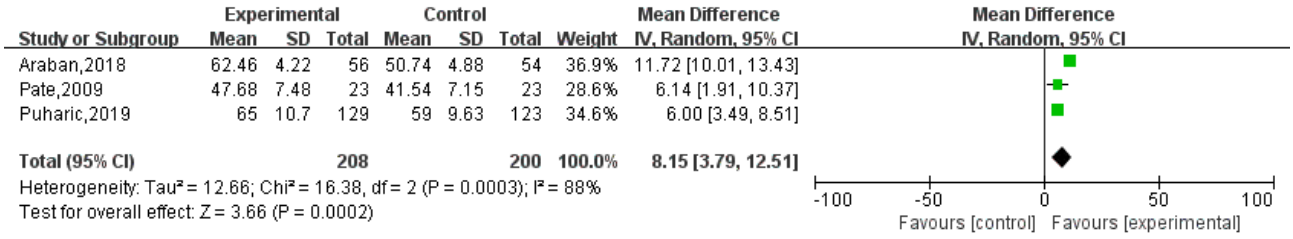
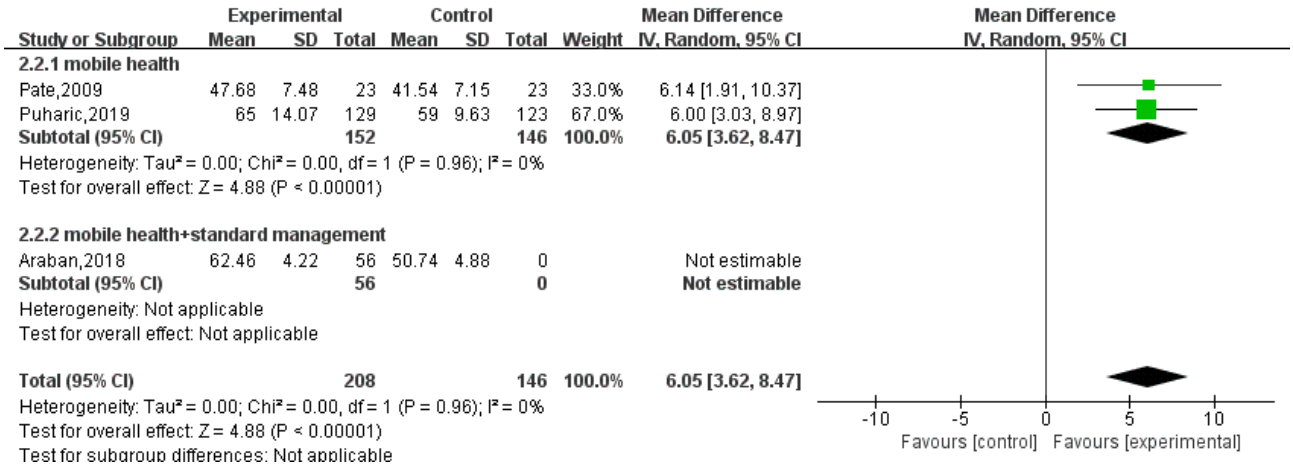


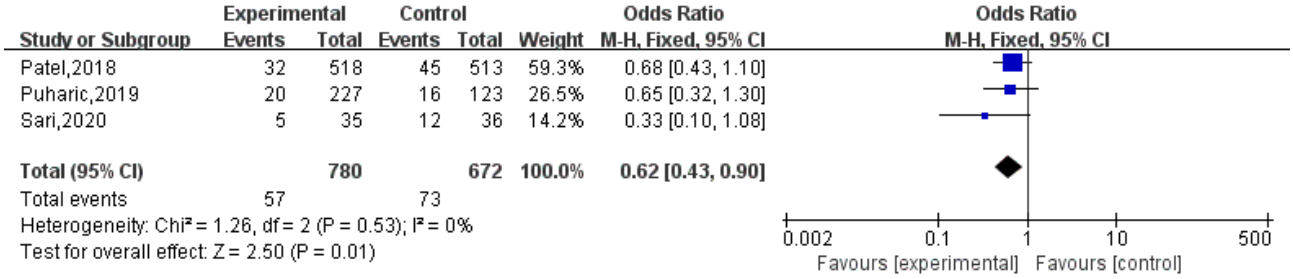
Figure 6. Forest plot of the breastfeeding self-efficacy subgroup.



Infant Hospitalization

A total of three [5,12,20] studies reported the rate of infant health problems. The results of fixed effects model analysis showed that mHealth-based interventions significantly reduced health problems in infants compared with usual care (OR 0.62, 95% CI 0.43-0.90; *P*=.01; I^2=0%). The sensitivity analysis showed that the results were unstable. The sensitivity analysis was performed by removing the studies one by one. However, after removing the study by Sari et al, there was no significant difference between the intervention and control groups (OR 0.67, 95% CI 0.46-0.99; *P*=.05; I^2=0%). The forest plot is shown in Figure 7.

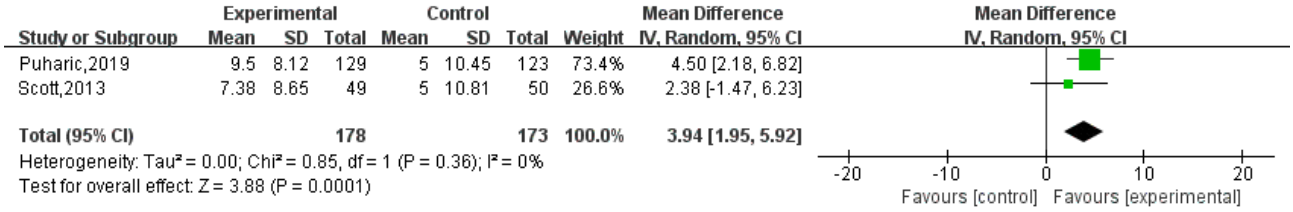
Figure 7. Forest plot of health problems of infants.



Participants' Attitudes Toward Breastfeeding

A total of two [5,16] studies reported participants' attitudes toward breastfeeding. The results of random effects model analysis showed that mHealth-based interventions significantly improved participants' attitudes toward breastfeeding compared with usual care (MD 3.94, 95% CI 1.95-5.92; *P*<.001; I^2=0%). The forest plot is shown in Figure 8.

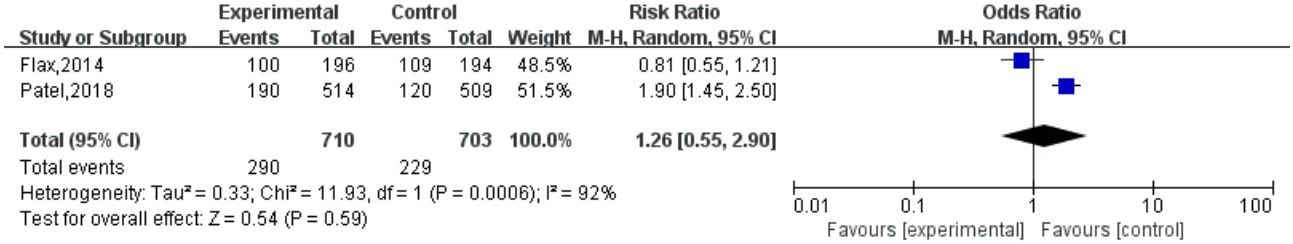
Figure 8. Forest plot of breastfeeding attitudes.



Initiation of Breastfeeding Within an Hour of Birth

Two studies [14,20] reported the rate of initiation of breastfeeding within an hour of birth. There was no significant difference in the initiation of breastfeeding within an hour of birth between the intervention group and the usual care group (OR 1.26, 95% CI 0.55-2.90; *P*=.59; I^2=92%). The forest plot is shown in Figure 9.

Figure 9. Forest plot of initiation of breastfeeding within an hour of birth.



Discussion

Principal Findings

In this meta-analysis, we included 15 RCTs comprising 4293 patients. The purpose of this meta-analysis was to evaluate whether mHealth-based interventions can improve the current breastfeeding situation compared with usual care. The meta-analysis showed that these interventions could improve the rate of exclusive breastfeeding at 1, 2, 3, and 6 months after delivery, improve breastfeeding efficacy, and reduce health problems in infants. Since breastfeeding efficacy has a great impact on postpartum breastfeeding, using mHealth interventions to enhance breastfeeding efficacy could greatly improve the breastfeeding status. As for breastfeeding attitude and the proportion of rapid initiation of breastfeeding, there was no significant difference between the groups. Thus, interventions based on mHealth are effective for improving the breastfeeding status.

In terms of health problems in infants, sensitivity analysis showed that the results were unstable. This may be related to the inconsistent follow-up duration. One paper assessed the rate over 3 months, one assessed the rate from 3 to 6 months, and one assessed the rate in the first 6 months. It may also be related to the different intervention modes used, which were telephone support and other interventions, telephone and SMS support, and internet-based support.

In terms of the exclusive breastfeeding rate, this study found that mHealth-based interventions increased the rate, and this is consistent with a study by Lee et al [21]. However, our meta-analysis investigated several factors, whereas the previous meta-analysis [21] mainly studied the effect on the health status of mothers and babies. Another meta-analysis [22] assessed the initiation of breastfeeding, breastfeeding efficacy, breastfeeding attitude, and breastfeeding duration, but there were many differences between the two studies. First, our study included 15 RCTs, whereas only six RCTs were included in the previous meta-analysis [22]. Second, our study compared the rate of exclusive breastfeeding from childbirth to 6 months postpartum, allowing the effects of mHealth-based interventions to be directly seen. The previous meta-analysis [22] only compared the duration of exclusive breastfeeding (not intuitive enough). Third, whether mobile medicine can increase the breastfeeding rate within 1 hour of birth has not been found in previous studies. Fourth, the previous meta-analysis [22] showed that mHealth interventions did not improve the efficacy of breastfeeding, whereas our study, which included more RCTs, did find an improvement with the use of mHealth interventions. We therefore conclude that mHealth is very important for promoting breastfeeding. Fifth, we explored whether interventions in different periods have an impact on the results.

The two existing measures to improve the breastfeeding status have their own advantages and limitations. One way to effectively convey health information to mothers who wish to breastfeed is mHealth-based interventions. The verbal and nonverbal communication behaviors of mHealth used by the provider can be used to build trust with the patients to improve satisfaction and adherence to the treatment plan [23]. mHealth can be widely applied to areas with low income and low medical levels in order to reduce medical expenses of postpartum women and improve their attitude toward breastfeeding. Second, it can improve maternal well-being and reduce anxiety by providing maternal and childcare information during pregnancy. With the improvement in health knowledge, maternal mental health may also be improved [22]. Third, fathers can also actively participate in pregnancy and postpartum care through the use of mobile apps. An increase in paternal participation can improve maternal confidence and attitudes toward breastfeeding, which can greatly increase the rate of health care use [24]. Finally, mHealth procedures can be used to collect pregnancy and child health data to facilitate the development of related research [21].

Limitations

Breastfeeding can not only reduce the risk of breast cancer and ovarian cancer, but also promote the healthy growth of babies [1]. Increasing numbers of mothers are realizing the importance of breastfeeding, but due to a lack of knowledge about breastfeeding, the overall level of breastfeeding in China and foreign countries is slightly low [2,3]. With the development of the economy and society, the popularity of the internet and mobile devices is increasing, which will provide an opportunity to increase the breastfeeding rate. Our research showed that the use of mHealth to convey key breastfeeding information to mothers during pregnancy or after delivery can help increase the breastfeeding rate. Therefore, we can use mHealth to provide pregnant or postpartum women with relevant knowledge and solve the problems that they encounter in the process of breastfeeding, so as to improve breastfeeding confidence and attitude, and achieve an increase in the breastfeeding rate. Women who want to increase the breastfeeding rate can positively seek help from people with knowledge of breastfeeding through telephone, text messages, the internet, and other tools before or after delivery, which can solve the problems that they encounter during breastfeeding and improve breastfeeding self-efficacy and attitude. For example, they can improve self-breastfeeding knowledge by watching breastfeeding videos on the internet or using electronic devices to communicate with breastfeeding professionals about the problems they encounter.

The study has several limitations. First, there was insufficient literature on several outcomes, which may lead to bias. For example, the outcome of rapid initiation of breastfeeding requires more data to obtain more reliable results. Second, the results of the sensitivity analysis of some outcomes were not stable, which may lead to bias, and they need to be further verified by more studies. Third, several articles were not highly representative. For example, the research subjects in the study by Flax et al [14] were women with microfinance, and since this population is not representative, these results should not necessarily be extrapolated to the whole population. Third, the source of heterogeneity needs to be further assessed in future research.

Conclusion

Our study found that interventions based on mHealth can improve the rate of exclusive breastfeeding, the breastfeeding attitude of mothers, and breastfeeding efficiency, and reduce health problems in infants. In view of the universality of mobile devices, mHealth can be used to promote the health of pregnant mothers and infants. The meta-analysis found limited improvement in rapid initiation of breastfeeding with mHealth interventions. More clinical studies are needed to confirm this view. In general, interventions based on mHealth can improve the breastfeeding status.

Acknowledgments

This study was funded by the Natural Science Foundation of Fujian Province, China (2018Y0037) and Fujian Provincial Health Technology Project, China (2019-CX-19).

Authors' Contributions

JZ initiated the study. JQ, TW, and ML performed data extraction and analyses. JQ drafted the first version of the manuscript. JZ and TW critically reviewed the manuscript and revised it. All authors made substantial contributions to the concept and design of the study, interpreted the data, and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy of the databases.
[DOCX File , 17 KB - mhealth_v9i7e26098_app1.docx]

Multimedia Appendix 2
Forest plots of subgroup analysis.
[DOCX File , 174 KB - mhealth_v9i7e26098_app2.docx]

References

1. Kramer MS, Kakuma R. Optimal duration of exclusive breastfeeding. Cochrane Database Syst Rev 2012 Aug 15(8):CD003517 [FREE Full text] [doi: 10.1002/14651858.CD003517.pub2] [Medline: 22895934]
2. Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. The Lancet 2016 Jan 30;387(10017):475-490. [doi: 10.1016/S0140-6736(15)01024-7] [Medline: 26869575]
3. Ke J, Ouyang Y, Redding SR. Family-Centered Breastfeeding Education to Promote Primiparas' Exclusive Breastfeeding in China. J Hum Lact 2018 May;34(2):365-378. [doi: 10.1177/0890334417737293] [Medline: 29161530]
4. Cavalcanti DS, Cabral CS, de Toledo Vianna RP, Osório MM. Online participatory intervention to promote and support exclusive breastfeeding: Randomized clinical trial. Matern Child Nutr 2019 Jul;15(3):e12806 [FREE Full text] [doi: 10.1111/mcn.12806] [Medline: 30825414]
5. Puharić D, Malički M, Borovac JA, Šparac V, Poljak B, Aračić N, et al. The effect of a combined intervention on exclusive breastfeeding in primiparas: A randomised controlled trial. Matern Child Nutr 2020 Jul;16(3):e12948 [FREE Full text] [doi: 10.1111/mcn.12948] [Medline: 31943761]
6. Araban M, Karimian Z, Karimian Kakolaki Z, McQueen KA, Dennis C. Randomized Controlled Trial of a Prenatal Breastfeeding Self-Efficacy Intervention in Primiparous Women in Iran. J Obstet Gynecol Neonatal Nurs 2018 Mar;47(2):173-183. [doi: 10.1016/j.jogn.2018.01.005] [Medline: 29406289]
7. Bonuck K, Stuebe A, Barnett J, Labbok MH, Fletcher J, Bernstein PS. Effect of primary care intervention on breastfeeding duration and intensity. Am J Public Health 2014 Mar;104 Suppl 1:S119-S127. [doi: 10.2105/AJPH.2013.301360] [Medline: 24354834]
8. Efrat MW, Esparza S, Mendelson SG, Lane CJ. The effect of lactation educators implementing a telephone-based intervention among low-income Hispanics: A randomised trial. Health Educ J 2015 Jul;74(4):424-441 [FREE Full text] [doi: 10.1177/0017896914542666] [Medline: 26941454]
9. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
10. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane. URL: https://training.cochrane.org/handbook [accessed 2021-06-30]
11. Uscher-Pines L, Ghosh-Dastidar B, Bogen DL, Ray KN, Demirci JR, Mehrotra A, et al. Feasibility and Effectiveness of Telelactation Among Rural Breastfeeding Women. Acad Pediatr 2020 Jul;20(5):652-659. [doi: 10.1016/j.acap.2019.10.008] [Medline: 31629118]
12. Sari C, Altay N. Effects of providing nursing care with web-based program on maternal self-efficacy and infant health. Public Health Nurs 2020 May;37(3):380-392. [doi: 10.1111/phn.12712] [Medline: 32017251]
13. Wen LM, Rissel C, Xu H, Taki S, Buchanan L, Bedford K, et al. Effects of Telephone and Short Message Service Support on Infant Feeding Practices, "Tummy Time," and Screen Time at 6 and 12 Months of Child Age: A 3-Group Randomized Clinical Trial. JAMA Pediatr 2020 Jul 01;174(7):657-664 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.0215] [Medline: 32282034]
14. Flax VL, Negerie M, Ibrahim AU, Leatherman S, Daza EJ, Bentley ME. Integrating group counseling, cell phone messaging, and participant-generated songs and dramas into a microcredit program increases Nigerian women's adherence to international breastfeeding recommendations. J Nutr 2014 Jul;144(7):1120-1124 [FREE Full text] [doi: 10.3945/jn.113.190124] [Medline: 24812071]

15. Pate BL. Effectiveness of web-based programs in improving breastfeeding self-efficacy. CiteSeerX. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.5559&rep=rep1&type=pdf [accessed 2021-06-30]
16. Scott BL. The effectiveness of internet and instant messaging approaches in promoting intention to breastfeed. DOC88. URL: http://www.doc88.com/p-7002366055936.html [accessed 2021-06-30]
17. Ahmed AH, Roumani AM, Szucs K, Zhang L, King D. The Effect of Interactive Web-Based Monitoring on Breastfeeding Exclusivity, Intensity, and Duration in Healthy, Term Infants After Hospital Discharge. J Obstet Gynecol Neonatal Nurs 2016;45(2):143-154 [FREE Full text] [doi: 10.1016/j.jogn.2015.12.001] [Medline: 26779838]
18. Tahir NM, Al-Sadat N. Does telephone lactation counselling improve breastfeeding practices? A randomised controlled trial. Int J Nurs Stud 2013 Jan;50(1):16-25. [doi: 10.1016/j.ijnurstu.2012.09.006] [Medline: 23084438]
19. Simonetti V, Palma E, Giglio A, Mohn A, Cicolini G. A structured telephonic counselling to promote the exclusive breastfeeding of healthy babies aged zero to six months: a pilot study. Int J Nurs Pract 2012 Jun;18(3):289-294. [doi: 10.1111/j.1440-172X.2012.02040.x] [Medline: 22621300]
20. Patel A, Kuhite P, Puranik A, Khan SS, Borkar J, Dhande L. Effectiveness of weekly cell phone counselling calls and daily text messages to improve breastfeeding indicators. BMC Pediatr 2018 Oct 30;18(1):337 [FREE Full text] [doi: 10.1186/s12887-018-1308-3] [Medline: 30376823]
21. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: Systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
22. Lau Y, Htun TP, Tam WS, Klainin-Yobas P. Efficacy of e-technologies in improving breastfeeding outcomes among perinatal women: a meta-analysis. Matern Child Nutr 2016 Jul 21;12(3):381-401 [FREE Full text] [doi: 10.1111/mcn.12202] [Medline: 26194599]
23. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
24. Chan KL, Chen M. Effects of Social Media and Mobile Health Apps on Pregnancy Care: Meta-Analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]

Abbreviations

OR: odds ratio
MD: mean difference
mHealth: mobile health
RCT: randomized controlled trial
WHO: World Health Organization

Edited by L Buis; submitted 27.11.20; peer-reviewed by S Mukherjee, M Herron, N Allen; comments to author 01.02.21; revised version received 20.03.21; accepted 12.04.21; published 16.07.21.

Please cite as:
Qian J, Wu T, Lv M, Fang Z, Chen M, Zeng Z, Jiang S, Chen W, Zhang J
The Value of Mobile Health in Improving Breastfeeding Outcomes Among Perinatal or Postpartum Women: Systematic Review and Meta-analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2021;9(7):e26098
URL: https://mhealth.jmir.org/2021/7/e26098
doi:10.2196/26098
PMID:34269681

Review

Smartphone-Based Interventions for Physical Activity Promotion: Scoping Review of the Evidence Over the Last 10 Years

Alex Domin[1], BA, MPhil; Donna Spruijt-Metz[2], MFA, PhD; Daniel Theisen[3], MSc, PhD; Yacine Ouzzahra[4], BSc, MSc, PhD; Claus Vögele[1], BSc, MSc, PhD

[1]Research Group: Self-Regulation and Health, Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]USC mHealth Collaboratory, Center for Economic and Social Research, University of Southern California, Los Angeles, CA, United States

[3]ALAN – Maladies Rares Luxembourg, Kockelscheuer, Luxembourg

[4]Research Support Department, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding Author:
Alex Domin, BA, MPhil
Research Group: Self-Regulation and Health
Department of Behavioural and Cognitive Sciences
University of Luxembourg
Maison des Sciences Humaines
11, Porte des Sciences
Esch-sur-Alzette, L-4366
Luxembourg
Phone: 352 46 66 44 9389
Fax: 352 46 66 44 39389
Email: alex.domin@uni.lu

Abstract

Background: Several reviews of mobile health (mHealth) physical activity (PA) interventions suggest their beneficial effects on behavior change in adolescents and adults. Owing to the ubiquitous presence of smartphones, their use in mHealth PA interventions seems obvious; nevertheless, there are gaps in the literature on the evaluation reporting processes and best practices of such interventions.

Objective: The primary objective of this review is to analyze the development and evaluation trajectory of smartphone-based mHealth PA interventions and to review systematic theory- and evidence-based practices and methods that are implemented along this trajectory. The secondary objective is to identify the range of evidence (both quantitative and qualitative) available on smartphone-based mHealth PA interventions to provide a comprehensive tabular and narrative review of the available literature in terms of its nature, features, and volume.

Methods: We conducted a scoping review of qualitative and quantitative studies examining smartphone-based PA interventions published between 2008 and 2018. In line with scoping review guidelines, studies were not rejected based on their research design or quality. This review, therefore, includes experimental and descriptive studies, as well as reviews addressing smartphone-based mHealth interventions aimed at promoting PA in all age groups (with a subanalysis conducted for adolescents). Two groups of studies were additionally included: reviews or content analyses of PA trackers and meta-analyses exploring behavior change techniques and their efficacy.

Results: Included articles (N=148) were categorized into 10 groups: commercial smartphone app content analyses, smartphone-based intervention review studies, activity tracker content analyses, activity tracker review studies, meta-analyses of PA intervention studies, smartphone-based intervention studies, qualitative formative studies, app development descriptive studies, qualitative follow-up studies, and other related articles. Only 24 articles targeted children or adolescents (age range: 5-19 years). There is no agreed evaluation framework or taxonomy to code or report smartphone-based PA interventions. Researchers did not state the coding method, used various evaluation frameworks, or used different versions of behavior change technique taxonomies. In addition, there is no consensus on the best behavior change theory or model that should be used in smartphone-based interventions for PA promotion. Commonly reported systematic practices and methods have been successfully identified. They include PA recommendations, trial designs (randomized controlled trials, experimental trials, and rapid design trials), mixed

methods data collection (surveys, questionnaires, interviews, and focus group discussions), scales to assess app quality, and industry-recognized reporting guidelines.

Conclusions: Smartphone-based mHealth interventions aimed at promoting PA showed promising results for behavior change. Although there is a plethora of published studies on the adult target group, the number of studies and consequently the evidence base for adolescents is limited. Overall, the efficacy of smartphone-based mHealth PA interventions can be considerably improved through a more systematic approach of developing, reporting, and coding of the interventions.

(JMIR Mhealth Uhealth 2021;9(7):e24308) doi:10.2196/24308

KEYWORDS

scoping review; smartphone application; physical activity; behavior change; mobile health; research design; mHealth; adolescents; adults; BCT; mobile phonescoping review; smartphone application; physical activity; behavior change; mobile health; research design; mHealth; adolescents; adults; BCT; mobile phone

Introduction

Background

Physical inactivity has been identified as a *global pandemic* and is reported to be the fourth leading cause of death worldwide [1]. There is strong evidence that physical inactivity shortens life expectancy and increases the risk of noncommunicable diseases such as breast and colon cancers, type 2 diabetes, and coronary heart disease, resulting in 5.3 million deaths annually worldwide [2]. Moreover, the world economy suffers great financial losses because of physical inactivity, bearing a yearly estimated burden of US $53.8 billion health care costs worldwide [3]. To avoid these health and financial consequences, it is important to pursue pre-emptive strategies to identify and mitigate the causes of low levels of physical activity (PA).

At the same time, the world is facing another life-threatening pandemic caused by SARS-CoV-2 or COVID-19 [4]. The World Health Organization (WHO) declared the virus outbreak as a pandemic on March 11, 2020, and more than 5.5 million cases of COVID-19 worldwide have been reported since, resulting in more than 346,600 deaths as of May 26, 2020 [5]. As a response to this crisis, many governments introduced confinements, curfews, or quarantines as compulsory or recommended containment and prevention measures [6]. Several studies have since found that home quarantine introduces a shift in lifestyle toward limited socialization and reduced PA, which may contribute to an exacerbation of already reduced PA levels in the population and its associated health risks [7,8].

Although confinement measures have been introduced to reduce the spread of the virus, with some success in flattening the curve, these interventions to contain the COVID-19 outbreak have unsurprisingly resulted in an increased use of digital communication technologies, such as in mobile health (mHealth) and telehealth approaches in the domains of PA and medicine [9-12]. In light of these developments, and the resulting increase in the importance of digital technologies for health, it has become even more evident that it is crucial to significantly advance the field of mHealth PA technologies by identifying knowledge gaps, evaluating reporting processes, and establishing best practices. This scoping review, therefore, focuses on the analysis of the development and evaluation trajectory of mHealth PA interventions and on the review of systematic theory- and evidence-based practices and methods that are implemented along this trajectory. We describe the advantages and disadvantages of theory- and evidence-based practices and methods to present recommendations on how to improve and accelerate the overall process of the development and evaluation of mHealth PA interventions. The overall aim of this review is to provide guidance in the field of smartphone- and wearable-based mHealth PA interventions.

A major decline in PA levels occurs during the transition from childhood to adolescence [13,14]. A high percentage of the global population of adolescents does not reach the levels of PA recommended by the WHO [15,16]. Insufficient levels of PA tend to track through childhood and adolescence into adulthood [17-19]. According to the report *Health at a Glance: Europe 2016* from the Organization for Economic Co-operation and Development, 36% of the adult population of the European Union does not meet the recommended levels of PA. According to the same report, the majority of reported adolescents in the European Union, by the age of 15, do not even reach 30% of the recommended PA time [16]. Given the scale of the problem and the fact that higher PA is associated with physical [20] and mental [21] health benefits, it is important to develop interventions that can effectively support and promote PA, which can reach large numbers of people easily and that can do this low-touch or remotely, and at low cost.

Face-to-face interventions are resource intensive and limited because of their attachment to their specific environment and multicomponent nature [22]. They can be difficult to access depending on circumstances such as a busy schedule, illness, childcare, lack of safe and attractive spaces to exercise, or, as has now been demonstrated, disasters such as the COVID-19 pandemic. Smartphones and affordable wearable sensors have become ubiquitous in the lives of today's population [23]. These devices could be beneficial for the development and delivery of remote PA interventions [22,24,25]. The advantages of smartphones and devices integrated into smartphone platforms include the ability to schedule the delivery of intervention content that can take into account the time of day and momentary environment of the user. These technologies offer the possibility of high-level personalization toward the user and the unobtrusive and in situ collection of behavioral data [26]. Therefore, smartphone-based interventions are accessible, scalable, comparatively inexpensive, and can deliver low-touch or completely remote interventions. These features make

smartphone-based interventions more advantageous for self-monitoring of PA compared with stand-alone pedometers and are preferable over computer-based interventions [27,28]. Several reviews of smartphone-based interventions have outlined their acceptability, efficacy, and effectiveness in increasing health behaviors in several age groups [24,29,30].

Despite their strong potential, the evidence concerning smartphone-based interventions to improve PA and decrease sedentary behaviors (SBs) is only emerging, and the literature is poorly systematized, which results in methodological inconsistencies and significant gaps in our understanding of the developments in the field of mHealth PA interventions.

Prior Work

There are four recent scoping reviews, which attempted to address these gaps [31-34]. Lee et al [34] aimed to identify the *efficacy* and *effectiveness* of mHealth PA interventions in adolescents; Aromatario et al [33] investigated how researchers conducting studies with mHealth PA and diet apps as a main component assess the app *conditions ofeffectiveness* across age groups; McCallum et al [32] explored the extent to which evaluations of mHealth PA apps and wearables affect the *effectiveness, engagement, and acceptability* of these apps, and Ly [31] reviewed the literature with the aim of presenting *an account of the current knowledge* on the use of mHealth interventions to enhance PA levels in young adults. These reviews included studies evaluating a range of different target populations with various states of health or ill health (without illness; chronic illness [33], including attention deficit/hyperactivity disorder [34]; cancer and diabetes [32]; and acute illness [33]) while targeting either PA alone [31], diet alone [32], or a combination of the two [33]. Finally, yet importantly, almost all reviews (excluding Aromatario et al [33], who focused on mHealth app only) included studies with various modes of delivery of the intervention, such as smartphone apps, websites, SMS text message, tablets, and PDAs.

Although these reviews are informative and have their strengths in different areas, they still fail to provide answers to several questions. First, behavior change components of mHealth interventions are often conceptualized as behavior change techniques (BCTs), which are described systematically in various BCT taxonomies [35-37]. However, there is no consensus on a universally accepted behavior change taxonomy. Therefore, it remains unclear why certain authors prefer one taxonomy to another. Second, studies on smartphone-based interventions fall under the domain of mHealth, which is commonly defined as a *medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, PDAs, and other wirelessdevices* [38]. This definition is currently argued to be outdated, as PDAs were largely discontinued after the extensive adoption of smartphones in the early 2010s, resulting in patient monitoring and other devices becoming less popular and attractive in health care compared with smartphones [39-41]. As a result, recent publications in the mHealth domain are mostly related to smartphone-based interventions, and it remains unclear whether it is advantageous to include outdated devices in current reviews [42]. Third, most of the published studies on smartphone-based interventions include exclusively adults, despite the importance of PA levels during adolescence. Finally, and most importantly, current reviews do not provide an exhaustive review of systematic practices and methods along the trajectory of the development and evaluation of smartphone-based mHealth interventions for PA promotion, as they mostly focus on reviewing specific aspects of interventions, such as effectiveness and validity. Overall, these reviews lack a clear representation of the mHealth PA development trajectory and the tools available for researchers along this trajectory (eg, taxonomies, theories). We argue that a clearer understanding of these would significantly improve the quality of development, end product, and reporting of mHealth interventions and would contribute to the development of *theory-based* rather than *theory-inspired* interventions.

Goal of This Review

This scoping review addresses these issues. It includes studies describing or evaluating smartphone apps alone or in combination with wearables as a primary intervention component to enhance PA levels, focusing on studies with healthy individuals without chronic or acute conditions (excluding cardiovascular diseases and obesity), and targeting studies with PA as a primary outcome. Although we included all age groups to provide a comprehensive review, we focused on one part of the analysis on studies involving adolescents, as the biggest impact on future generations' health is to be expected from changing their behavior. The primary objective of this scoping review is to *analyze the development and evaluation trajectory of mHealth PA interventions* and to *reviewsystematic theory- and evidence-based practices and methods* that are implemented along this trajectory. The secondary objective of this review is to *identify the range of evidence (both quantitative and qualitative) available on smartphone-based mHealth PA interventions* to provide a comprehensive tabular and narrative *review of the available literature* in terms of its nature, features, and volume.

This review is guided by the following research questions: (1) What kind of literature is available in the field, and how can the existing literature be categorized? (2) Which theories and techniques are implemented in smartphone-based PA interventions to support behavior changes, and how are these theories and techniques systematized? (3) Which practices and methods are used to systematically develop and evaluate smartphone-based PA interventions? and (4) Which devices and primary outcomes are used for data collection and analysis in smartphone-based PA interventions?

Methods

Study Design

Methodological guidelines for scoping reviews developed by Arksey and O'Malley [43], extended by Levac et al [44] and Peters et al [45], were accommodated, and the methodology adopted by McCallum et al [32] was implemented. In accordance with these guidelines, studies were not rejected based on their research design or quality.

Identification of Relevant Articles

The literature search was conducted from September 2017 to August 2018 in three databases: MEDLINE/PubMed, ScienceDirect, and ResearchGate. The search was limited to studies published in 2008 and later, as Apple App Store and Google Play (formerly known as the Android Market) started in July and October of that year. Only publications in English were considered. Full papers, study protocols, conference proceedings, dissertations, and books were considered eligible. Reference lists of germane articles and review studies were manually searched to identify potentially relevant articles. The articles were initially screened by the first author (AD). As per best review practice, an assistant reviewer independently reviewed the eligibility of articles for inclusion in the review. Inconsistencies were resolved by discussion and consensus between the 2 reviewers.

Search strategies for MEDLINE/PubMed were developed using a combination of thesaurus and free terms based on Boolean logic (Table 1).

Table 1. Search builder for MEDLINE/PubMed.

Search lines	Search terms	Filtered by
Line 1	mobile phone OR cell phone OR smartphone OR smart phone OR smart-phone OR mobile device OR iphone OR mobile technology OR mhealth OR android	Title or abstract
2. AND	app OR apps OR application OR intervention OR trial OR behavior OR behaviour	Title or abstract
3. AND	physical activity OR exercise OR fitness	Title or abstract
4. NOT	heart attack OR heart failure OR cancer OR diabetes OR diabetic OR injury OR injuries OR alcohol OR sexual OR e-learning OR home OR HIV OR pain OR sleep OR smoke OR smoking OR epileptic OR rehabilitation OR asthma	Title

The use of this search builder was not possible for ResearchGate and ScienceDirect. Consequently, various combinations of the following search terms were used: *mobile phone*, *cell phone*, *smartphone*, *smart phone*, *smart-phone*, *mobile device*, *iphone*, *mobile technology*, *mhealth*, *android*, *app*, *apps*, *application*, *intervention*, *trial*, *behavior*, *behaviour*, *physical activity*, *exercise*, and *fitness*.

To select articles that were related to mHealth interventions with the primary outcome in PA, the following terms were used to manually filter out articles from the initial search results: *weight*, *eat*, *nutrition*, *diet*, and *game*.

Study Selection

Although all age groups were included, an additional subanalysis for adolescents' target groups was conducted (specifically accounting for BCTs effective for this target population). This was also done to contrast the differences in BCTs used in adolescents and other target populations. Studies were included if (1) the primary component of the intervention involved a mobile app targeting PA and SB and (2) the study used smartphones with available embedded sensors alone (stand-alone intervention) or in conjunction with other external components, for example, accelerometers, pedometers, and websites accessed through desktop computers (multicomponent interventions). Studies were excluded from the review if (1) the intervention was limited to using text messages only, (2) the app was used for data collection only (eg, phone-based questionnaires), (3) the intervention included any mobile device other than smartphone or PA tracker, for example, PDAs, (4) the intervention targeted other preventive health issues, such as alcohol abuse, smoking, and sport injuries, and (5) they focused on patients with chronic conditions other than cardiovascular diseases and obesity, for example, diabetes mellitus. This review includes experimental and descriptive studies, as well as reviews addressing smartphone-based mHealth interventions aimed at promoting PA. Two additional groups of studies were included: reviews or content analyses of PA trackers and meta-analyses exploring BCTs and their efficacy. This approach was used to obtain additional evidence from the domains, which are closely related to smartphone-based mHealth PA promotions, to provide theoretical evidence related to the field and to present the latest developments in the domain. Instead of considering studies using combined interventions designed to reduce body weight (ie, PA promotion and dietary interventions), we aimed to include studies promoting PA and reducing sedentary time, as it is difficult to disentangle the effects of specific intervention BCTs on particular behaviors in studies targeting several health behaviors. For example, a BCT such as *adding objects to the environment* as a part of one intervention may be successful in terms of changing eating behaviors, while having a neutral or even negative effect on PA outcomes. Therefore, we tried to avoid drawing conclusions on the effectiveness of BCTs across interventions targeting PA behavior only and interventions targeting PA, eating, and other behaviors.

Data Extraction, Collation, Summary, and Reporting of Results

A data extraction form was developed specifically for this review and served as a basis for Tables S1-S10 presented in Multimedia Appendix 1 [2,22,24-30,35-37,46-181]. A mixed methods descriptive approach was adopted to analyze the extracted data [32]. The identified articles were categorized into 10 groups: commercial smartphone app content analyses, smartphone-based intervention review studies, activity tracker content analyses, activity tracker review studies, meta-analyses of PA intervention studies, smartphone-based intervention studies, qualitative formative studies, app development descriptive studies, qualitative follow-up studies, and other related articles.

For all groups of publications, data were extracted for author, year, target group, and targeted behavior. Depending on the

group, data were further extracted for several additional categories, as follows:

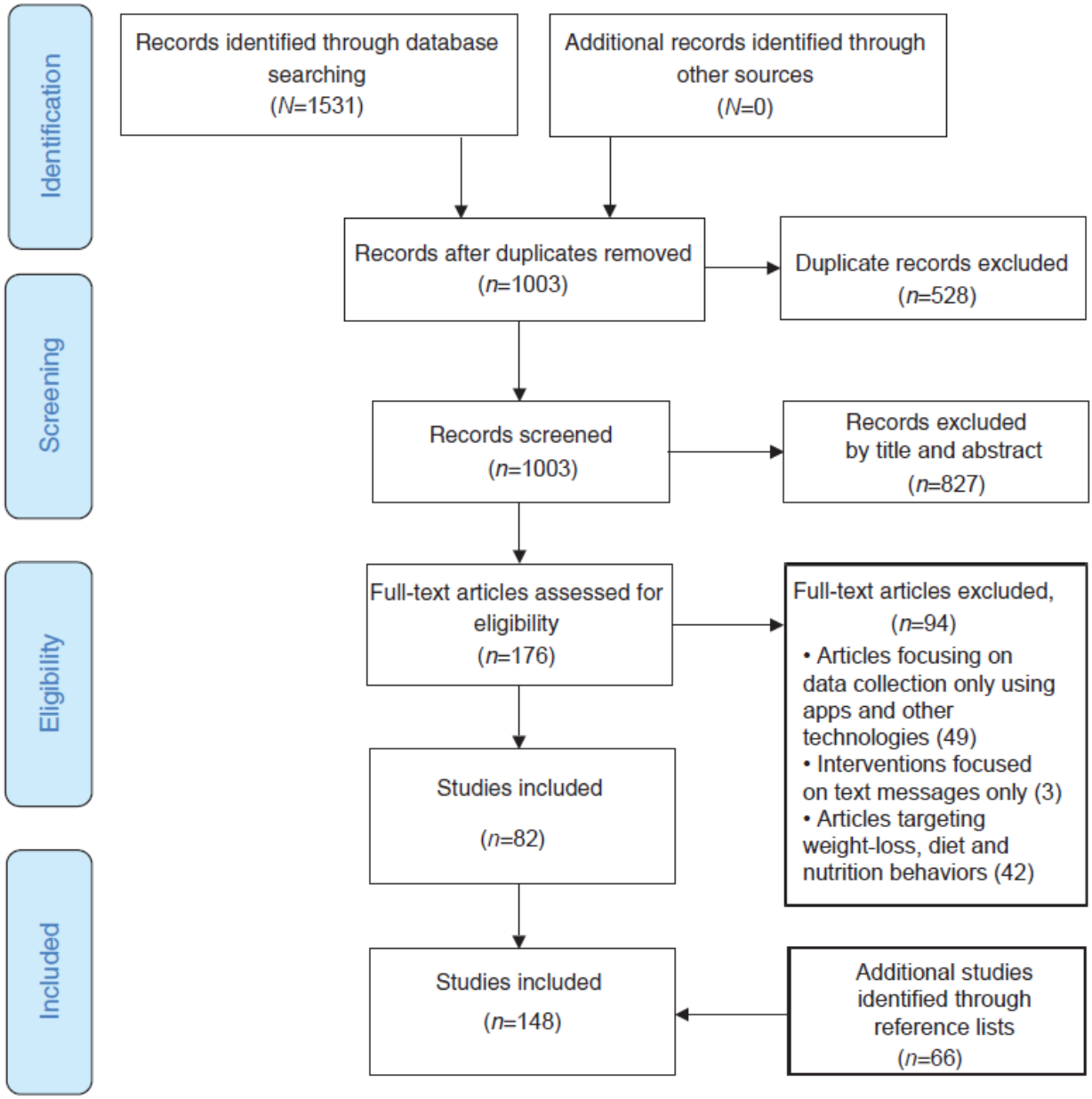
- For commercial smartphone app content analyses, data were extracted for evaluation framework or taxonomy used for coding, number of apps included, app market name and category, and findings related to the theoretical background.
- For smartphone-based intervention review studies, data were extracted for taxonomy used for coding, information on BCTs, identified psychological theories, number of studies included, objective, industry-recognized reporting guidelines.
- For activity tracker content analyses, data were extracted for evaluation criteria or taxonomy used for coding, number of trackers included, number of BCTs included (mean value), BCTs present in all included devices, and BCTs present in none of the included devices.
- For activity tracker reviews, data were extracted for evaluation criteria or taxonomy used for coding, number of studies included, industry-recognized reporting guidelines.
- For meta-analyses, data were extracted for taxonomy used for coding, BCTs associated with more effective interventions, BCTs associated with less effective interventions, and industry-recognized reporting guidelines.
- For smartphone-based intervention studies, data were extracted for pilot, protocol, sample size, theoretical background, study design, study duration, stand-alone or multicomponent intervention, principal outcome measures, industry-recognized reporting guidelines, and PA recommendations.
- For qualitative formative studies and qualitative follow-up studies, data were extracted for sample size, theoretical background, and method of data collection.
- For app development descriptive studies, data were extracted for sample size, theoretical background, commonly reported systematic theory or evidence-based practices, and methods for development, evaluation, and reporting.
- For all other related articles, data were extracted for keyword, title, type of study or methodology, and objective and narratively described further.

Results

Summary of Search Results

A total of 1531 articles were identified during the initial database search. The searches of the MEDLINE and PubMed and ScienceDirect databases yielded 785 and 546 results, respectively. ResearchGate database search results were restricted to 200 because the database search engine generated an unlimited number of search results. After the removal of duplicates, 1003 articles were screened for their titles and abstracts, resulting in 176 full-text articles. Of these, 94 full-text articles were excluded for the following reasons. The resulting 82 articles were hand-searched for references to relevant articles, leading to the identification of an additional 66 articles. As a result, 148 articles were included in the review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Categorization of the Literature Available in the Field

To categorize the included studies, we used the stepwise approach developed by Whittaker et al [182]. Whittaker et al [182] organized the study methods according to the research and evaluation steps in the development of an mHealth intervention. We mapped the identified studies in the same fashion, aligning the development and evaluation trajectory of mHealth PA interventions with the study types used along this trajectory. Although the approach developed by Whittaker et al [182] was the most fitting, it did not accommodate all the identified study types; therefore, a more fine-graded stepwise trajectory was developed. After the literature search was completed, 148 included studies were divided into 10 groups according to the study type (Figure 2), which was in line with the adopted development and evaluation trajectory. If a study could not be allocated to one specific category, it was included in the group *Related Articles* section. After conducting the analyses of the included studies, commonly reported systematic theory or evidence-based practices and methods for development, evaluation, and reporting of mHealth PA interventions were identified (Multimedia Appendix 1). This was the categorization principle used in this review.

To improve further categorization attempts, we refined the outcome of our analysis, which resulted in the table presented below (Table 2). This categorization system may be advantageous for future studies.

Figure 2. Map of search results by number of studies. PA: physical activity.

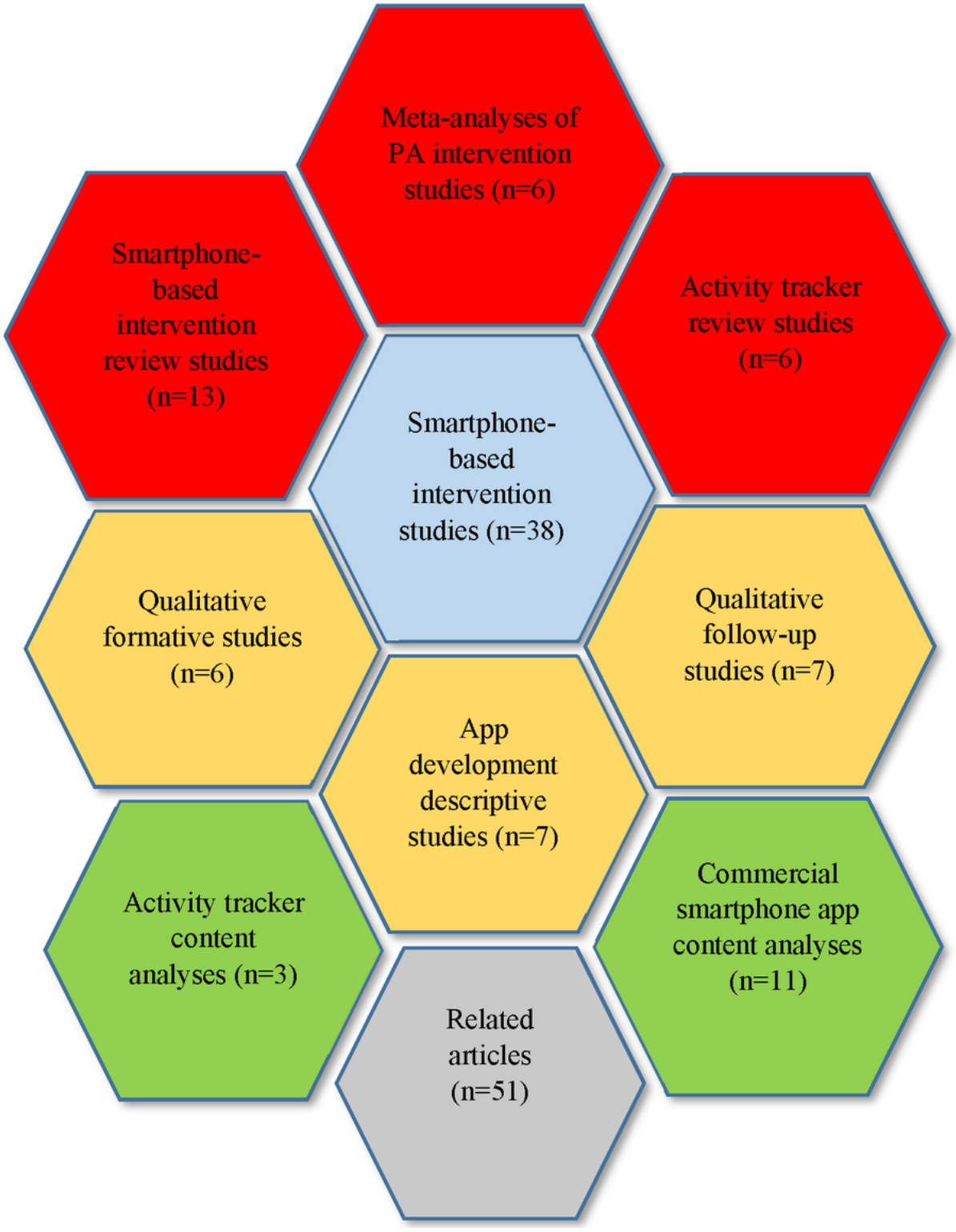


Table 2. Possible categorization approach for smartphone-based interventions for physical activity promotion.

Steps in the development and evaluation process [182] and development and evaluation trajectory of mHealth[a] PA[b] interventions, study type	Common reported systematic theory or evidence-based practices and methods for development, evaluation, and reporting	Purpose
Formative research		
Summarizing findings		
Commercial smartphone app content analyses	• BCT[c] taxonomies • Scales to assess app quality • PA recommendations	• To critically evaluate the material that has already been published • To provide an overview of the current state of knowledge
Smartphone-based intervention review studies	• Industry-recognized reporting guidelines • Behavior change theories or models • BCT taxonomies	
Activity tracker review studies	• Industry-recognized reporting guidelines	
Activity tracker content analyses	• BCT taxonomies	
Synthesizing findings		
Meta-analyses of PA intervention studies	• BCT taxonomies • Industry-recognized reporting guidelines	• To assess the strength of evidence present through establishing statistical significance
Qualitative formative research		
Qualitative formative studies (assessing general topic perception by target users)	• Mixed methods data collection (surveys, questionnaires, interviews, and focus groups discussions)	• To inform the development of the intervention
Pretesting		
Describing an intervention		
App development descriptive studies	• BCT taxonomies • Behavior change theories or models • PA recommendations • Scales to assess app quality	• To describe the intervention development process and intervention features • To control acceptability, engagement, and experiences of proposed intervention to target audience • To improve and refine intervention on the basis of qualitative feedback
Pilot study		
Pilot testing		
Pilot trials	• Behavior change theories or models • PA recommendations • Industry-recognized reporting guidelines • Trial designs (RCTs[d], experimental trials, and rapid design trials)	• To examine content of intervention • To examine feasibility of a trial approach, trial processes (eg, recruitment, registration, data collection), methods
Trial protocol		
Study protocols	• Behavior change theories or models • PA recommendations • Industry-recognized reporting guidelines • Trial designs (RCTs, experimental trials, and rapid design trials)	• To describe processes of trials (eg, recruitment, registration, data collection)

Steps in the development and evaluation process [182] and development and evaluation trajectory of mHealth[a] PA[b] interventions, study type	Common reported systematic theory or evidence-based practices and methods for development, evaluation, and reporting	Purpose
RCTs		
Testing		
Clinical trials	• Behavior change theories or models • PA recommendations • Industry-recognized reporting guidelines • Trial designs (RCTs, experimental trials, and rapid design trials)	• To examine the effect of the intervention as a whole package or the effect of one of its components
Qualitative follow-up		
Qualitative follow-up evaluation		
Qualitative follow-up studies (assessing the developed intervention by target users)	• Mixed methods data collection (surveys, questionnaires, interviews, and focus groups discussions) • Behavior change theories or models	• To control acceptability, engagement, and experiences of proposed intervention to target audience • To control implementation issues • To control the effect of the intervention after dissemination

[a]mHealth: mobile health.

[b]PA: physical activity.

[c]BCT: behavior change technique.

[d]RCT: randomized controlled trial.

Study Characteristics

Overview

The characteristics of the included studies are presented in Tables S1-S10 of Multimedia Appendix 1. All included articles (n=148) were separated according to the subject of the article in the following groups: commercial smartphone app content analyses (n=11), smartphone-based intervention review studies (n=13), activity tracker content analyses (n=3), activity tracker review studies (n=6), meta-analyses of PA intervention studies (n=6), smartphone-based intervention studies (n=38), qualitative formative studies (n=6), app development descriptive studies (n=7), qualitative follow-up studies (n=7), and related articles (n=51). All articles were published between 2008 and 2018. The most common targeted behaviors were PA, SB, and dietary behavior, although the majority of the included articles targeted a single health behavior, namely PA. Although the majority of studies included adult populations (125), 24 articles targeted children and adolescents (age range: 5-19 years).

Commercial Smartphone App Content Analyses

Articles were allocated to this group if the objective of the study was to analyze the content of commercial apps presented on digital distribution platforms (ie, App Store, Google Play, and Microsoft Store). The included studies (n=11) were published from 2012 to 2018, and most of them targeted the general population (n=7) and adults (n=2), whereas only 2 targeted children and adolescents. More than half of the content analyses targeted PA behavior (n=7); the other reported lifestyle-related health behaviors, outcomes and aims were SB, diet, health and fitness, and obesity prevention. Sample sizes ranged between 25 and 3336 (mobile) apps, and the most common digital distribution platform was App Store (n=11). A total of 6 studies used different variations of the BCT taxonomy (26, 40, and 93 BCTs) as an evaluation or coding framework. The average number of the BCTs in those studies ranged from fewer than 4 to 8.1, and the most common BCTs for adults included *provide instruction*, *provide feedback on performance*, *prompt specific goal setting*, *prompt self-monitoring of behavior* (26 BCTs taxonomy [35]), *provide instruction on how to perform the behavior*, *provide feedback on performance*, *goal setting (behavior)*, *prompt self-monitoring of behavior* (40 BCTs taxonomy [36]), *instruction on how to perform the behavior,feedback on behavior*, *goal setting (behavior*), and *self-monitoring of behavior* (93 BCTs taxonomy [37]). Interestingly, this supports the study reporting that the average number of BCTs used in gamified apps aimed at health promotion was higher (14 BCTs) [46] than in nongamified health promotion apps. For children and adolescents, only one study reported the most frequently used BCTs [47]. They were *providing instructions*, *general encouragement*, *contingent rewards*, and *feedback on performance* (26 BCTs taxonomy). The two most recent studies [47,48] used the Mobile App Rating Scale (MARS) to assess the quality of apps. On a 5-point scale, the overall app quality was moderate: the total MARS score ranged from 3.6 to 3.88 points.

Smartphone-Based Intervention Review Studies

This group included intervention studies aimed at reviewing smartphone-based intervention publications. The included reports (n=13) were published between 2013 and 2017 and targeted the general population (n=6), adults (n=4), and children

and adolescents (n=3). More than half of the reviews targeted PA behavior exclusively (n=7), whereas the other reported lifestyle-related health behaviors and outcomes and aims were SB, diet, weight reduction, obesity combatting, healthy nutrition, and overweight prevention. The number of articles included in these reviews ranged from 7 to 52. Only two studies used the taxonomy of BCTs (26 and 93 BCTs) to code the included interventions [49,50]. These studies reported that for adults the following BCTs were most frequently employed: *goal setting (behavior)*, *self-monitoring of behavior*, *social support (unspecified)*, *feedback on behavior*, *instruction on how to perform the behavior*, *adding objects to the environment*, *information about health consequences*, and *prompts or cues* (93 BCTs taxonomy). For adolescents, *prompt self-monitoring of behavior* and *provision of feedback on performance* techniques were most often applied (26 BCTs taxonomy). The other 4 studies provided information about behavioral components without mentioning any taxonomy used for coding [24,29,51,52]. Self-monitoring, cues to action, feedback, and social support were identified as the most commonly used BCTs [29]. The most efficacious and helpful BCTs were reported to be goal setting, self-monitoring, performance feedback, motivational cuing, rewards, social support, and coaching [24,51,52]. The majority of the identified reviews (n=9) reported the theoretical background of smartphone-based interventions. The most frequently used theoretical framework was the Social Cognitive Theory (n=7), followed by the transtheoretical model (n=4), Self-Determination Theory (n=4), and the Theory of Planned Behavior (n=2). The other reported models and theoretical approaches included the Persuasive Systems Design Model, the Control Systems Theory of Self-regulation, the Behavior Change Wheel, the Five A's Model, the Fogg Behavior Model, Learning Theory or operant conditioning, Social Influence Theory, the Theory of Reasoned Action, and Cognitive Behavior Therapy. Of 13 reviews, 3 used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines [24,25,50].

Activity Tracker Content Analyses

Articles were included in this group if the objective of the content analysis was to analyze the theoretical components included in the activity trackers. The included studies (n=3) were published between 2014 and 2017 and targeted the general population [53-55]. The majority of the content analyses targeted PA behavior exclusively (n=2), whereas the other reported lifestyle-related health behaviors were SB and sleep. The number of included activity trackers per article ranged from 3 to 13. All 3 studies used the taxonomy of BCTs (40 and 93 BCTs) to code the included interventions [53-55]. According to these content analyses, the average number of BCTs included in the activity monitors ranged between 9 and 25 BCTs (40 BCTs taxonomy). There was an agreement between 2 studies about BCTs present in all included devices, which were *provide information about others' approval*, *provide normative information about others' behavior*, *prompt review of behavioral goal*, *provide rewards contingent on successful behavior*, *prompt self-monitoring of behavior*, *prompting focus on past success*, *provide feedback on performance*, *facilitate social comparison*, and *plan social support or social change* (40 BCTs taxonomy) [53,54]. According to the same studies, *prompt anticipated regret*, *fear arousal*, *prompt self-talk*, *prompt use of imagery*, and *general communication skills training* BCTs were not present in any of the included devices (40 BCTs taxonomy).

Activity Tracker Review Studies

Review studies in this group aimed to provide evidence on the effectiveness, efficacy, feasibility, validity, or reliability of activity trackers. The included studies (n=6) were published between 2012 and 2018 and targeted adults (n=5) and children and adolescents (n=1) [56-61]. The majority of the reviews targeted PA behavior exclusively (n=5), whereas the other reported lifestyle-related health behavior was sleep. The number of articles included per review ranged between 5 and 134 publications. Five studies used PRISMA or PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) reporting guidelines [56-60].

Meta-analyses of PA Intervention Studies

Articles were included in this group if the objective of the meta-analysis was to analyze PA intervention studies and to define the BCTs that were associated with more or less effective interventions. It is important to note that all identified meta-analyses (n=6) reviewed only *classic* interventions and did not include smartphone-based interventions [62-67]. Only one meta-analysis reviewed PA smartphone-based interventions [50]. However, this meta-analysis was excluded because it did not provide information on the effectiveness of BCTs because of the small number of included studies.

Articles included in this group were published between 2009 and 2017 and mainly targeted adults with one exception, where the targeted group included children and adolescents [65]. All the included meta-analyses targeted PA, with 3 studies additionally targeting healthy eating (HE) and diet [62,65,67]. Two studies used 26 BCTs taxonomy for coding, 3 studies used 40 BCTs taxonomy for coding, and 1 study used the latest 93 BCTs taxonomy for coding. Reported results for BCTs associated with interventions that are more effective were divergent, with *self-monitoring* and *feedback* reported to be effective according to 4 and 3 meta-analyses, respectively. Every meta-analysis reported different results for BCTs associated with less effective interventions. For the adolescent target group, BCTs (26 BCTs taxonomy) associated with more effective interventions include *provide information on consequences*, *provide information about others' approval*, *prompt intention formation*, *prompt self-monitoring of behavior*, and *agree on behavioral contract* [65]. The *provide instruction* BCT was associated with less effective interventions in adolescents [65]. Only the latest meta-analysis used PRISMA reporting guidelines [67].

Smartphone-Based Intervention Studies (Study Protocols, Pilot Trials, and Clinical Trials)

The smartphone-based intervention study group included 38 articles representing 32 research studies published between 2008 and 2018. The majority of these studies targeted adults (n=20), whereas 12 targeted adolescents, and the sample size ranged from 8 to 700 participants, and the duration of interventions ranged from 2-32 weeks (most common duration: 8 weeks).

The participants' ages ranged from 8-81 years. A total of 14 studies exclusively targeted PA behavior; the other reported lifestyle-related health behaviors, aims, concepts, outcomes, and conditions included weight loss, SB, cardiorespiratory fitness, diet, sleep, fitness, and obesity. The most common study design was a two-arm randomized controlled trial (RCT; n=10); for other study designs, the number of intervention groups ranged between 1 and 4. There was a preponderance in the number of multicomponent interventions (n=19) over stand-alone interventions (n=13). The interventions mainly used newly designed smartphone apps (n=29) rather than commercially available apps (n=3), the theoretical background of which was unknown. The most common outcome measures were minutes spent with moderate-to-vigorous PA (MVPA) and a daily step count. In total, 14 studies did not report a theoretical background. For adults, the most frequently used theoretical framework was Social Cognitive Theory (n=11), followed by Self-Regulatory Theory (n=3) and the Fogg Behavior Model (n=2). Of the 12 studies including adolescents, several (n=4) did not report any theoretical background, and among those who did, Self-Determination Theory (n=6) was the most frequently used. The other reported theoretical frameworks and models include the Theory of Meaning Behavior, the Five Factor Model of Personality, the Health Belief Model, the Technology Acceptance Model, the Theory of Motivation in Videogames, the Transtheoretical Model of Health Behavior Change, the Functional Triad, the Transcontextual Model of Motivation, the Synergy Hypothesis, Learning Theory, Basic Psychological Needs Theory, the COM-B (Capability, Opportunity, Motivation, Behaviour) model, and the Behavior Change Wheel. A total of 10 studies used CONSORT (Consolidated Standards of Reporting Trials) reporting guidelines, 1 study used SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) reporting guidelines, and 2 studies used both CONSORT and SPIRIT reporting guidelines.

Qualitative Formative Studies

Articles in this group used a qualitative approach to examine users' views of and preferences for app features in terms of usability and attractiveness, among others, that can inform the development of future mHealth PA interventions. The identified studies (n=6) were published between 2011 and 2016 and included adults (n=5) and adolescents (n=1) target populations [68-73]. Sample sizes ranged between 14 and 120 participants, and research designs included focus groups (n=3), web-based surveys (n=3), and individual interviews (n=3). The studies focused on the perception of apps targeting PA (n=3), health behavior change in general (n=2), and health and fitness (n=1). The app features that were evaluated by participants in the majority of these studies (n=4) were social networking, context sensing or personalization, design, self-monitoring, and goal setting.

Social networking, that is, exposing one's health behavior through integration of the PA app in social networks (eg, Facebook), was generally perceived negatively. Context sensing or personalization, self-monitoring, and goal setting were perceived as valued features in smartphone apps. The design of the app appeared to be a crucial feature, in that users preferred a simple and structured layout, which was easy to use, playful, and fun. Apps were not used or uninstalled if they contained unnecessary features, required excessive data entry for sign up, had complicated operating procedures, and required instructions that were time-consuming or burdensome.

App Development Descriptive Studies

Articles were included in this group if the objective of the study was to describe the intervention development process and intervention features. The identified studies (n=7) were published between 2012 and 2018 and included adults (n=5), adolescents (n=1), and general populations (n=1) [74-80]. The intervention groups differed in terms of sample size, ranging from 10 to 68 participants. The developed interventions mostly targeted PA (n=6), and other related behaviors, such as SB (n=2) and weight loss (n=1). Common reported systematic theory or evidence-based practices and methods for development, evaluation, and reporting included PA recommendations (n=1) [76], BCT taxonomies (n=2) [78,80], and MARS (n=1) [79]. The most frequently used theoretical framework was Social Cognitive Theory (n=5); the other reported theoretical frameworks and models included Health Belief Model, Theory of Planned Behavior, Technology Acceptance Model, Fogg Behavior Model and Self-Determination Theory.

Qualitative Follow-up Studies

This group of studies aimed at assessing the acceptability, engagement, and experiences of the target audience with the intervention and the effect of the intervention after dissemination. The identified studies (n=7) were published between 2012 and 2017 and included adults (n=4) and adolescents (n=3) as target populations [81-87]. Sample sizes ranged between 5 and 68 participants and research designs included surveys (n=3), interviews (n=2), focus groups (n=1), and questionnaires (n=1). The apps included in the studies targeted PA (n=5), fitness (n=1), and well-being (n=1). The theoretical frameworks and models were reported only in 2 studies, which included the Theory of Planned Behavior in both studies (n=2) [84,87]. The other reported theoretical frameworks and models included the Theory of Meaning Behavior, the 5 Factor Model of Personality, and the Functional Triad [84,87].

Related Articles

The related articles group included 51 articles published between 2008 and 2018. These articles were mainly identified through manual reference searches, and although they were relevant to the topic of this review, they did not fit into the other groups presented above. The study types included methodological, theoretical, conceptual studies; reports; recommendations from workshops; other literature reviews (reviews of methodological, theoretical, and conceptual studies); and reviews and trials on related topics (eg, gamification) that represented theoretical and methodological findings and recommendations that were grouped into several topical subgroups: activity tracking, automation, BCT, behavior change theory, GPS, just-in-time adaptive interventions, mHealth apps, PA, profiling, and RCT alternatives for mHealth. Relevant information from these articles was analyzed and presented narratively in the *Discussion* section.

Discussion

Theories and Techniques Implemented in Smartphone-Based PA Interventions to Support Behavior Changes: Current Situation and Recommendations

The science of behavior change has advanced significantly in recent years. Nevertheless, many challenges remain concerning the standardization of the development and reporting of methods of behavior change interventions. As presented in the tables of Multimedia Appendix 1, there is a plethora of approaches in developing smartphone-based PA interventions; however, most of them have been developed and reported without an explicit theoretical foundation. This has been described as the development of *theory-inspired* interventions (in which the theoretical background is often chosen depending on the experiences and preferences of researchers and developers), rather than *theory-based* interventions (in which the chosen theoretical background was measured and tested in the intervention or conditions) [88].

To accomplish a more standardized methodological approach, several frameworks have been developed by Michie et al [88] in the domains of smoking, PA and HE, alcohol consumption, and safer sex. For the PA and HE domains, these authors developed the *Behavior Change Technique Taxonomy*, in which a BCT is defined as "an observable and replicable component designed to change behavior" [37]. It is the smallest component compatible with retaining the postulated active ingredients and can be used alone or in combination with other BCTs [89]. The taxonomy itself has been described as "an extensive, integrated, hierarchical classification system for reliably specifying intervention components (BCTs)" [88]. There are three versions of the BCT taxonomy: *A Taxonomy of Behavior Change Techniques* (26 BCTs), developed for coding PA and HE interventions for adults in 2008; *The CALO-RE (Coventry, Aberdeen & London–Refined) taxonomy* (40 BCTs), developed in 2011, which is the extended version of the previous taxonomy; and finally *The Behavior Change Technique Taxonomy (v1)*, which is the latest (developed in 2013) 93 BCTs cross-domain taxonomy, and which is recommended to be used instead of the previous versions, which are considered as "domain specific proto-versions" [35-37,89].

In addition, Michie et al [88,183] created a compendium of 83 theories of behavior and behavior change, containing more than 1700 theoretical constructs, some of which can be potentially considered as so called "theoretical mechanisms of action." In this context, mechanisms of action are conceptualized as "a range of theoretical constructs, defined broadly as the processes through which a behavior change technique affects behavior" [88].

To overcome the unsystematic intervention development and reporting, it is also important to understand how BCTs can be linked to theoretical mechanisms of action, which is currently being investigated [88,90]. Such a link will provide a basis for a systematic and transparent method for developing behavior change interventions. Until then, the *Behavior Change Wheel* was considered the most appropriate development framework for selecting appropriate BCTs for specific behavior change interventions. This framework was also developed by Michie et al [91,184], introducing a synthesis of 19 behavior change frameworks, providing a systematic guide for designing and evaluating behavior change interventions and policies.

Several important tendencies were identified in all the included groups of studies. First, studies aiming to promote PA via smartphone-based interventions in adolescents are underrepresented in comparison with those targeting adults. While analyzing the studies including adolescents, Schoeppe et al [24] confirmed that there was no difference in the BCTs incorporated in apps for adolescents compared with those used in apps for adults. This is surprising, as adolescents' motivations, social environment, and financial opportunities, among others, are much different from those of adults [185].

Second, the tables in Multimedia Appendix 1 demonstrate that there is no agreed evaluation framework or taxonomy to code or report smartphone-based PA interventions. Researchers did not state the coding methods [24,29,51,52], used various evaluation frameworks [92-94], or used different versions of the BCT taxonomy by Michie et al [66,89], who developed all versions of the BCT taxonomy, recommend using the latest version, which consists of 93 BCTs, and although several authors justified their preference for the specific version of the taxonomy (eg, O'Brien et al [66] stated that the 40 BCTs CALO-RE taxonomy was used "as it was specifically developed for use with PA and dietary interventions"), it is evident that such an approach is disadvantageous because it hinders the systematic accumulation of evidence. However, this is not surprising, as the field of mHealth is still a fairly young field of research, where new, dynamic theories and models of behavior that better fit the capabilities of mobile systems have yet to be developed or are currently under development [186]. These new developments in behavioral models and mechanisms of action must be taken into account for the field to progress. Thus, while striving for uniformity in reporting, researchers should periodically *upgrade* their reporting methods while maintaining a balance between systematic and innovative approaches. It is therefore important to realize that as the field grows, the taxonomy will be extended and modified, and it will be subject to further refinement and development, as stated by Michie et al [37].

Third, the tables in Multimedia Appendix 1 demonstrate that there seems to be no consensus about the optimal behavior change theory or model that should be used in smartphone-based interventions for PA promotion. Until now, there has been no clear evidence for the best behavior change model; however, the results show that Social Cognitive Theory seems to be the most favored among researchers. Progress in this field of research will be hampered if theoretical models on which interventions are based are not selected according to explicit criteria but on personal preferences. Although some researchers continue in their work to eventually provide systematic solutions, the most coherent approach at this time seems to consist of selecting BCTs based on the features and goals of the designed intervention using the Behavior Change Wheel framework [88,184].

Reviews of the commercial app market (Table S1 in Multimedia Appendix 1) also suggest that there is a lack of theory-based and evidence-based apps [92,95-98]. One should differentiate those terms; although the first usually refers to the systematic selection of BCTs and theories or models, the second refers to the compliance of apps with various national and international PA norms.

As previously outlined, the included meta-analyses did not analyze smartphone-based interventions. We decided to include them here, based on the rationale of Brannon and Cushing [65], who state that *classic* and smartphone-based interventions do not represent disparate bodies of evidence. After comparing BCTs used in commercial, research-driven apps, activity trackers, and meta-analyses (corresponding tables in the Multimedia Appendix 1), it becomes clear that there is no agreed theoretical or evidence base for the choice of BCTs in smartphone-based interventions. However, it is still important to report on and verify accumulated evidence for the field to progress, while considering its inconclusive nature. For adolescents, a comparison of the review by Schoeppe [24] and Brannon and Cushing meta-analysis [65] shows that BCTs associated with effective interventions *provide information about others' approval* and *prompt self-monitoring of behavior*.

There is a clear need to conduct meta-analyses on mHealth studies. Until now, such a meta-analysis has been conducted once by Direito et al [50]; however, the author concentrated more on RCTs of mHealth technologies, rather than smartphone-based mHealth interventions tested with more suitable trial designs. As a result, only RCTs were selected for this meta-analysis, despite the latest considerations that RCTs may not provide the most advantageous design for the evaluation of mHealth interventions [99]. Second, out of the 21 studies included, only 5 described smartphone-based interventions, whereas the rest included interventions delivered through a website, SMS text messages, and PDA devices, that is, modes of delivery that are often considered outdated in the mHealth domain.

In general, it is also important to consider the mechanisms of action and the parameters of effectiveness of coded BCTs. Although the current approach applied for coding, using the taxonomy of Michie et al [37], does not consider the context of BCTs, Kok et al [100] argued that this is crucial. They state that the taxonomy developed by Michie et al [37] is useful for coding, but is not a good basis for intervention development, as it may contain ineffective and even countereffective methods (BCTs) [100]. Kok et al [100] define the *parameters of effectiveness* as "the conditions that must be satisfied in practical applications for the method (BCT) to be effective" and add that if parameters of effectiveness for the particular method (BCT) are violated, it may become less effective or even countereffective. Consequently, an alternative, that is, *A Taxonomy of Behavior Change Methods*, has been designed to take parameters of effectiveness into consideration, while developing an intervention [100]. Various researchers, including Michie et al [88], support the idea that BCTs should not be treated in a vacuum, considering their context and possible combinations [101-103].

When selecting a theoretical model, many researchers seem to assume that the basic motivation of the user is to become more physically active, which is not always the case [104]. Therefore, which models consider the level of motivation before favoring a specific framework is worth assessing [104]. Interestingly, the present analysis of intervention studies shows that only interventions targeting adolescents used Self-Determination Theory, whereas for adults, Social Cognitive Theory was by far the most frequently used model. Finally, researchers have recently started to question whether the theoretical models developed before the invention of smartphones, and digitalization in general, are still applicable [51,105,106,186]. Such critiques are justifiable, as digital devices such as smartphones provide unprecedented opportunities for observation, data collection, and just-in-time interventions and, therefore, the interaction between the user and the device delivering the intervention [26].

Other Commonly Reported Systematic Theory or Evidence-Based Practices and Methods for the Development, Evaluation, and Reporting of Smartphone-Based PA Interventions: Current Situation and Recommendations

Overview

As is evident from the results in this review and the tables presented in Multimedia Appendix 1, commonly reported systematic practices and methods could be successfully identified. They include PA recommendations, trial designs (RCT trials, experimental trials, and rapid design trials), mixed methods data collection (surveys, questionnaires, interviews, and focus group discussions), scales to assess app quality, and industry-recognized reporting guidelines. Nevertheless, there seems to be no consensus on which practices and methods are preferable to use, which reflects the same tendency as outlined for theories and techniques. To advance this field of research, researchers and developers should consider using existing practices and methods depending on the aims and features of the developed intervention. The more systematic the development process, the higher the replicability of the results. As a result of the current review, we provide a list of best practices and methods that can be used during the development evaluation and reporting of PA mHealth interventions.

WHO Global Recommendations on PA for Health

These are evidence-based recommendations of the WHO that "address the links between the frequency, duration, intensity, type and total amount of PA needed for the prevention of NCDs" [15]. Alternatively, researchers can use other public PA guidelines used by national agencies and health institutions, such as Canadian Physical Activity Guidelines for Adults, Physical Activity Guidelines for Americans, American College of Sports Medicine Guidelines, Center for Disease Control Guidelines, American Heart Association Guidelines, UK Department of Health Guidelines, Institute of Medicine Guidelines, and US Department of Health and Human Services Guidelines [97,98,107]. Applying one or several of these guidelines will help researchers to understand PA norms that,

for instance, can be used as a PA goal for participants or inclusion and exclusion criteria.

MARS and User Version of the MARS

The MARS scale has been developed quite recently in many of the most recent mHealth research studies [47,79,108]. This is a "reliable, multidimensional measure for trialing, classifying, and rating the quality of mobile health apps" [109]. The MARS scale is useful if the researcher wishes to reliably rate or see the possible flaws of the developed mobile app.

Industry-Recognized Reporting Guidelines

The following industry-recognized reporting guidelines have been illustrated:

- PRISMA: this is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses [187].
- PRISMA-P: this is a set of items aimed at facilitating the development and reporting of systematic review protocols [188].
- CONSORT statement: this internationally acknowledged tool can be used to assess the quality of RCT studies and to design or report an RCT of the highest quality and standard [189].
- SPIRIT: a guideline for minimum content of a clinical trial protocol [190].

Rapid Design Trials

Although RCT study designs are widely considered a *gold standard* for intervention research in many areas, it has been suggested that they may not be the best approach for the evaluation of mHealth interventions for several reasons. First, the duration of the completion of an RCT is long (5.5 years on average from recruitment to publication of the trial results [99,110]), which in the modern ever-developing digital world may be the cause for an app becoming obsolete. Second, an RCT is a rigid design requiring interventions to remain unchanged and stable during the entire duration of the trial. This creates a problem, as software is meant to change, progress, evolve, and adapt to its user in short periods [32,51,99,108,110-114]. Therefore, mHealth interventions could make use of flexible evaluation designs and methodologies, providing timely information and being responsive and agile. Consequently, alternative designs and methodologies for the evaluation of mHealth interventions have been proposed [32,99,108,110-113]:

1. Continuous Evaluation of Evolving Behavioral Intervention Technologies
2. Sequential Multiple Assignment Randomized Trial
3. The Multiphase Optimization Strategy
4. Microrandomized trial (MRT)
5. Step-wedge design (ie, cluster randomized design)
6. n-of-1 trials
7. Practice-Based-Evidence methodology
8. Trial of Intervention Principles framework
9. Collaborative Adaptive Interactive Technology framework

However, these designs have rarely been implemented. According to the most recent review of PA apps, only 2 of 111 included studies used rapid research designs [32]. The methodology of the most recent rapid design, MRT, is currently being developed, and the first protocols and trials have been recently published [115,191].

Qualitative Studies

Although there is no one recommended methodology, the most commonly reported methods in identified studies include surveys, questionnaires, interviews, and focus group discussions. On the basis of this review, we cannot recommend any specific method, yet there is a clear need for more systematic reporting of results. Nevertheless, the studies summarized in Tables S7-S9 in Multimedia Appendix 1 provide some indication of the most efficacious and user-attractive features in mobile apps aimed at PA promotion.

1. Design simplicity: Ease of use and navigation through the app, absence of unnecessary features, unambiguous information, and a structured layout were all listed as features that positively affected participants' engagement. Apps with excessive data entry for sign up, presenting features that required instructions, and complicated operating procedures were negatively perceived by users [28,29,69-71,81,82,116].
2. Personal approach for each user- tailored coaching, goals, feedback, and notifications: Users perceive a personalized approach as an important factor for motivation and engagement. Therefore, it is important to consider sociodemographic user differences [117,118]. Moreover, the users themselves prefer to be in control of the app's features, having the ability to hide or add them [28,69,71,72,81,82,116].
3. Reward: A transparent reward system was positively recognized by users [28,71,82,116].
4. Self-monitoring and goal setting: These app features were the key features enjoyed or rated positively by app users [68,70,72,82].
5. Gamification: This feature can positively affect user engagement by bringing more enjoyment to exercise or activity [119,120].
6. Social networking: This feature was perceived differently in various apps:
   - *Peer-to-peer influence* was delivered through encouragement, praise, and competition with the participants' peers. As indicated by Klasnja and Pratt [121], the results presented by different researchers are inconclusive: studies report both positive and null effects [120-125].
   - *Social support from family and friends*: As the reviews show, the effect on participants depends on the behavioral goals of friends or family members: if the goals differ, the effect of social support seems to be low [121].
   - *Social modeling* (eg, tips for health-related resources from successful peers) seems to have a positive effect on participants [121].
   - *Integration with social networks* (eg, Facebook) was perceived negatively by app users [69,71,81,82].

These findings demonstrate that a chosen method of social support can significantly affect the acceptability and usefulness

of the app among users. Overall, it is important to underline the necessity of pretesting the app with a specific target audience to optimally refine the app's features and components.

Devices and Primary Outcomes Used for Data Collection and Analysis in Smartphone-Based PA Interventions: Current Situation and Recommendations

Smartphone-based interventions can be divided into stand-alone interventions, where only the app is used and multicomponent interventions, where the app is one of several intervention components. The choice of intervention components affects the intervention outcomes, and, if a multicomponent approach is chosen, may lead to the inclusion of various devices as additional components of the intervention.

For the majority of researchers, the selection of smartphone-based intervention components depends on several factors, such as the accuracy of data collection, device compatibility with the user, and durability. As can be seen from Table S6 in Multimedia Appendix 1, all smartphone-based intervention data collection components can be divided into three groups: smartphones, commercial activity trackers, and medical-grade activity trackers. The selection of a data collection device is usually well aligned with the chosen outcome measures.

As presented in Table S6 of Multimedia Appendix 1, a stand-alone intervention that includes only a smartphone with the installed app and inbuilt accelerometer can track the most common PA outcome measures, that is, minutes spent in MVPA and SB, and daily step count [126]. These data can also be collected with a range of precision levels (depending, for instance, on the use of built-in GPS sensors, which can provide data that are more accurate) [127-129]. The drawbacks of solely using smartphones include the short battery life of the device, only moderate accuracy levels, moderate durability, and limited exposure time (the user will not usually carry the phone during certain periods of the day) [26,121].

The validation reviews presented in Table S4 of Multimedia Appendix 1 demonstrate that commercially available, usually wrist-worn activity trackers can help collect similar data with higher accuracy levels, although still in a moderate range [56,60]. In addition, some built-in sensors (eg, heart rate [HR]) can provide supplementary data and improve the accuracy of MVPA measures. They avoid most of the drawbacks of smartphone devices, as they provide a long battery life for the device, high device durability, and extended exposure time. Commercial activity trackers show good potential in the implementation of theory-based practices and improve the data collection procedure in human physiology research for both adults and adolescents [58,120,130-132,192].

Medical-grade activity trackers (hip, waist, or wrist worn), for example, ActiGraph devices, provide the highest measurement accuracy levels; however, they also have certain drawbacks. The hip and waist location can lead to low user compatibility levels and reduced exposure, whereas HR can only be measured with a wireless HR monitor [193].

Consequently, while developing PA interventions, researchers should consider these factors and choose the device according to the characteristics most suitable for their projects. It is important to note the findings of a recent review, which confirms that multicomponent interventions tend to be associated with higher intervention efficacy [24]. Although some researchers chose to use simplistic outcome measures such as a daily step count, studies show that a multidimensional approach with several outcome measures is more comprehensive [107].

Advancing mHealth Further: Technological Advances Applied in Smartphone-Based mHealth Interventions

Researchers working in the smartphone-based mHealth field often face problems with participants' engagement: the long-term retention levels are usually quite low at 18 months follow-up measurements [133]. One way to solve this problem is to make the intervention more attractive to the participants by personalizing it. Personalized smartphone-based PA mHealth interventions may be more effective and preferred by participants over interventions with a generic program and advices or notifications [28,51]. Various researchers have suggested that personalization or tailoring of PA interventions will positively affect participants' perception and engagement [59,134-136]. Some studies have attempted to personalize the intervention components manually and using automated approaches [137,138]. Manual automation (where the researcher inputs a large amount of collected data individually into every participant's profile) has shown positive trends. Nevertheless, depending on the number of variables, the number of entries required for each participant, and the number of participants, this approach might be too time-consuming to become impractical [139,140]. More automated approaches, specifically machine learning or data mining, require minimum assistance during the utilization period, and are therefore promising for solving big data challenges, including behavior change interventions [113,194]. Rabbi et al [138,195] have already successfully implemented machine-learning solutions in various smartphone-based mHealth interventions, demonstrating their potential. However, machine learning science is in its early stage of development, and some questions still need to be answered and tested. One of them concerns the level of automation: Which one should the researcher choose for his or her particular intervention? Although full manual tracking is considered outdated and has a high data collection burden, fully automated tracking requires high data collection accuracy and may lower participants' self-awareness; therefore, semiautomated tracking is currently the best solution [141,142].

As the articles listed in Table S10 of Multimedia Appendix 1 show, numerous technological advances are increasingly being used in smartphone-based PA mHealth interventions. One example concerns HR monitoring. Previously, HR monitoring PA interventions used separate devices. Currently, commercial activity trackers include built-in HR sensors, which can increase participants' acceptance of the intervention. Another example concerns the use of a smartphone's inbuilt GPS sensor, which can provide high accuracy of movement speed and location, among others; however, it is highly energy consuming and drains smartphone batteries very fast. The latest power management algorithms help to reduce the resource demands

of continuous sensing, which ensures longer usability time, providing researchers with additional data collection opportunities [143].

Implications for Future Research

On the basis of this review and in light of the widely used international reporting guidelines, several recommendations for future research can be inferred:

- Support uniformity of reporting by describing interventions and procedures in an adequate and consistent manner, using industry-recognized reporting guidelines, such as PRISMA, CONSORT, and SPIRIT [196].
- Develop and code interventions in a more systematic way, using recommended practices while taking into account new models that offer additional opportunities in behavior research [186]. Currently, the systematic approach is either not applied or various frameworks are being used (eg, different versions of taxonomy by Michie), which slows or even prevents knowledge transfer and evidence accumulation. After the first results will be yielded in the development of a methodology for linking BCTs to theoretical mechanisms and the Human Behavior-Change Project, more systematic solutions will become available [88,194].
- Meta-analyses, including modern mHealth solutions (eg, smartphones) and excluding outdated devices or methods (intervention based solely on SMS, PDAs, etc), provided there is a sufficient number of studies meeting the inclusion criteria.
- Profit from interdisciplinary collaboration while developing mHealth interventions. Various researchers and research groups working on the development of PA mHealth interventions have underlined the positive effect of collaboration between related stakeholders and experts in the domains of behavior change, software development, machine learning or data science, physiology, and public health [29,65,70,94,102,106,113,135,144]. A recent systematic review demonstrated that the collaboration of experts from various research domains greatly enhances the quality of the produced publications and research work in general [145].
- Perform more studies designed for adolescents, accounting for differences in levels of motivation and lifestyle compared with adults.
- Implement rapid study designs while evaluating the intervention (eg, MRT, Multiphase Optimization Strategy, Sequential Multiple Assignment Randomized Trial, etc) [32].
- Implement wearable activity monitors with built-in sensors (eg, HR and GPS) will provide more opportunities for data collection. Both commercial and research-grade trackers are advantageous. However, the collaboration of two domains, for instance ActiGraph and Garmin, is yet to bring fruitful results [197].
- Implement the latest findings of machine learning or data mining and artificial intelligence domains into behavior change interventions [88,138,194].
- Improve engagement with smartphone-based mHealth interventions by testing and implementing meaningful gamification and social networking features [120].
- Build the reward and engagement engine of the app in a way that users will become autonomously physically active over time and do not depend on an app, a tracker, or an intervention in perpetuity.

Strengths and Limitations

The strength of this scoping review is the comprehensive search strategy, which allows the majority of published related articles to be included. Therefore, the scope of the review is wider than the scope of systematic reviews on smartphone-based mHealth interventions for PA promotion. However, a scoping review does not consider the methodological quality assessment of the included studies. Consequently, several studies had moderate methodological quality, which calls for their findings into question. It is important to emphasize that the included interventions developed and evaluated apps and activity trackers that provide sensor-based feedback on PA. Smartphone-based interventions related to chronic diseases other than cardiovascular diseases and obesity (eg, diabetes mellitus), preventive health issues (eg, alcohol abuse, smoking, and sports injuries), weight loss, diet, and nutrition were not included in this review. Finally, yet most importantly, only smartphone-based mHealth interventions were included in this review.

Conclusions

Smartphone-based mHealth interventions aimed at PA promotion in adolescents and adults show promising results for effective behavior change. Although there is a plethora of published studies with adults, the number of studies and, consequently, the evidence base for adolescents is very limited. In the past few years, a growing number of researchers have developed multicomponent mHealth interventions that, in addition to the app, include commercial or research-grade activity trackers, which can provide additional insight into a participant's lifestyle. Overall, the efficacy of smartphone-based mHealth PA interventions can be considerably improved through a more systematic approach to developing, reporting, and coding of the interventions. Specifically, researchers should aim to develop theory-based rather than theory-inspired interventions, which is currently challenging, as there is no consensus on development, evaluation, or coding practice. Finally, the current stage of behavior science advocates an interdisciplinary approach to the development of behavior change interventions, including innovative approaches such as machine learning and data mining.

Acknowledgments

The first author is supported by an Industrial Fellowship from the Luxembourg National Research Fund (reference number 12674722) and by Actimage Luxembourg SA. The authors are grateful for the support of the graduate psychology student Claudia Manuela Vila Verde Gonçalves for reviewing the literature.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Articles included in the scoping review.
[DOCX File , 235 KB - mhealth_v9i7e24308_app1.docx]

References

1. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
2. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
3. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
4. WHO Director-general's Opening Remarks at the Media Briefing on Covid-19. World Health Organization. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-05-26]
5. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE). Johns Hopkins University. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-05-26]
6. Coronavirus: Half of Humanity Now on Lockdown as 90 Countries Call for Confinement. EuroNews. URL: https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou [accessed 2020-05-26]
7. Mattioli AV, Ballerini Puviani M, Nasi M, Farinetti A. COVID-19 pandemic: the effects of quarantine on cardiovascular risk. Eur J Clin Nutr 2020 Jun;74(6):852-855 [FREE Full text] [doi: 10.1038/s41430-020-0646-z] [Medline: 32371988]
8. Goethals L, Barth N, Guyot J, Hupin D, Celarier T, Bongue B. Impact of home quarantine on physical activity among older adults living at home during the covid-19 pandemic: qualitative interview study. JMIR Aging 2020 May 7;3(1):e19007 [FREE Full text] [doi: 10.2196/19007] [Medline: 32356777]
9. SURVEY REPORT:Americans' Perceptions of Telehealth in the Era of COVID-19. SYKES. URL: https://www.sykes.com/reports/2020-telehealth-survey/ [accessed 2020-05-26]
10. Telehealth Market in US - Industry Outlook and Forecast 2020-2025. Research and Markets. URL: https://tinyurl.com/2shxvcxm [accessed 2020-05-26]
11. Mhealth Solutions for Managing the Covid-19 Outbreak. Andalusian Agency for Healthcare Quality. URL: https://mhealth-hub.org/mhealth-solutions-against-covid-19 [accessed 2020-05-26]
12. Coronavirus Gives People a Reason to Use Telehealth, But Doubts Remain. Xtelligent Healthcare Media. URL: https://mhealthintelligence.com/news/coronavirus-gives-people-a-reason-to-use-telehealth-but-doubts-remain [accessed 2020-05-26]
13. A National Survey of Children and Young People's Physical Activity and Dietary Behaviours in New Zealand: 2008/09. Clinical Trials Research Unit The University of Auckland. 2010. URL: http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/media-use-attitudes-14/Childrens_2014_Report.pdf [accessed 2020-05-26]
14. Farooq MA, Parkinson KN, Adamson AJ, Pearce MS, Reilly JK, Hughes AR, et al. Timing of the decline in physical activity in childhood and adolescence: gateshead millennium cohort study. Br J Sports Med 2018 Aug;52(15):1002-1006 [FREE Full text] [doi: 10.1136/bjsports-2016-096933] [Medline: 28288966]
15. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010.
16. Health at a Glance: Europe 2016 – State of Health in the EU Cycle. Paris, France: OECD Publishing; 2016.
17. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273. [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]
18. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195 [FREE Full text] [doi: 10.1159/000222244] [Medline: 20054224]

19. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: a systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
20. 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]
21. Kwak L, Kremers SP, Bergman P, Ruiz JR, Rizzo NS, Sjöström M. Associations between physical activity, fitness, and academic achievement. J Pediatr 2009 Dec;155(6):914-8.e1. [doi: 10.1016/j.jpeds.2009.06.019] [Medline: 19643438]
22. Direito A, Jiang Y, Whittaker R, Maddison R. Smartphone apps to improve fitness and increase physical activity among young people: protocol of the Apps for IMproving FITness (AIMFIT) randomized controlled trial. BMC Public Health 2015 Jul 11;15:635 [FREE Full text] [doi: 10.1186/s12889-015-1968-y] [Medline: 26159834]
23. Parasuraman S, Sam AT, Yee SW, Chuon BL, Ren LY. Smartphone usage and increased risk of mobile phone addiction: a concurrent study. Int J Pharm Investig 2017;7(3):125-131 [FREE Full text] [doi: 10.4103/jphi.JPHI_56_17] [Medline: 29184824]
24. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
25. Quelly SB, Norris AE, DiPietro JL. Impact of mobile apps to combat obesity in children and adolescents: a systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17. [doi: 10.1111/jspn.12134] [Medline: 26494019]
26. 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 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
27. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
28. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):- [FREE Full text] [Medline: 27034992]
29. 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]
30. 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]
31. Ly H. The impact of utilizing mobile phones to promote physical activity among post-secondary students: a scoping review. Mhealth 2016;2:47 [FREE Full text] [doi: 10.21037/mhealth.2016.12.03] [Medline: 28293617]
32. 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]
33. Aromatario O, Van Hoye A, Vuillemin A, Foucaut A, Crozet C, Pommier J, et al. How do mobile health applications support behaviour changes? A scoping review of mobile health applications relating to physical activity and eating behaviours. Public Health 2019 Oct;175:8-18 [FREE Full text] [doi: 10.1016/j.puhe.2019.06.011] [Medline: 31374453]
34. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
35. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
36. 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]
37. 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]
38. -. Mhealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2011.
39. Smith A, Faithe W. CompTIA Strata Study Guide Authorized Courseware. Hoboken, USA: John Wiley & Sons; Nov 28, 2011.
40. Park Y. Emerging new era of mobile health technologies. Healthc Inform Res 2016 Oct;22(4):253-254 [FREE Full text] [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
41. Lee J. 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]

42. Ali EE, Chew L, Yap KY. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan 5;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
43. Dupuy B, Raffestin S, Matamouros S, Mani N, Popoff MR, Sonenshein AL. Regulation of toxin and bacteriocin gene expression in Clostridium by interchangeable RNA polymerase sigma factors. Mol Microbiol 2006 May;60(4):1044-1057 [FREE Full text] [doi: 10.1111/j.1365-2958.2006.05159.x] [Medline: 16677313]
44. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
45. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
46. 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]
47. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
48. 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]
49. 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]
50. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. Mhealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
51. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
52. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
53. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
54. 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]
55. 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]
56. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. Maturitas 2018 Jun;112:85-93. [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
57. de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
58. 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]
59. 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]
60. 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]
61. van Remoortel H, Giavedoni S, Raste Y, Burtin C, Louvaris Z, Gimeno-Santos E, PROactive consortium. Validity of activity monitors in health and chronic disease: a systematic review. Int J Behav Nutr Phys Act 2012 Jul 9;9:84 [FREE Full text] [doi: 10.1186/1479-5868-9-84] [Medline: 22776399]
62. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

63. 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]
64. 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]
65. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
66. O'Brien N, McDonald S, Araújo-Soares V, Lara J, Errington L, Godfrey A, et al. The features of interventions associated with long-term effectiveness of physical activity interventions in adults aged 55-70 years: a systematic review and meta-analysis. Health Psychol Rev 2015;9(4):417-433. [doi: 10.1080/17437199.2015.1012177] [Medline: 25689096]
67. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
68. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
69. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
70. Ehlers DK, Huberty JL. Middle-aged women's preferred theory-based features in mobile physical activity applications. J Phys Act Health 2014 Sep;11(7):1379-1385. [doi: 10.1123/jpah.2012-0435] [Medline: 24368818]
71. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 6;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
72. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mhealth technology to promote physical activity and to sustain behavior change. JMIR Mhealth Uhealth 2016 Jan 20;4(1):e5 [FREE Full text] [doi: 10.2196/mhealth.4814] [Medline: 26792225]
73. Arteaga SM, Kudeki M, Woodworth A, Kurniawan S. Mobile System to Motivate Teenagers' Physical Activity. In: 9th International Conference on Interaction Design and Children. 2010 Presented at: IDC'10; June 9-11, 2010; Barcelona, Spain. [doi: 10.1145/1810543.1810545]
74. Ayubi SU, Parmanto B. PersonA: persuasive social network for physical activity. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:2153-2157. [doi: 10.1109/EMBC.2012.6346387] [Medline: 23366348]
75. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug 22;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
76. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
77. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]
78. Alnasser A, Sathiaseelan A, Al-Khalifa A, Marais D. Development of 'Twazon': an Arabic app for weight loss. JMIR Res Protoc 2016 May 16;5(2):e76 [FREE Full text] [doi: 10.2196/resprot.5497] [Medline: 27185568]
79. Vos S, Janssen M, Goudsmit J, Lauwerijssen C, Brombacher A. From problem to solution: developing a personalized smartphone application for recreational runners following a three-step design approach. Procedia Eng 2016 Mar;147(3):799-805 [FREE Full text] [doi: 10.1016/j.proeng.2016.06.311]
80. Middelweerd A, Te Velde SJ, Mollee JS, Klein MC, Brug J. App-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults (active2gether): descriptive study of the development and content. JMIR Res Protoc 2018 Dec 21;7(12):e185 [FREE Full text] [doi: 10.2196/resprot.7169] [Medline: 30578198]
81. Chang T, Kaasinen E, Kaipainen K. What Influences Users' Decisions to Take Apps Into Use? In: 11th International Conference on Mobile and Ubiquitous Multimedia. 2012 Dec 04 Presented at: MUM'12; June 9-12, 2012; Ulm, Germany. [doi: 10.1145/2406367.2406370]
82. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 1;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
83. Arteaga SM, González VM, Kurniawan S, Benavides RA. Mobile games and design requirements to increase teenagers' physical activity. Pervasive Mob Comput 2012 Dec;8(6):900-908. [doi: 10.1016/j.pmcj.2012.08.002]

84. Arteaga S, Kurniawan S. Designing an Application to Motivate Teenagers' Physical Activity. In: IEEE International Workshop on Mobile Consumer Health Care Networks, Systems and Services. 2012 Presented at: IEEE'12; June 10-15, 2012; Ottawa, Canada. [doi: 10.1109/icc.2012.6364719]
85. Lu F, Turner K. Improving Adolescent Fitness Attitudes With a Mobile Fitness Game to Combat Obesity in Youth. In: IEEE International Games Innovation Conference. 2013 Presented at: IEEE'13; September 23-25, 2013; Vancouver, Canada. [doi: 10.1109/igic.2013.6659162]
86. Casey M, Hayes PS, Glynn F, OLaighin G, Heaney D, Murphy AW, et al. Patients' experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care. Br J Gen Pract 2014 Aug;64(625):e500-e508 [FREE Full text] [doi: 10.3399/bjgp14X680989] [Medline: 25071063]
87. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]
88. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
89. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
90. Hoj TH, Covey EL, Jones AC, Haines AC, Hall PC, Crookston BT, et al. How do apps work? An analysis of physical activity app users' perceptions of behavior change mechanisms. JMIR Mhealth Uhealth 2017 Aug 3;5(8):e114 [FREE Full text] [doi: 10.2196/mhealth.7206] [Medline: 28778846]
91. 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]
92. Cowan LT, van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
93. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
94. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]
95. 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]
96. Yang C, 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]
97. Modave F, Bian J, Leavitt T, Bromwell J, Harris CI, Vincent H. Low quality of free coaching apps with respect to the American college of sports medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]
98. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
99. 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]
100. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
101. 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]
102. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Mar;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
103. Dusseldorp E, van Genugten L, van Buuren S, Verheijden MW, van Empelen P. Combinations of techniques that effectively change health behavior: evidence from Meta-CART analysis. Health Psychol 2014 Dec;33(12):1530-1540. [doi: 10.1037/hea0000018] [Medline: 24274802]

104. Hardcastle SJ, Hancox J, Hattar A, Maxwell-Smith C, Thøgersen-Ntoumani C, Hagger MS. Motivating the unmotivated: how can health behavior be changed in those unwilling to change? Front Psychol 2015;6:835 [FREE Full text] [doi: 10.3389/fpsyg.2015.00835] [Medline: 26136716]
105. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
106. Spring B, Gotsis M, Paiva A, Spruijt-Metz D. Healthy apps: mobile devices for continuous monitoring and intervention. IEEE Pulse 2013;4(6):34-40 [FREE Full text] [doi: 10.1109/MPUL.2013.2279620] [Medline: 24233190]
107. Thompson D, Peacock O, Western M, Batterham AM. Multidimensional physical activity: an opportunity, not a problem. Exerc Sport Sci Rev 2015 Apr;43(2):67-74 [FREE Full text] [doi: 10.1249/JES.0000000000000039] [Medline: 25607280]
108. White BK, Burns SK, Giglia RC, Scott JA. Designing evaluation plans for health promotion mHealth interventions: a case study of the milk man mobile app. Health Promot J Austr 2016 Mar;27(3):198-203. [doi: 10.1071/HE16041] [Medline: 27784506]
109. 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]
110. 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]
111. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17(Suppl 1):5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
112. Monroe CM, Thompson DL, Bassett DR, Fitzhugh EC, Raynor HA. Usability of mobile phones in physical activity–related research: a systematic review. Am J Heal Educ 2015 Jul 6;46(4):196-206. [doi: 10.1080/19325037.2015.1044141]
113. 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]
114. VonHoltz LA, Hypolite KA, Carr BG, Shofer FS, Winston FK, Hanson CW, et al. Use of mobile apps: a patient-centered approach. Acad Emerg Med 2015 Jun;22(6):765-768 [FREE Full text] [doi: 10.1111/acem.12675] [Medline: 25998446]
115. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartsteps. Ann Behav Med 2019 May 3;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
116. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution-using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015 Jun;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]
117. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
118. Ernsting C, Dombrowski SU, Oedekoven M, O'Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101. [doi: 10.2196/jmir.6838] [Medline: 28381394]
119. Gillman AS, Bryan AD. Effects of performance versus game-based mobile applications on response to exercise. Ann Behav Med 2016 Mar;50(1):157-162. [doi: 10.1007/s12160-015-9730-3] [Medline: 26362539]
120. Lyons E, Swartz M. Motivational dynamics of wearable activity monitors. Acsms Heal Fit J 2017;21(5):26. [doi: 10.1249/fit.0000000000000324]
121. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Mar;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
122. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. A randomized controlled trial testing a social network intervention to promote physical activity among adolescents. BMC Public Health 2018 Apr 23;18(1):542 [FREE Full text] [doi: 10.1186/s12889-018-5451-4] [Medline: 29685112]
123. Harries T, Eslambolchilar P, Rettie R, Stride C, Walton S, van Woerden HC. Effectiveness of a smartphone app in increasing physical activity amongst male adults: a randomised controlled trial. BMC Public Health 2016 Sep 2;16:925 [FREE Full text] [doi: 10.1186/s12889-016-3593-9] [Medline: 27590255]
124. Belmon LS, Middelweerd A, te Velde SJ, Brug J. Dutch young adults ratings of behavior change techniques applied in mobile phone apps to promote physical activity: a cross-sectional survey. JMIR Mhealth Uhealth 2015 Nov 12;3(4):e103 [FREE Full text] [doi: 10.2196/mhealth.4383] [Medline: 26563744]
125. Feldman DI, Robison WT, Pacor JM, Caddell LC, Feldman EB, Deitz RL, et al. Harnessing mHealth technologies to increase physical activity and prevent cardiovascular disease. Clin Cardiol 2018 Jul;41(7):985-991 [FREE Full text] [doi: 10.1002/clc.22968] [Medline: 29671879]

126. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. PLoS One 2016;11(6):e0156370 [FREE Full text] [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
127. Benson AC, Bruce L, Gordon BA. Reliability and validity of a GPS-enabled iPhone 'app' to measure physical activity. J Sports Sci 2015;33(14):1421-1428. [doi: 10.1080/02640414.2014.994659] [Medline: 25555093]
128. Wiehe SE, Hoch SC, Liu GC, Carroll AE, Wilson JS, Fortenberry JD. Adolescent travel patterns: pilot data indicating distance from home varies by time of day and day of week. J Adolesc Health 2008 Apr;42(4):418-420. [doi: 10.1016/j.jadohealth.2007.09.018] [Medline: 18346668]
129. Wiehe SE, Carroll AE, Liu GC, Haberkorn KL, Hoch SC, Wilson JS, et al. Using GPS-enabled cell phones to track the travel patterns of adolescents. Int J Health Geogr 2008 May 21;7:22 [FREE Full text] [doi: 10.1186/1476-072X-7-22] [Medline: 18495025]
130. Wright SP, Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 1;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
131. Aitken M, Lyle J. Patient Adoption of Mhealth: Use, Evidenceremaining Barriers to Mainstream Acceptance. IQVIA.: IMS Institute for Healthcare Informatics; 2015. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2020-05-26]
132. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. J Am Med Assoc 2015 Mar 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
133. Lubans DR, Smith JJ, Plotnikoff RC, Dally KA, Okely AD, Salmon J, et al. Assessing the sustained impact of a school-based obesity prevention program for adolescent boys: the ATLAS cluster randomized controlled trial. Int J Behav Nutr Phys Act 2016 Aug 20;13:92 [FREE Full text] [doi: 10.1186/s12966-016-0420-8] [Medline: 27542825]
134. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for improving fitness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
135. Hartin PJ, Nugent CD, McClean SI, Cleland I, Tschanz JT, Clark CJ, et al. The empowering role of mobile apps in behavior change interventions: the gray matters randomized controlled trial. JMIR Mhealth Uhealth 2016 Aug 2;4(3):e93 [FREE Full text] [doi: 10.2196/mhealth.4878] [Medline: 27485822]
136. Jee H. Review of researches on smartphone applications for physical activity promotion in healthy adults. J Exerc Rehabil 2017 Mar;13(1):3-11 [FREE Full text] [doi: 10.12965/jer.1732928.464] [Medline: 28349027]
137. 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):1-9 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
138. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
139. Kukafka R, Khan SA, Kaufman D, Mark J. An evidence-based decision aid to help patients set priorities for selecting among multiple health behaviors. AMIA Annu Symp Proc 2009 Nov 14;2009:343-347 [FREE Full text] [Medline: 20351877]
140. Kukafka R, Jeong IC, Finkelstein J. Optimizing decision support for tailored health behavior change applications. Stud Health Technol Inform 2015;216:108-112. [Medline: 26262020]
141. Li I, Dey A, Forlizzi J. Using context to reveal factors that affect physical activity. ACM Trans Comput Hum Interact 2012 Mar;19(1):1-21 [FREE Full text] [doi: 10.1145/2147783.2147790]
142. Choe EK, Abdullah S, Rabbi M, Thomaz E, Epstein DA, Cordeiro F, et al. Semi-automated tracking: a balanced approach for self-monitoring applications. IEEE Pervasive Comput 2017 Jan;16(1):74-84. [doi: 10.1109/mprv.2017.18]
143. Lu H, Yang J, Liu Z, Lane N, Choudhury T, Campbell A. The Jigsaw continuous sensing engine for mobile phone applications. In: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems. 2010 Presented at: SenSys'10; November 3-5, 2010; Zürich, Switzerland p. 71-84. [doi: 10.1145/1869983.1869992]
144. Middelweerd A, Mollee JS, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
145. Recio GM, García-Hernández L, Luque RM, Salas-Morera L. The role of interdisciplinary research team in the impact of health apps in health and computer science publications: a systematic review. Biomed Eng Online 2016 Jul 15;15(Suppl 1):77 [FREE Full text] [doi: 10.1186/s12938-016-0185-y] [Medline: 27454164]
146. Blackman KC, Zoellner J, Kadir A, Dockery B, Johnson SB, Almeida FA, et al. Examining the feasibility of smartphone game applications for physical activity promotion in middle school students. Games Health J 2015 Oct;4(5):409-419. [doi: 10.1089/g4h.2014.0120] [Medline: 26287931]
147. 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]

148. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
149. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
150. Dallinga JM, Mennes M, Alpay L, Bijwaard H, de la Faille-Deutekom MB. App use, physical activity and healthy lifestyle: a cross sectional study. BMC Public Health 2015 Aug 28;15:833 [FREE Full text] [doi: 10.1186/s12889-015-2165-8] [Medline: 26316060]
151. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
152. Duncan MJ, Vandelanotte C, Trost SG, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Jul 30;16:670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
153. Matsumoto T, Doi K, Kano A, Nakamura H, Nakanishi T. [Evaluation of the potential benefit of computer-aided diagnosis (CAD) for lung cancer screenings using photofluorography: analysis of an observer study]. Nihon Igaku Hoshasen Gakkai Zasshi 1993 Oct 25;53(10):1195-1207. [Medline: 8255750]
154. Garde A, Umedaly A, Abulnaga SM, Robertson L, Junker A, Chanoine JP, et al. Assessment of a mobile game ("Mobilekids Monster Manor") to promote physical activity among children. Games Health J 2015 Apr;4(2):149-158. [doi: 10.1089/g4h.2014.0095] [Medline: 26181809]
155. Gaudet J, Gallant F, Bélanger M. A Bit of Fit: Minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
156. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. SMART MOVE - a smartphone-based intervention to promote physical activity in primary care: study protocol for a randomized controlled trial. Trials 2013 May 29;14:157 [FREE Full text] [doi: 10.1186/1745-6215-14-157] [Medline: 23714362]
157. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):384-391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
158. Hirsch JA, James P, Robinson JR, Eastman KM, Conley KD, Evenson KR, et al. Using MapMyFitness to place physical activity into neighborhood context. Front Public Health 2014;2:19 [FREE Full text] [doi: 10.3389/fpubh.2014.00019] [Medline: 24653982]
159. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
160. Lu F, Lemonde M. Reducing adolescent obesity with a social networking mobile fitness application. In: Proceedings of the IEEE 16th International Conference on e-Health Networking, Application & Services. 2014 Presented at: HealthCom 2014 : IEEE 16th International Conference on e-Health Networking, Application & Services; Oct 15-18, 2014; Natal, Brazil p. 429-434. [doi: 10.1109/HealthCom.2014.7001881]
161. Lu F, Turner K, Murphy B. Reducing adolescent obesity with a mobile fitness application: study results of youth age 15 to 17. In: Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013). 2013 Presented at: IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013); Oct 9-12, 2013; Lisbon, Portugal p. 554-558. [doi: 10.1109/HealthCom.2013.6720738]
162. Lubans DR, Smith JJ, Peralta LR, Plotnikoff RC, Okely AD, Salmon J, et al. A school-based intervention incorporating smartphone technology to improve health-related fitness among adolescents: rationale and study protocol for the NEAT and ATLAS 2.0 cluster randomised controlled trial and dissemination study. BMJ Open 2016 Jun 27;6(6):e010448 [FREE Full text] [doi: 10.1136/bmjopen-2015-010448] [Medline: 27354070]
163. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
164. Pellegrini CA, Steglitz J, Johnston W, Warnick J, Adams T, McFadden HG, et al. Design and protocol of a randomized multiple behavior change trial: Make Better Choices 2 (MBC2). Contemp Clin Trials 2015 Mar;41:85-92 [FREE Full text] [doi: 10.1016/j.cct.2015.01.009] [Medline: 25625810]
165. Pirolli P, Mohan S, Venkatakrishnan A, Nelson L, Silva M, Springer A. Implementation intention and reminder effects on behavior change in a mobile health system: a predictive cognitive model. J Med Internet Res 2017 Nov 30;19(11):e397 [FREE Full text] [doi: 10.2196/jmir.8217] [Medline: 29191800]
166. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MD, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence

to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
167. Recio-Rodriguez JI, Gómez-Marcos MA, Agudo-Conde C, Ramirez I, Gonzalez-Viejo N, Gomez-Arranz A, EVIDENT 3 investigators. EVIDENT 3 Study: A randomized, controlled clinical trial to reduce inactivity and caloric intake in sedentary and overweight or obese people using a smartphone application: study protocol. Medicine (Baltimore) 2018 Jan;97(2):e9633 [FREE Full text] [doi: 10.1097/MD.0000000000009633] [Medline: 29480874]
168. Recio-Rodríguez JI, Martín-Cantera C, González-Viejo N, Gómez-Arranz A, Arietaleanizbeascoa MS, Schmolling-Guinovart Y, EVIDENT Group. Effectiveness of a smartphone application for improving healthy lifestyles, a randomized clinical trial (EVIDENT II): study protocol. BMC Public Health 2014 Mar 15;14:254 [FREE Full text] [doi: 10.1186/1471-2458-14-254] [Medline: 24628961]
169. Rospo G, Valsecchi V, Bonomi AG, Thomassen IW, van Dantzig S, La Torre A, et al. Cardiorespiratory improvements achieved by American College of Sports Medicine's exercise prescription implemented on a mobile app. JMIR Mhealth Uhealth 2016 Jun 23;4(2):e77 [FREE Full text] [doi: 10.2196/mhealth.5518] [Medline: 27339153]
170. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
171. Shin DW, Joh HK, Yun JM, Kwon HT, Lee H, Min H, et al. Design and baseline characteristics of participants in the Enhancing Physical Activity and Reducing Obesity through Smartcare and Financial Incentives (EPAROSFI): a pilot randomized controlled trial. Contemp Clin Trials 2016 Mar;47:115-122. [doi: 10.1016/j.cct.2015.12.019] [Medline: 26744232]
172. Shin DW, Yun JM, Shin JH, Kwon H, Min HY, Joh HK, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
173. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Rationale and study protocol for the 'active teen leaders avoiding screen-time' (ATLAS) group randomized controlled trial: an obesity prevention intervention for adolescent boys from schools in low-income communities. Contemp Clin Trials 2014 Jan;37(1):106-119. [doi: 10.1016/j.cct.2013.11.008] [Medline: 24291151]
174. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):723-731. [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
175. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mhealth intervention for large, sustained change in multiple diet and activity risk behaviors: the make better choices 2 randomized controlled trial. J Med Internet Res 2018 Jun 19;20(6):e10528 [FREE Full text] [doi: 10.2196/10528] [Medline: 29921561]
176. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
177. Toscos T, Faber A, Connelly K, Upoma AM. Encouraging physical activity in teens Can technology help reduce barriers to physical activity in adolescent girls? In: Proceedings of the 2008 Second International Conference on Pervasive Computing Technologies for Healthcare. 2008 Presented at: 2008 Second International Conference on Pervasive Computing Technologies for Healthcare; Jan. 30-Feb. 1, 2008; Tampere, Finland p. 218-221. [doi: 10.1109/PCTHEALTH.2008.4571073]
178. van Dantzig S, Geleijnse G, van Halteren A. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246. [doi: 10.1007/s00779-012-0588-0]
179. Voth EC, Oelke ND, Jung ME. A theory-based exercise app to enhance exercise adherence: a pilot study. JMIR Mhealth Uhealth 2016 Jun 15;4(2):e62 [FREE Full text] [doi: 10.2196/mhealth.4997] [Medline: 27307134]
180. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
181. Watterson TA. Changes in attitudes and behaviors toward physical activity, nutrition, and social support for middle school students using the AFIT app as a suppliment to instruction in a physical education class. Graduate Theses and Dissertation, University of South Florida. 2012. URL: https://core.ac.uk/download/pdf/154469964.pdf [accessed 2021-07-02]
182. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17(Suppl 1):11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
183. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behaviour Change Theories. London, UK: Silverback Publishing; 2014.
184. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. London, UK: Silverback Publishing; 2014.
185. Wood D, Crapnell T, Lau L, Bennett A, Lotstein D, Ferris M. Emerging Adulthood as a Critical Stage in the Life Course. New York, USA: Springer; 2018.

186. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015 Sep;5(3):335-346 [FREE Full text] [doi: 10.1007/s13142-015-0324-1] [Medline: 26327939]
187. 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]
188. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
189. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
190. Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PK, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Mar 5;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
191. Rabbi M, Kotov MP, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, et al. Toward increasing engagement in substance use data collection: development of the substance abuse research assistant app and protocol for a microrandomized trial using adolescents and emerging adults. JMIR Res Protoc 2018 Jul 18;7(7):e166 [FREE Full text] [doi: 10.2196/resprot.9850] [Medline: 30021714]
192. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
193. Activity Monitor Comparison. ActiGraph. 2019. URL: https://actigraphcorp.com/activity-monitor-comparison/ [accessed 2020-05-26]
194. Michie S, Thomas J, Johnston M, Aonghusa PM, Shawe-Taylor J, Kelly MP, et al. The human behaviour-change project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implement Sci 2017 Oct 18;12(1):121 [FREE Full text] [doi: 10.1186/s13012-017-0641-5] [Medline: 29047393]
195. Rabbi M, Aung MS, Gay G, Reid MC, Choudhury T. Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: pilot study on adults. J Med Internet Res 2018 Oct 26;20(10):e10147 [FREE Full text] [doi: 10.2196/10147] [Medline: 30368433]
196. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
197. Garmin® Health and Actigraph™ Collaborate on Wearable Solutions for Clinical Trials. ActiGraph. 2018. URL: https://www.actigraphcorp.com/press-release/garmin-health-and-actigraph-collaborate-on-wearable-solutions-for-clinical-trials/ [accessed 2020-05-26]

Abbreviations

BCT: behavior change technique
CALO-RE: Coventry, Aberdeen & London–Refined
COM-B: Capability, Opportunity, Motivation, Behaviour
CONSORT: Consolidated Standards of Reporting Trials
HE: healthy eating
HR: heart rate
MARS: Mobile App Rating Scale
mHealth: mobile health
MRT: microrandomized trial
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols
RCT: randomized controlled trial
SB: sedentary behavior
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
WHO: World Health Organization

Edited by L Buis; submitted 16.09.20; peer-reviewed by M Rodrigues, J Alvarez Pitti, M Hayman; comments to author 24.12.20; revised version received 12.02.21; accepted 16.04.21; published 21.07.21.

Please cite as:
Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C
Smartphone-Based Interventions for Physical Activity Promotion: Scoping Review of the Evidence Over the Last 10 Years
JMIR Mhealth Uhealth 2021;9(7):e24308
URL: https://mhealth.jmir.org/2021/7/e24308
doi:10.2196/24308
PMID:34287209

Original Paper

Determining the Evolution of Headache Among Regular Users of a Daily Electronic Diary via a Smartphone App: Observational Study

Bianca Raffaelli[1], MD; Jasper Mecklenburg[1], MD; Lucas Hendrik Overeem[1], MSc; Simon Scholler[2], MSc; Markus A Dahlem[2], MSc; Tobias Kurth[3], MD; Ana Sofia Oliveira Gonçalves[3], MSc; Uwe Reuter[1], MD; Lars Neeb[1], MD

[1]Department of Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany

[2]Newsenselab GmbH, Berlin, Germany

[3]Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Bianca Raffaelli, MD
Department of Neurology
Charité – Universitätsmedizin Berlin
1 Charitéplatz
Berlin, 10117
Germany
Phone: 49 30 450 660888
Fax: 49 30 450 560932
Email: bianca.raffaelli@charite.de

Abstract

Background: Smartphone-based apps represent a major development in health care management. Specifically in headache care, the use of electronic headache diaries via apps has become increasingly popular. In contrast to the soaring volume of available data, scientific use of these data resources is sparse.

Objective: In this analysis, we aimed to assess changes in headache and migraine frequency, headache and migraine intensity, and use of acute medication among people who showed daily use of the headache diary as implemented in the freely available basic version of the German commercial app, M-sense.

Methods: The basic version of M-sense comprises an electronic headache diary, documentation of lifestyle factors with a possible impact on headaches, and evaluation of headache patterns. This analysis included all M-sense users who had entered data into the app on a daily basis for at least 7 months.

Results: We analyzed data from 1545 users. Mean MHD decreased from 9.42 (SD 5.81) at baseline to 6.39 (SD 5.09) after 6 months (P<.001; 95% CI 2.80-3.25). MMD, AMD, and migraine intensity were also significantly reduced. Similar results were found in 985 users with episodic migraine and in 126 users with chronic migraine.

Conclusions: Among regular users of an electronic headache diary, headache and migraine frequency, in addition to other headache characteristics, improved over time. The use of an electronic headache diary may support standard headache care.

(JMIR Mhealth Uhealth 2021;9(7):e26401) doi:10.2196/26401

KEYWORDS

headache; migraine; mobile app; headache app; electronic diary; app; pain; frequency; intensity

Introduction

The rapid development of modern technologies has led to major advances in health care [1,2]. The trend toward eHealth and, in particular, mobile health (mHealth) has entered the headache field and includes the use of smartphone-based apps. In recent years, headache apps have increased not just in number but also in functionality. Functions comprise headache documentation in so called e-diaries, the recording of lifestyle factors that may potentially influence headaches, and the automatic incorporation of external information, such as weather conditions. Some apps also provide digital-based education programs and self-managed nonpharmacological interventions, like relaxation methods.

Tracking headaches and associated symptoms is a pillar of headache management. Digital headache documentation has proven to have a higher compliance than do paper headache diaries [3,4]. From a clinical perspective, proper documentation can facilitate health care providers in diagnosing headache disorders, evaluating the course of the disease, and assessing treatment effects. It also provides patients with a better understanding and awareness for their headaches [5]. Through a comprehensive analysis of headache attacks and associated factors, digital algorithms can recognize and offer a clear presentation of headache patterns and may improve the self-management of headaches. Conversely, the constant dealing with headache symptoms and potential triggers may lead to an increased focus on the disease. In this way, these apps could even become stressors themselves and cause a worsening of headaches [6].

Hence, apps may offer new opportunities but also entail possible risks in the outpatient care of migraine patients [7]. Previous studies indicate that the complementary use of smartphone apps and other internet-based technologies could lead to a better treatment of headache patients compared to standard care [8]. However, evidence of positive health outcomes through the use of mHealth is scarce [9]. Most available headache apps lack scientific evaluation, and uniform quality standards have not yet been established [10,11].

In this analysis, we aimed to assess the evolution of headache characteristics among regular users of the freely available basic version of a commercial headache app (M-sense). We tested whether headache and migraine frequency, headache and migraine intensity, and days with intake of acute medication differed between baseline and after 6 months. In extension analyses, we compared outcomes at baseline to those at 12 months.

Methods

M-sense

M-sense is a certified German headache app developed by Newsenselab. In Germany, Austria, and Switzerland, it has been available for Android since 2016 and for iOS since 2017 [12]. From September 2016 to May 2020, the app was downloaded approximately 250,000 times and had nearly 85,000 registered users.

There are 2 versions of the app: a free “basic” version, and an additional, paid, “active” version. Our study focused on the basic version. In this version, users can document their headache information in a diary according to a predefined scheme. This includes the start and end of the headache, maximal pain intensity (on a 11-point numeric rating scale from 0 to 10), unilateral headache (yes or no), throbbing headache (yes or no), worsening through physical activity (yes or no), nausea (yes or no), vomitus (yes or no), photophobia (yes or no), phonophobia (yes or no), and migraine aura (yes or no). For each headache attack, users can record the intake of acute medication including the name and dose of the drug and time of intake. A standardized, validated algorithm based on the International Classification of Headache Disorders (ICHD-3) criteria [13] classifies single headache attacks as migraine, tension-type headache (TTH), or nonmigraine or non-TTH [14]. A detailed description of the classification algorithm can be found in the publication by Roesch et al [14]. For the purposes of this study, we assessed migraine days and headache days. A migraine day was defined as a calendar day on which the user experienced a qualified migraine attack. To identify if an attack should be counted as migraine or TTH, we applied all relevant ICHD-3 guidelines [13]. A qualified migraine attack needed to meet the following ICHD-3 criteria for migraine: a headache with or without aura lasting at least 4 hours with both features A (at least 2 of the following: unilateral location, pulsatile quality, moderate or severe pain intensity, and aggravation caused by physical activity or avoidance of physical activity) and B (during headache, at least 1 of the following: nausea and/or vomiting and/or photophobia and phonophobia). Any headache attacks accompanied by intake of a migraine-specific medication (eg, triptans) to treat a headache or accompanied by aura were defined as migraine regardless of the duration and pain features or associated symptoms. The ICHD-3 classifications of TTH and probable migraine differ depending on whether or not a patient has received a diagnosis. During the onboarding process of the app, users were asked if they received a headache diagnosis by a health care professional. In users with a previous migraine diagnosis, all headaches that fulfilled the criteria of both TTH and probable migraine were counted as migraine according to the ICHD-3. A headache day was defined as any calendar day on which the user experienced a headache attack (including migraine). The app provides an exportable, graphic overview of headache attacks, subdivided into migraine, TTH, and others.

Users can also enter daily data on predefined lifestyle factors with a possible impact on headaches (eg, sleep duration, sleep quality, or stress level). In addition, the app collects weather data. Users can see the course of these factors over time and their mean values on the days immediately preceding a headache and on headache days. After 60 days of daily entries, the app analyzes correlations between the entered factors and headaches and, if necessary, a further medical consultation is recommended.

Population and Outcomes

This analysis included all M-sense users who entered data in the app every day for at least 7 months on a daily basis. No further inclusion or exclusion criteria were applied.

The primary outcome was the number of monthly headache days (MHD). Secondary outcomes included monthly migraine days (MMD), monthly days with intake of acute headache medication (AMD), mean monthly headache intensity, and mean monthly migraine intensity.

The outcomes were selected based on recent guidelines for migraine trials [15]. As recommended in the guidelines for trials of behavioral treatments for recurrent headache, we focused on all monthly headache days and not only on moderate or severe monthly headache days, in addition to monthly migraine days [16].

A month was defined as 4 weeks (28 days), beginning with the first day of app use. Both triptan and nontriptan pain medication (eg, nonsteroidal anti-inflammatory drugs or paracetamol) counted as acute headache medication.

We defined the first month after the app installation as baseline. We compared the primary and secondary outcome measures between the baseline phase and the sixth month after baseline (ie, the seventh month of app use).

After analyzing the entire population, we performed subgroup analyses for users with headache and migraine frequency compatible with the diagnosis of episodic or chronic migraine according to the ICHD-3 criteria [13] and based on the first 90 days of app use. Another concomitant headache diagnosis (eg, TTH in patients with episodic migraine) was possible.

Additionally, we conducted an analysis for users who continued to use the app for at least 13 months. In this population, we compared the aforementioned parameters between the baseline month and the 12th month after baseline (ie, the 13th month of app use). We further recorded the headache frequency and medication use 1, 7, and 13 months after users had begun accessing the app.

Data Processing

For data protection reasons, Newsenselab provided a sample data set containing dummy data in the same format as the real data set. We tested our analysis code locally against the sample data set. The Newsenselab team (SS and MAD) then ran the analysis code on the real, personal data and provided the aggregated output data. Through this procedure, the research team had no access to the individual data sets but only to aggregated results for the predefined sample and outcome variables.

Ethics

By installing the app, the users agreed to the general terms and conditions of M-sense as well as the privacy policy. These include the storage of personal and health data on an Amazon Web Services (Amazon.com, Inc) server in encrypted form and the transfer of these data in anonymous form to third parties for medical research purposes.

We used only aggregated data. In accordance with the local legislation and institutional requirements, use of aggregated data does not require institutional review board approval.

Statistical Analysis

The statistical analysis was performed using R version 3.6.2 (The R Foundation for Statistical Computing). Demographics and monthly headache characteristics were summarized with descriptive statistical methods. For categorical variables, we report absolute frequencies and percentages. For numerical variables, we report mean and SD.

Outcome measures were compared between baseline and the last considered month using paired samples *t* tests. Due to the large sample size, calculation for normal distribution was not necessary [17]. To compare whether the headache frequencies at baseline, month 6, and month 12 differed, we used analysis of variance (ANOVA) for repeated measures. A 2-tailed *P* value ≤.05 was considered statistically significant.

Results

Demographics and Headache Diagnosis

Between September 2016 and May 2020, 1545 users recorded headache information every day for 7 months (Figure 1). Of the users with available data, most were female (920/1047, 87.87%) with a mean age of 37.2 years (SD 11.1).

The app supported the diagnosis of “episodic migraine” in 985 cases and “chronic migraine” in 126 cases. The other users reported headaches compatible with the diagnosis of TTH or a not-further-classified headache disorder. Demographic features for all patients, patients with episodic migraine, and patients with chronic migraine are shown in Table 1.

Figure 1. Flowchart of the user selection process.

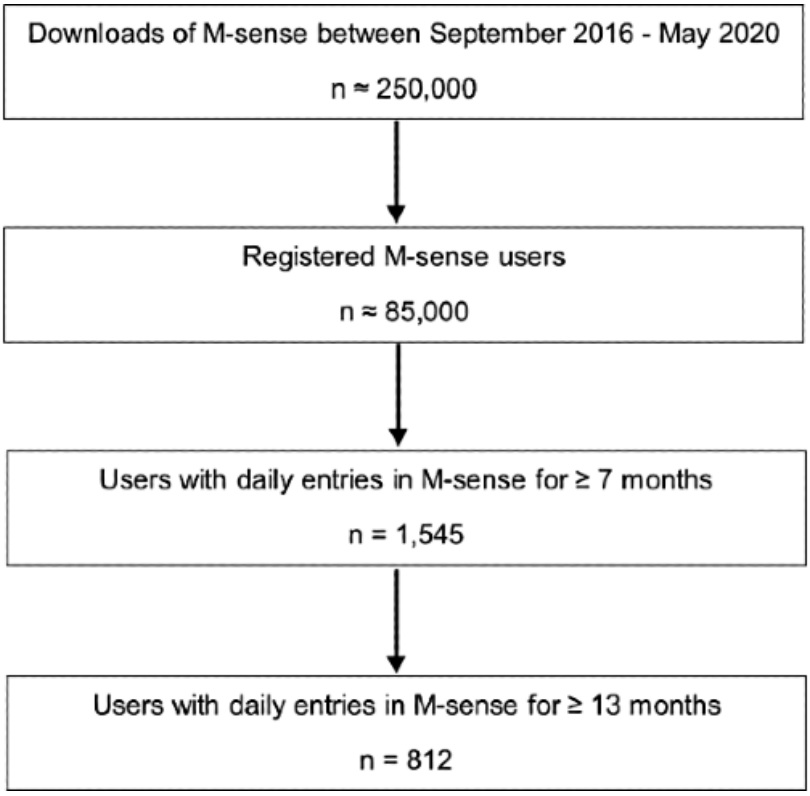


Table 1. Demographic features of M-sense users with daily data entries for 6 months after baseline. Percentages are calculated for the available data.

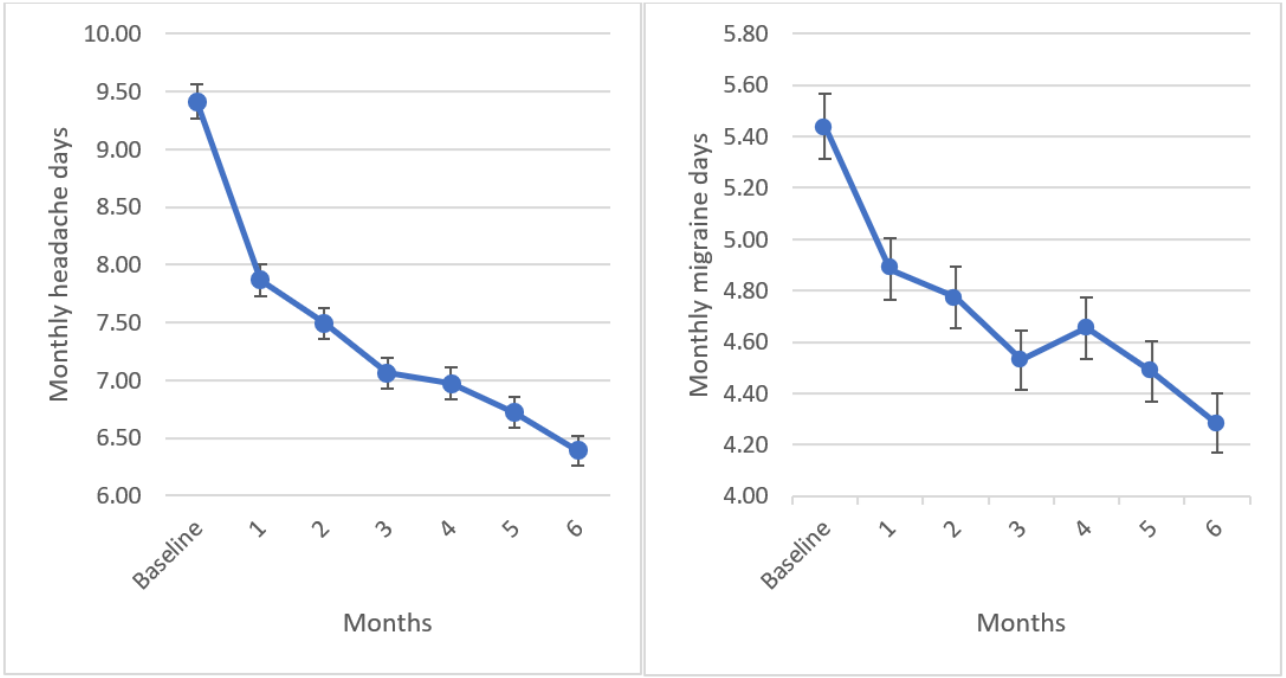
Characteristic	All users	Episodic migraine	Chronic migraine
Distribution, n	1545	985	126
Female sex, n (%)	920 (87.9)	607 (89.8)	85 (91.4)
Missing data for sex, n	498	309	33
Age (years), mean (SD)	37.2 (11.1)	37.7 (11.1)	36.7 (11.7)
Missing data for age, n	393	252	23

Changes in Headache Characteristics Over 6 Months of App Use

During the first month of use, users reported on average 9.42 MHD (SD 5.81), which decreased to 6.39 (SD 5.09) after 6 months (P<.001; 95% CI 2.80-3.25). MMD decreased from 5.44 (SD 4.98) during baseline to 4.28 (SD 4.56) after 6 months (P<.001; 95% CI 1.04-1.48). Figure 2 shows MMD and MHD during the first 6 months of use after baseline.

AMD and migraine intensity also reduced from baseline to the sixth month after baseline (P<.001 for both). Headache intensity showed a numerically small, yet statistically significant increase from a mean 4.57 (SD 1.50) during baseline to a mean 4.71 (SD 1.85) in the sixth month after baseline (P<.001; 95% CI –0.19 to –0.06).

Figure 2. Monthly headache days and monthly migraine days (mean and SE) during baseline and in the first 6 months of M-sense use after baseline.



Subgroup analysis revealed a significant reduction in MMD, MHD, and AMD for both users with episodic and chronic headache in the sixth month after baseline. The slight increase in headache intensity remained significant only for users with episodic migraine, rising from a mean 4.77 (SD 1.47) during baseline to a mean 4.88 (SD 1.73) in the sixth month after baseline (P=.02; 95% CI –0.19 to –0.02), but not for users with chronic migraine. Headache characteristics for all patients, patients with episodic migraine, and patients with chronic migraine are shown in Table 2.

Table 2. Headache or migraine frequency and intensity, and use of acute medication during baseline at the third and the sixth month after baseline for all M-sense users, users with episodic migraine, and users with chronic migraine.

Outcomes	Baseline, mean (SD)	Month 6, mean (SD)	P value[a]	95% CI
All M-sense users (N=1545)				
Monthly headache days	9.42 (5.81)	6.39 (5.09)	<.001	2.80 to 3.25
Monthly migraine days	5.44 (4.98)	4.28 (4.56)	<.001	0.96 to 1.35
Monthly days with acute medication use	5.62 (4.64)	4.35 (4.08)	<.001	1.08 to 1.45
Mean monthly headache intensity	4.57 (1.50)	4.71 (1.85)	<.001	–0.19 to –0.06
Mean monthly migraine intensity	5.42 (1.78)	5.22 (1.85)	<.001	0.09 to 0.27
M-sense users with episodic migraine (n=985)				
Monthly headache days	8.23 (4.01)	5.63 (3.71)	<.001	2.35 to 2.85
Monthly migraine days	5.64 (3.61)	4.38 (3.54)	<.001	1.04 to 1.48
Monthly days with acute medication use	5.64 (4.43)	4.39 (3.91)	<.001	1.03 to 1.47
Mean monthly headache intensity	4.77 (1.47)	4.88 (1.73)	.02	–0.19 to –0.02
Mean monthly migraine intensity	5.36 (1.64)	5.19 (1.86)	.002	0.06 to 0.26
M-sense users with chronic migraine (n=126)				
Monthly headache days	21.08 (4.90)	15.03 (7.03)	<.001	4.91 to 7.18
Monthly migraine days	15.84 (6.02)	12.05 (7.21)	<.001	2.54 to 5.04
Monthly days with acute medication use	8.79 (6.09)	6.71 (6.03)	<.001	1.29 to 2.85
Mean monthly headache intensity	5.40 (1.69)	5.50 (1.90)	.21	–0.34 to 0.08
Mean monthly migraine intensity	5.83 (1.46)	5.81 (1.72)	.22	–0.19 to 0.22

[a]P values as calculated with paired samples t test.

Extended Analysis up to Month 12

For the extension up to 1 year, we analyzed the data of 812 users. Of the 559 users with available data, most (500/559, 89.4%) were female and on average 37.16 (SD 11.04) years old. The app algorithm supported the diagnosis of episodic migraine in 504 cases, chronic migraine in 68 cases, and another headache diagnosis in 240 cases.

These users reported a mean MHD of 9.27 (SD 5.88) during the baseline month, which gradually declined to 6.03 (SD 5.13) after 12 months (P<.001; 95% CI 2.90 to 3.57), as shown in Figure 3. MMD were reduced from a mean 5.20 (SD 4.95) during baseline to a mean 4.14 (SD 4.50) after 12 months (P<.001; 95% CI 0.79-1.33). Users also documented a decreased number of AMD (baseline: mean 5.54, SD 4.67; month 12: mean 4.16, SD 3.96; P<.001; 95% CI 1.19-1.64). Mean migraine intensity decreased significantly from baseline to month 12 (P=.004), while mean headache intensity was higher in month 12 (P=.02). Separate analyses for users with episodic and chronic migraine revealed similar results (Table 3).

Figure 3. Monthly headache days, monthly migraine days, and monthly days with acute medication use (mean and SE) during baseline, and after 6 and 12 months of app use. Outcomes were compared using repeated-measures analysis of variance. AMD: days with acute medication use; HMD: monthly headache days; MMD: monthly migraine days .*P<.001.

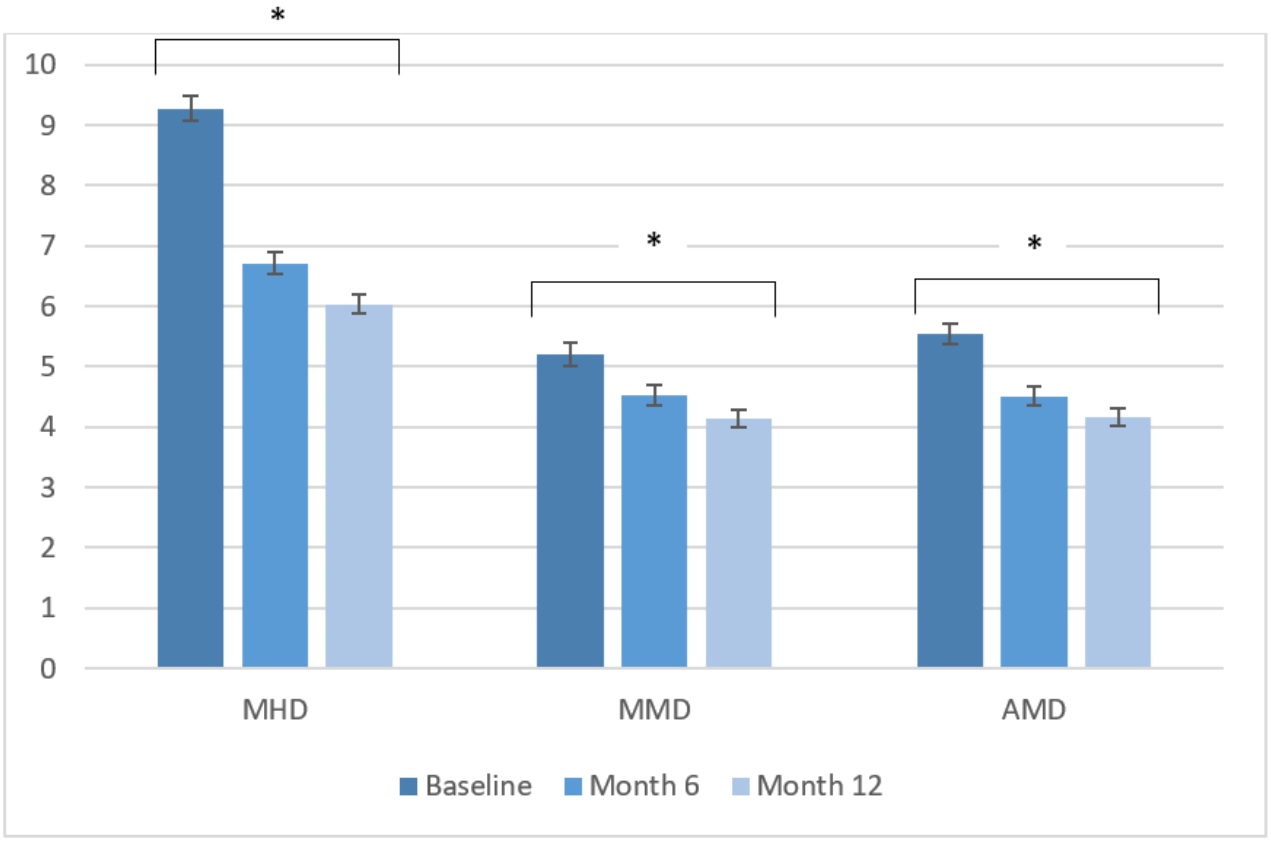


Table 3. Headache and migraine frequency and intensity, and use of acute medication during baseline and the 12th month after baseline for all M-sense users, users with episodic migraine, and users with chronic migraine.

Outcomes	Baseline, mean (SD)	Month 12, mean (SD)	*P* value[a]	95% CI
All M-sense users (N=812)				
Monthly headache days	9.27 (5.88)	6.03 (5.13)	<.001	2.90 to 3.57
Monthly migraine days	5.20 (4.95)	4.14 (4.50)	<.001	0.79 to 1.33
Monthly days with acute medication use	5.54 (4.67)	4.16 (3.96)	<.001	1.19 to 1.64
Mean monthly headache intensity	4.40 (1.52)	4.66 (1.78)	.02	–0.24 to –0.03
Mean monthly migraine intensity	5.33 (1.86)	5.11 (1.74)	.004	0.06 to 0.32
M-sense users with episodic migraine (n=504)				
Monthly headache days	8.09 (4.03)	5.42 (3.77)	<.001	2.29 to 3.04
Monthly migraine days	5.48 (3.53)	4.35 (3.69)	<.001	0.80 to 1.46
Monthly days with acute medication use	5.65 (4.67)	4.31 (4.03)	<.001	0.99 to 1.70
Mean monthly headache intensity	4.68 (1.52)	5.04 (1.74)	.26	–0.18 to 0.05
Mean monthly migraine intensity	5.23 (1.86)	5.04 (1.74)	.02	0.03 to 0.34
M-sense users with chronic migraine (n=68)				
Monthly headache days	21.31 (4.99)	14.21 (7.42)	<.001	5.28 to 8.92
Monthly migraine days	15.38 (6.36)	11.06 (7.30)	<.001	2.65 to 5.99
Monthly days with acute medication use	8.87 (5.63)	6.18 (5.17)	<.001	1.60 to 3.78
Mean monthly headache intensity	5.32 (1.70)	5.44 (1.89)	.45	–0.41 to 0.18
Mean monthly migraine intensity	5.80 (1.59)	5.82 (1.76)	.99	–0.28 to 0.28

[a]*P* values as calculated with paired samples *t* test.

Discussion

We observed a significant improvement of headaches in patients who used the electronic headache diary on the M-sense app on a daily basis. The frequency of monthly migraine and headache days, days with use of acute medication, and migraine intensity declined over 6 months of daily app use. This decrease remained for regular users of the app after 12 months. These changes applied to the entire user population and to the subpopulations with episodic migraine and chronic migraine.

The observed changes in headache frequency are comparable to those of open-label studies with nonpharmacological treatments for headache prevention. Previous research on cognitive-behavioral treatment, relaxation training, or aerobic exercise in patients with primary headache disorders reported a reduction of 2 to 3 MHD after approximately 6 months of treatment, which is in the range of our analysis [18-20]. Of note, our population achieved this clinically meaningful improvement without any specific intervention apart from the basic M-sense app.

The main function of the basic version of the app is the electronic headache diary. The documentation of headache episodes in an app offers various advantages over a paper-and-pencil calendar. Most people have access to their smartphone at all times. This allows a more rapid “real-time” documentation in an app, while paper documentation is usually performed at a later time point and may be affected by a recall bias [3]. Backfilling of entries for several days is a common issue in paper calendars. In a study on chronic pain, 80 participants were asked to complete standardized pain assessments 3 times a day either in a paper or in an electronic calendar for 3 weeks [3]. The electronic diary was completed in 94% of cases within a predefined 30-minute time window, while this was the case for only 11% of entries in the paper calendar [3]. The group with the paper calendar documented about one-third of data after more than 1 day [3].

In our analysis, users documented headache information every day. Although completing diaries daily could be challenging [21], this allows for better data quality and validity. Completers of a daily internet-based migraine diary reported that it was a “major commitment but worthwhile” [22] and that it contributed to a better understanding of their headache disorder. An interactive visual presentation also facilitates data interpretation [23]. The graphic visualization of headache attacks in a monthly calendar may contribute to recognizing headache patterns over time and to raising self-awareness of the headache disorder [24]. Self-awareness in turn can lead to behavioral changes and a better treatment choice and adherence [25,26].

Another core function of the basic M-sense app is the recording of lifestyle factors that may trigger migraine attacks. The identification of individual trigger factors might also induce changes in behavior and influence the course of the headache disorders [27]. However, we cannot provide data on the number of individuals who used the trigger analysis on a regular basis. It was our primary goal to observe headache characteristics in users of a headache app in a real-world setting and not to analyze

the single, specific functions. Therefore, we considered all offered app functions as part of a whole and not separately. Future research analyzing or comparing single app functions could provide further insights into the effectiveness of different features for different groups of users. Individual customization options might then enable users to adapt the app interface to their needs and obtain the largest possible benefits.

Comparisons between our analysis and other research on headache apps are only possible to a limited extent. In a review from 2016, Mosadeghi et al [28] identified only 6 studies about mobile apps for headache disorders. None of these studies assessed the change of headache characteristics during the use of a headache app and only evaluated the usability and feasibility of the apps through user surveys. In 2019, Göbel et al [29] reported results regarding the use of Migräne-App. Like M-sense, this app includes the documentation of headache information in a diary and the registration of weather data. Further features include a search function for headache experts, instructions for progressive muscle relaxation, a media library with educational resources, and an aura simulation. In contrast to M-sense, it does not record or analyze other lifestyle factors. Via an online survey, users of Migräne-App were questioned about their app use, their satisfaction with the app, and also about changes in headache patterns during the app use. Similar to our results, 1464 users reported a reduction of about 3 MHD and 1 AMD after using the app for more than 1 year. In contrast to our analysis, users manually entered their headache data from the app in an online survey, while we analyzed data directly from the app and therefore encountered a much lower risk of bias.

The main strengths of this analysis are the large sample size and the minimization of missing data, as we selected only users with daily app entries for up to 1 year. However, this could have led to a selection bias in favor of users who experienced noticeable benefits through the app, as users who experienced no benefit might have stopped using the app earlier. Furthermore, the daily use of an app also indicates a dedication to managing disease, which might have led to other preventive actions outside the use of the app. We had no further information about other pharmaceutical and nonpharmaceutical preventive therapies used by patients during the observation period that might have contributed to the change of headache frequency and intensity. We also cannot provide data about the frequency of app use in nondaily users or about how many users stopped using the app, as we received aggregated data only for the predefined population. The assessment of app-related risks was not within the scope of our analysis.

Another limitation concerns the classification of single headache attacks via the app algorithm. During app installation, users could indicate a pre-existing headache diagnosis. In users without a known headache diagnosis, TTH could be overrepresented because headache attacks with the characteristics of both probable migraine and TTH were classified by the algorithm as TTH according to the "general rule of hierarchy" in the ICHD-3 criteria [13]. However, in patients with a previous migraine diagnosis, these attacks were classified as migraine [14]. Further limitations relate to the definition of the episodic migraine and chronic migraine subgroups based on the classification by the app. To diagnose a headache disorder solely on the ground of the app data is not justified because characteristics and signs for other primary headaches or secondary headaches are not collected by M-sense and a physical examination is not included in the assessment. A definite headache diagnosis is only possible by the integration of the app data with the clinical assessment of a physician. Hence, we cannot rule out the possibility that some patients included in the study experienced headache disorders other than migraine or TTH.

Moreover, we included only those users who downloaded the basic version of M-sense. Other studies suggested a benefit of behavioral interventions through headache apps [30]. A multicenter, randomized controlled trial (SMARTGEM) is currently being conducted in Germany to test if the advanced active version of M-sense can confer a further advantage compared to the simple documentation of the basic version [31]. The first results are expected by the end of 2021. Through the Digital Healthcare Act (Digitale-Versorgung-Gesetz), physicians in Germany will be able to prescribe health care apps, and the costs will be reimbursed by statutory health insurance [32]. Therefore, evaluation of app-related health benefits will not only have clinical relevance but also major relevance in health economics.

In conclusion, our study found a decline in the mean values of headache and migraine characteristics after 6 months compared to baseline for people who regularly used an app as a headache diary. The reductions extended to 1 year for those who continued to regularly use the app. These results suggest that the regular use of an app to monitor headache and migraine characteristics may support standard headache care. Headache documentation in an app could help to raise awareness for headache disorders and to identify possible patterns and aggravating factors. It enables patients to take a more active role in managing their headache and feel more in control of their health, resulting in the improvement of headache over time. Based on our findings, the regular use of a headache app could represent an effective measure to complement the therapy of primary headaches.

Conflicts of Interest

BR reports research grants from Novartis, and personal fees from Novartis, Teva, and Allergan. JM reports personal fees from Novartis. SS is the Chief Data Officer of Newsenselab. MAD is the Executive Director and Chief Executive Officer of Newsenselab. TK reports having received honoraria from Eli Lilly, Newsenselab, Total, and The BMJ. UR received honoraria for consulting and lectures from Amgen, Allergan, Abbvie, Eli Lilly, Lundbeck, Novartis Pharma, electroCore, Medscape, Novartis, StreaMedUp, and Teva; UR received research funding from the German Federal Ministry of Education and Research and Novartis Pharma.

LN has contributed to the advisory boards of Hormosan, Lilly, Novartis, and Teva, and received speaking fees from Allergan, Bial, Hormosan, Lilly, Novartis, and Teva. LN received research funding from Deutsche Zentrum für Luft- und Raumfahrt. The other authors have no conflicts of interest to declare.

References

1. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Mar 27:appips201800519. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
2. Stubberud A, Linde M. Digital technology and mobile health in behavioral migraine therapy: a narrative review. Curr Pain Headache Rep 2018 Jul 31;22(10):66. [doi: 10.1007/s11916-018-0718-0] [Medline: 30066141]
3. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [Medline: 12689739]
4. Bandarian-Balooch S, Martin PR, McNally B, Brunelli A, Mackenzie S. Electronic-diary for recording headaches, triggers, and medication use: development and evaluation. Headache 2017 Nov;57(10):1551-1569. [doi: 10.1111/head.13184] [Medline: 28925498]
5. Dodick DW, Tepper SJ, Lipton RB, Buse DC, Stewart WF, Bayliss M, et al. Improving medical communication in migraine management: a modified Delphi study to develop a digital migraine tracker. Headache 2018 Oct;58(9):1358-1372. [doi: 10.1111/head.13426] [Medline: 30362524]
6. Neeb L, Dresler T. Sollten Patienten Kopfschmerz-Apps verwenden? Schmerzmed 2019 Jul 22;35(4):10-11. [doi: 10.1007/s00940-019-1105-6]
7. Dahlem MA, Roesch A, Neeb L, Kropp P, Dresler T. [What do migraine apps offer?]. MMW Fortschr Med 2018 Feb;160(2):51-54. [doi: 10.1007/s15006-018-0153-5] [Medline: 29417508]
8. Göbel H, Frank B, Heinze A, Zimmermann W, Göbel C, Göbel A, et al. [Healthcare behavior of migraine and headache patients when treatment is accompanied by the digital migraine app]. Schmerz 2019 Apr;33(2):147-155. [doi: 10.1007/s00482-018-0355-x] [Medline: 30649625]
9. 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]
10. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
11. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
12. M-sense migräne. M-Sense. URL: https://www.m-sense.de/ [accessed 2020-12-07]
13. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. Cephalalgia 2018 Jan;38(1):1-211. [doi: 10.1177/0333102417738202] [Medline: 29368949]
14. Roesch A, Dahlem MA, Neeb L, Kurth T. Validation of an algorithm for automated classification of migraine and tension-type headache attacks in an electronic headache diary. J Headache Pain 2020 Jun 12;21(1):75 [FREE Full text] [doi: 10.1186/s10194-020-01139-w] [Medline: 32532222]
15. Diener H, Tassorelli C, Dodick DW, Silberstein SD, Lipton RB, Ashina M, International Headache Society Clinical Trials Committee. Guidelines of the International Headache Society for controlled trials of preventive treatment of migraine attacks in episodic migraine in adults. Cephalalgia 2020 Sep;40(10):1026-1044. [doi: 10.1177/0333102420941839] [Medline: 32722936]
16. Penzien DB, Andrasik F, Freidenberg BM, Houle TT, Lake AE, Lipchik GL, et al. Guidelines for trials of behavioral treatments for recurrent headache, first edition: American Headache Society Behavioral Clinical Trials Workgroup. Headache 2005 May;45 Suppl 2:S110-S132. [doi: 10.1111/j.1526-4610.2005.4502004.x] [Medline: 15921503]
17. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-169. [doi: 10.1146/annurev.publhealth.23.100901.140546] [Medline: 11910059]
18. Christiansen S, Jürgens TP, Klinger R. Outpatient combined group and individual cognitive-behavioral treatment for patients with migraine and tension-type headache in a routine clinical setting. Headache 2015 Sep;55(8):1072-1091. [doi: 10.1111/head.12626] [Medline: 26377233]
19. Darabaneanu S, Overath CH, Rubin D, Lüthje S, Sye W, Niederberger U, et al. Aerobic exercise as a therapy option for migraine: a pilot study. Int J Sports Med 2011 Jun;32(6):455-460. [doi: 10.1055/s-0030-1269928] [Medline: 21472632]
20. Zsombok T, Juhasz G, Budavari A, Vitrai J, Bagdy G. Effect of autogenic training on drug consumption in patients with primary headache: an 8-month follow-up study. Headache 2003 Mar;43(3):251-257. [doi: 10.1046/j.1526-4610.2003.03049.x] [Medline: 12603644]
21. Minen MT, Morio K, Schaubhut KB, Powers SW, Lipton RB, Seng E. Focus group findings on the migraine patient experience during research studies and ideas for future investigations. Cephalalgia 2020 Jun;40(7):712-734. [doi: 10.1177/0333102419888230] [Medline: 31870189]

22. Moloney MF, Aycock DM, Cotsonis GA, Myerburg S, Farino C, Lentz M. An Internet-based migraine headache diary: issues in Internet-based research. Headache 2009 May;49(5):673-686. [Medline: 19472446]
23. Krause A, Gieschke R. Interactive visualization and communication for increased impact of pharmacometrics. J Clin Pharmacol 2010 Sep;50(9 Suppl):140S-145S. [doi: 10.1177/0091270010376964] [Medline: 20881227]
24. Fujita M, Fujiwara J. [Validity of graphic headache diary for children having primary headaches]. No To Hattatsu 2011 Nov;43(6):443-447. [Medline: 22180958]
25. Lipton RB, Stewart WF, Liberman JN. Self-awareness of migraine: interpreting the labels that headache sufferers apply to their headaches. Neurology 2002 May 14;58(9 Suppl 6):S21-S26. [doi: 10.1212/wnl.58.9_suppl_6.s21] [Medline: 12011270]
26. Lucas C, Géraud G, Valade D, Chautard M, Lantéri-Minet M. Recognition and therapeutic management of migraine in 2004, in France: results of FRAMIG 3, a French nationwide population-based survey. Headache 2006 May;46(5):715-725. [doi: 10.1111/j.1526-4610.2006.00430.x] [Medline: 16643573]
27. Pavlovic JM, Buse DC, Sollars CM, Haut S, Lipton RB. Trigger factors and premonitory features of migraine attacks: summary of studies. Headache 2014;54(10):1670-1679. [doi: 10.1111/head.12468] [Medline: 25399858]
28. Mosadeghi-Nik M, Askari MS, Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. J Telemed Telecare 2016 Dec;22(8):472-477. [doi: 10.1177/1357633X16673275] [Medline: 27799450]
29. Göbel H, Frank B, Heinze A, Zimmermann W, Göbel C, Göbel A, et al. [Healthcare behavior of migraine and headache patients when treatment is accompanied by the digital migraine app]. Schmerz 2019 Apr;33(2):147-155. [doi: 10.1007/s00482-018-0355-x] [Medline: 30649625]
30. Minen MT, Jalloh A, Ortega E, Powers SW, Sevick MA, Lipton RB. User design and experience preferences in a novel smartphone application for migraine management: a think aloud study of the relaxahead application. Pain Med 2019 Feb 01;20(2):369-377 [FREE Full text] [doi: 10.1093/pm/pny080] [Medline: 29868895]
31. Smartphone-gestützte Migränetherapie als neue Versorgungsform. SMARTGEM. URL: https://smartgem-projekt.de/ [accessed 2021-01-17]
32. Digitale-Versorgung-Gesetzt: Ärzte sollen Apps verschreiben können. Bundesministerium. URL: https://www.bundesgesundheitsministerium.de/digitale-versorgung-gesetz.html [accessed 2020-12-07]

Abbreviations

AMD: days with acute medication use
ANOVA: analysis of variance
ICHD-3: The International Classification of Headache Disorders
MHD: monthly headache days
mHealth: mobile health
MMD: monthly migraine days
TTH: tension-type headache

Edited by G Eysenbach; submitted 10.12.20; peer-reviewed by C Cheng, P Gazerani; comments to author 20.12.20; revised version received 18.01.21; accepted 17.04.21; published 07.07.21.

Please cite as:
Raffaelli B, Mecklenburg J, Overeem LH, Scholler S, Dahlem MA, Kurth T, Oliveira Gonçalves AS, Reuter U, Neeb L
Determining the Evolution of Headache Among Regular Users of a Daily Electronic Diary via a Smartphone App: Observational Study
JMIR Mhealth Uhealth 2021;9(7):e26401
URL: https://mhealth.jmir.org/2021/7/e26401
doi:10.2196/26401
PMID:34255716

Original Paper

Chemotherapy-Induced Peripheral Neuropathy Detection via a Smartphone App: Cross-sectional Pilot Study

Ciao-Sin Chen[1], PharmD; Judith Kim[1], PharmD; Noemi Garg[1], BSc; Harsha Guntupalli[1], PharmD; Reshma Jagsi[2], MD, DPhil; Jennifer J Griggs[3], MD, MPH; Michael Sabel[4], MD; Michael P Dorsch[1], PharmD, MS; Brian C Callaghan[5], MD, MS; Daniel L Hertz[1], PharmD, PhD

[1]Department of Clinical Pharmacy, University of Michigan College of Pharmacy, Ann Arbor, MI, United States

[2]Department of Radiation Oncology, University of Michigan Medical School, Ann Arbor, MI, United States

[3]Department of Internal Medicine, Hematology & Oncology Division, University of Michigan Medical School, Ann Arbor, MI, United States

[4]Department of Surgery, University of Michigan Medical School, Ann Arbor, MI, United States

[5]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Daniel L Hertz, PharmD, PhD
Department of Clinical Pharmacy
University of Michigan College of Pharmacy
428 Church St.
Room 2560C
Ann Arbor, MI, 48109-1065
United States
Phone: 1 7347630015
Fax: 1 734763 4480
Email: dlhertz@med.umich.edu

Abstract

Background: Severe chemotherapy-induced peripheral neuropathy (CIPN) can cause long-term dysfunction of the hands and feet, interfere with activities of daily living, and diminish the quality of life. Monitoring to identify CIPN and adjust treatment before it progressing to a life-altering severity relies on patients self-reporting subjective symptoms to their clinical team. Objective assessment is not a standard component of CIPN monitoring due to the requirement for specially trained health care professionals and equipment. Smartphone apps have the potential to conveniently collect both subjective and objective CIPN data directly from patients, which could improve CIPN monitoring.

Objective: The objective of this cross-sectional pilot study was to assess the feasibility of functional CIPN assessment via a smartphone app in patients with cancer that have received neurotoxic chemotherapy.

Methods: A total of 26 patients who had completed neurotoxic chemotherapy were enrolled and classified as CIPN cases (n=17) or controls (n=9) based on self-report symptoms. All participants completed CIPN assessments within the NeuroDetect app a single time, including patient-reported surveys (CIPN20 [European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-induced Peripheral Neuropathy 20-item scale] and PRO-CTCAE [Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events]) and functional assessments (Gait and Balance and 9-Hole Peg Test). Functional assessment data were decomposed into features. The primary analysis was done to identify features indicative of the difference between CIPN cases and controls using partial least squares analyses. Exploratory analyses were performed to test if any features were associated with specific symptom subtypes or patient-reported survey scores. Patient interviews were also conducted to understand the challenges they experienced with the app.

Results: Comparisons between CIPN cases and controls indicate that CIPN cases had shorter step length (P=.007), unique swaying acceleration patterns during a walking task, and shorter hand moving distance in the dominant hands during a manual dexterity task (variable importance in projection scores ≥2). Exploratory analyses showed similar signatures associated with symptoms subtypes, CIPN20, and PRO-CTCAE. The interview results showed that some patients had difficulties due to technical issues, which indicated a need for additional training or oversight during the initial app download.

Conclusions: Our results supported the feasibility of remote CIPN assessment via a smartphone app and suggested that functional assessments may indicate CIPN manifestations in the hands and feet. Additional work is needed to determine which functional assessments are most indicative of CIPN and could be used for CIPN monitoring within clinical care.

(JMIR Mhealth Uhealth 2021;9(7):e27502) doi:10.2196/27502

KEYWORDS

chemotherapy-induced peripheral neuropathy; smartphone; mobile health; gait; balance; 9-Hole Peg Test

Introduction

Up to 70% of patients receiving neurotoxic chemotherapy experience chemotherapy-induced peripheral neuropathy (CIPN) [1]. CIPN manifests in the feet or hands primarily as numbness or tingling, but it can also have motor or painful components [2,3]. In some patients, CIPN is irreversible, causing long-term interference with balance and dexterity, increased risk of falls [4], and negative effects on quality of life [5].

After neurotoxic chemotherapy, patients self-report symptoms of CIPN during regular appointments with their medical oncology team [6]. Patients who report CIPN symptoms may undergo objective assessment, but this is not standard of care due to the requirement for specially trained health care professionals and equipment [7]. There have been recent efforts to develop validated patient-reported outcome (PRO) questionnaires for CIPN assessment, including the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-induced Peripheral Neuropathy 20-item scale (CIPN20) [8]. PROs are an improvement over routine clinician assessment but are limited by a lack of objective assessment and the complete reliance on patients' willingness and ability to accurately report CIPN [9].

Objective evidence of neuropathy includes gait impairment and increased sway, which can be detected by wearable sensors [10,11]. Another common manifestation is reduced manual strength and dexterity, which can be assessed using functional testing with a 9-Hole Peg [12]. Collecting PRO during chemotherapy via a smartphone app has been integrated into clinical practice to monitor and reduce severe toxicity [13]. Smartphone apps that integrate PRO and objective assessment could improve CIPN detection with minimal inconvenience and cost to the patient or health care system [14]. The objectives of this pilot study were to assess the feasibility of patients downloading and completing objective CIPN assessments remotely within a smartphone app and to explore whether there were differences in the objective functional data between patients with and without CIPN.

Methods

Study Patients

Patients with cancer aged over 18 who had completed neurotoxic chemotherapy including a taxane, platinum, or vinca alkaloid and had access to an iPhone were enrolled in this cross-sectional study. All patients completed written informed consent. This study was approved by the University of Michigan Institutional Review Board (IRBMed).

Completion of NeuroDetect

All enrolled patients were emailed instructions on how to download NeuroDetect version 1.0 (available in the iOS app store) and were asked to complete all data collection assessments within the app a single time. Data collection included a baseline demographic survey, patient-reported CIPN assessment, and 2 functional CIPN assessments (the Gait and Balance and 9-Hole Peg Tests) available within Apple's ResearchKit.

Baseline Demographics

Patients self-reported standard demographics and information about their cancer, including date of diagnosis, cancer type, chemotherapy agents, whether they had surgery, surgery date, and if they were taking pain medication.

Patient-Reported CIPN

The CIPN20 is considered the gold standard for CIPN assessment [15]. CIPN20 asks patients to indicate the degree of 20 CIPN symptoms they experienced within the past week on a numerical scale from 1 to 4 (1=not at all, 2=a little, 3=quite a bit, and 4=very much). CIPN20 includes 9 sensory items, 8 motor items, and 3 autonomic items [8]. In exploratory analyses, to test if any features that are specifically associated with CIPN cases having specific types of symptoms, we created subscales including the first 8 sensory symptoms items (CIPN8) [16], the first 4 sensory symptoms items (CIPN4), the first 2 feet items (CIPN2Feet), the first 2 hands items (CIPN2Hands), and all 8 motor items (CIPNM) as summary indicators in this exploratory study. CIPN20 scores and all subscales were linearly converted to 0-100 [8].

In addition, patient-reported CIPN was collected via PRO-CTCAE, which collects the severity and interference of neuropathy symptoms [17]. Patients also completed a falls diary, which asked whether or not the patient has fallen since the last chemotherapy session [18,19].

Functional CIPN Analysis

In the Gait and Balance test, patients were asked to walk 7 steps in a straight line, turn around, and walk back to where they started with their smartphone in a pocket or bag. Walking forward and walking back were treated as 2 replications. User data were collected by the inertial accelerometer and pedometer and processed by iOS's sensor fusion algorithm [20]. Vertical acceleration data were obtained by rotating user acceleration with orientation information. Horizontal acceleration data were obtained by assuming the axis with a larger SD of the velocity, which was calculated as the sum of the integral of acceleration, was the moving direction of the patient. Horizontal acceleration data from patients who did not follow instructions to put their phone in their pocket and instead held their phone in their hands

were removed from the analysis. Data from phones that were moving within the pocket or bag during the test were also removed from the analysis due to the acceleration noise introduced by this movement. Data duration shorter than 3.50 seconds was considered as early termination of the Gait and Balance test and removed from the analysis. Step length was calculated by the distance and step count collected by the pedometer. A total of 87 gait features were generated by an open-source tool, mhealthtools [21], which decomposes the acceleration data into more sophisticated summary indicators using Fourier Transforms.

In the 9-Hole Peg Test, patients were asked to place a virtual peg in a virtual hole and then remove the virtual peg from the virtual hole as quickly as possible on their phone screen. For patient convenience, this task was repeated 4 times instead of the typical 9 times. The time required to complete the assessment and the total distance the peg was moved during the assessment were collected. Hand speed was calculated by dividing the distance by the time. Sixty hand features were generated, including statistical summaries such as means, medians, minimums, and maximums of dominant and subordinate hand movement of placing and removing the peg.

Structured Interviews

After completion, the patients were interviewed about issues they experienced with the app. The structured interview (Table 1) included 4 topics: enrollment and app downloading, informed consent forms, demographics and PRO questionnaires, and functional assessments.

Table 1. The structured interview guides.

Topics	Questions
Enrollment and app downloading	1. What challenges, if any, did you experience in enrolling in the study?
Informed consent forms	2. Did you experience any difficulty in reading and understanding the informed consent forms?
Demographics and patient-reported outcome questionnaires	3. Which study questions, if any, were difficult to answer? Why?
Functional assessments	4. Which functional assessments, if any, were difficult or uncomfortable to perform? Why?

Statistical Analysis

Patients were classified as CIPN cases if they reported “quite a bit” or “very much” for at least one of the first 4 questions of the CIPN20 questionnaire [8]. The first 4 questions ask about numbness or tingling in the feet or hands, which are the most common symptoms of CIPN [22,23]. This shortened version has been used to screen long-term CIPN in a large clinical trial [5]. Patients that did not complete CIPN20 within NeuroDetect were classified based on their answers during screening. To test the feasibility of functional assessments via a smartphone app for detecting CIPN, task features available in at least 70% of patients were compared between CIPN case and control groups. Additionally, alternative CIPN case classifications were explored by classifying patients as CIPN feet cases if they responded 3 or more on either of the 2 sensory CIPN feet questions within the CIPN2Feet. A similar approach was used to classify patients as CIPN hands cases. In all instances, any patients not included in the CIPN case group were included in the control group for comparisons. Exploratory analyses of secondary endpoints of cumulative CIPN20 score and PRO-CTCAE severity and interference were also conducted. Principal component analysis (PCA) was conducted using the prcomp function with scaling. Partial least squares discriminant analysis (PLSDA) and PLS regression analysis were conducted using the plsr function in pls package with the orthogonal scoring method and scaling. Features from PLS models with loadings ≥0.2 or ≤–0.2 or variable importance in projection (VIP) score ≥2 were considered important. Predictive performance of PLS models were evaluated with leave-one-out validation. Important features were tested with unpaired 2-sample *t* tests with uncorrected α=.05. Differences in demographics and patient-reported CIPN between CIPN cases and controls were examined by unpaired 2-sample *t* tests and chi-square tests with α=.05. All statistical analysis and visualization were done in R 3.6.3 (R Foundation for Statistical Computing).

Results

Enrolled Patients

A total of 26 patients who had completed neurotoxic chemotherapy enrolled and participated in the NeuroDetect assessments. Patients were classified based on the 4 CIPN20 questions. There were 2 patients that did not complete CIPN20, and they were classified as CIPN cases based on their answers during screening. A total of 17 CIPN cases and 9 controls were enrolled. CIPN cases were nominally older (54 vs 49) and completed chemotherapy more recently (2.53 vs 8.27 months), but none of the clinical factors were significantly different between groups (Table 2).

Table 2. Clinical data of patients included in the analysis (no variables were significantly different between the 2 groups).

Demographics	Controls (n=9)	Cases (n=17)	*P* value
Age (years), mean (SD)	49 (12.4)	54 (8.2)	.96
Race, n (%)			.66
White	8 (89)	14 (82)	
Other	1 (11)	3 (18)	
Height (m), mean (SD)	1.78 (0.14)	1.70 (0.10)	.81
Weight (kg), mean (SD)	83.1 (23.6)	75.3 (16.6)	.27
Cancer type, n (%)			.59[a]
Breast	4 (44)	5 (29)	
Ovarian	0 (0)	2 (12)	
Lung	0 (0)	1 (6)	
Colorectal	2 (22)	6 (35)	
Other[b]	4 (44)	8 (47)	
Neurotoxic chemotherapy agent , n (%)			.60
Taxane	5 (56)	6 (35)	
Platinum	2 (22)	5 (29)	
Taxane and platinum	2 (22)	6 (35)	
Time since treatment completion (months), mean (SD)	8.27 (16.7)	2.53 (2.70)	.08

[a]Chi-square test between chemotherapy-induced peripheral neuropathy cases and controls that had breast cancer, ovarian cancer, lung cancer, or colorectal cancer versus only other cancer types.

[b]Other cancer types include liver cancer, esophageal cancer, prostate cancer, cervical cancer, and pancreatic cancer.

Patient-Reported CIPN

The median CIPN20 scores were higher in CIPN cases than in controls (28.1 vs 10.0, *P*<.001; Table 3), which was implied by the grouping criteria. All of the CIPN20 subscale scores were also higher in CIPN cases, including CIPN8 (*P*<.001), CIPN4 (*P*<.001), CIPN2Feet (*P*<.001), CIPN2Hands (*P*=.004), and CIPNM (*P*=.02). The median PRO-CTCAE severity and interference were higher in cases (2 vs 3, *P*<.001 and 1 vs 2, *P*<.001, respectively). Only 1 CIPN case reported having fallen since the last chemotherapy session.

Table 3. Patient-reported CIPN.

Scale	Controls (n=9)	Cases (n=17)	Feet cases (n=13)	Hands cases (n=9)
CIPN20[a], median (SD)[b]				
CIPN20	10.0 (8.8)	28.1 (17.0)	29.8 (16.5)	43.3 (18.6)
CIPN8 (Sensory)	10.4 (12.1)	35.4 (21.3)	37.5 (20.8)	58.3 (23.0)
CIPNM (Motor)	8.3 (11.0)	20.9 (19.4)	25.0 (19.5)	33.3 (20.5)
PRO-CTCAE[c], median (SD)[d]				
Severity	2 (0.7)	3 (0.8)	3 (0.8)	3 (0.8)
Interference	1 (0.3)	2 (0.8)	2 (0.8)	2 (1.0)
Patients reporting a fall, n (%)	0 (0)	1 (6)	1 (8)	0 (0)

[a]CIPN20: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-induced Peripheral Neuropathy 20-item scale.

[b]There were 2 cases that did not complete CIPN20.

[c]PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

[d]There were 3 cases that did not complete PRO-CTCAE.

Functional CIPN Analysis

Twelve patients gave NeuroDetect permission to collect pedometer data during the Gait and Balance test (Figure 1). The median step lengths were shorter in cases compared with controls (0.54 vs 0.78 m; *P*=.007; Figure 2).

Figure 1. Sample size of each type of data. Excluded acceleration data were from 5 CIPN cases and 2 controls.

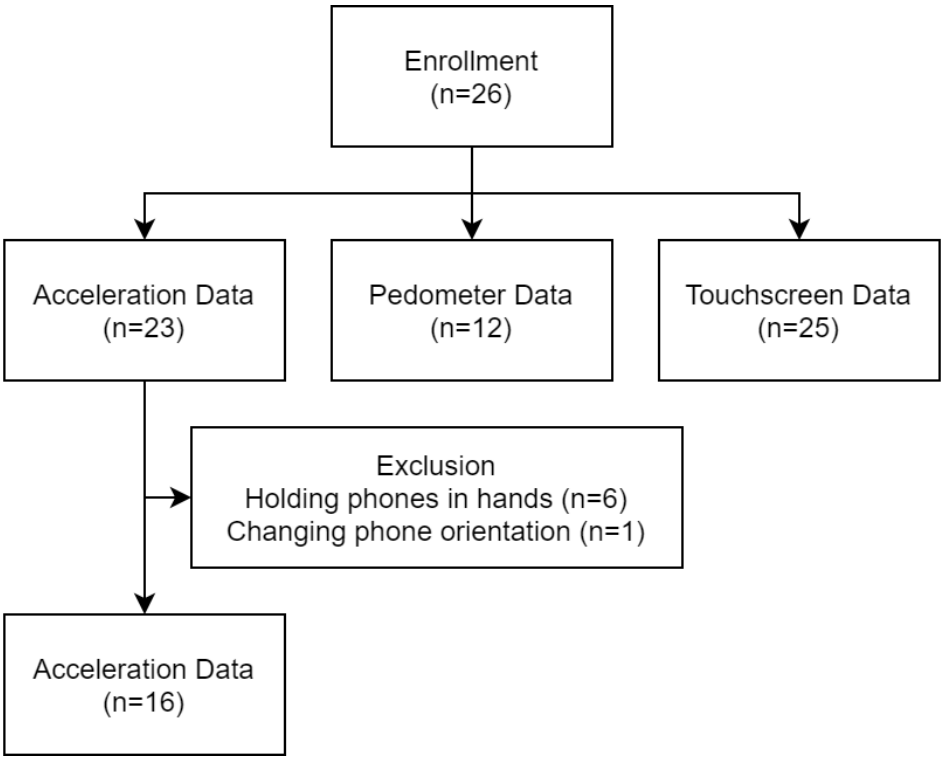
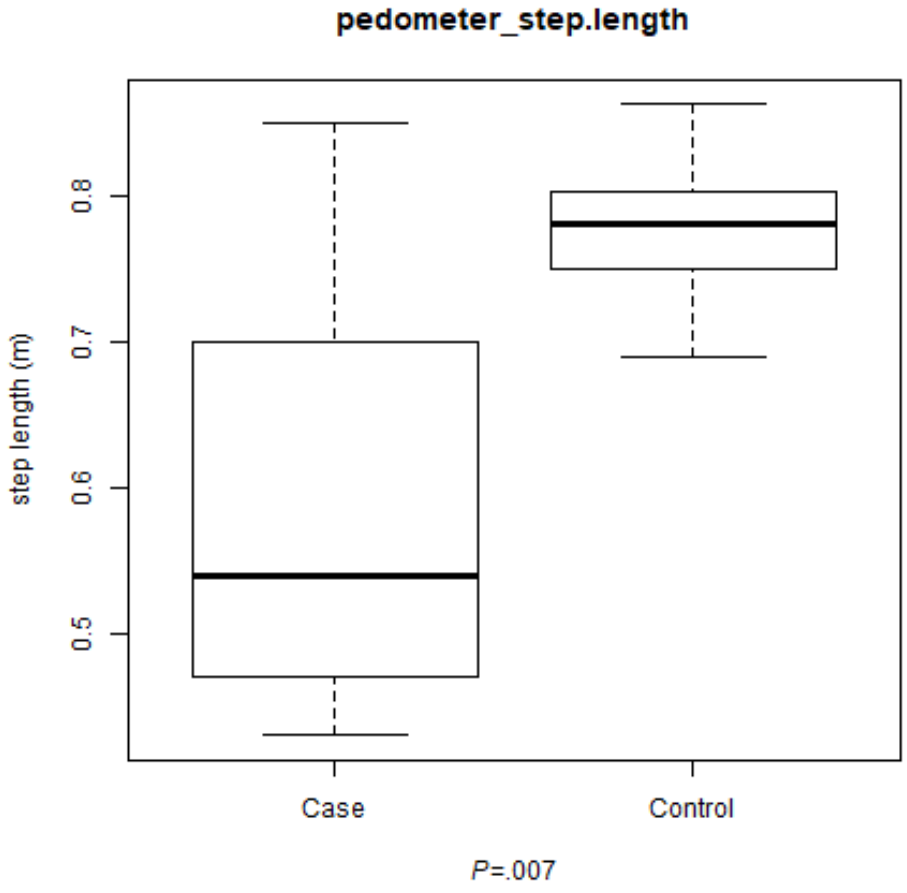
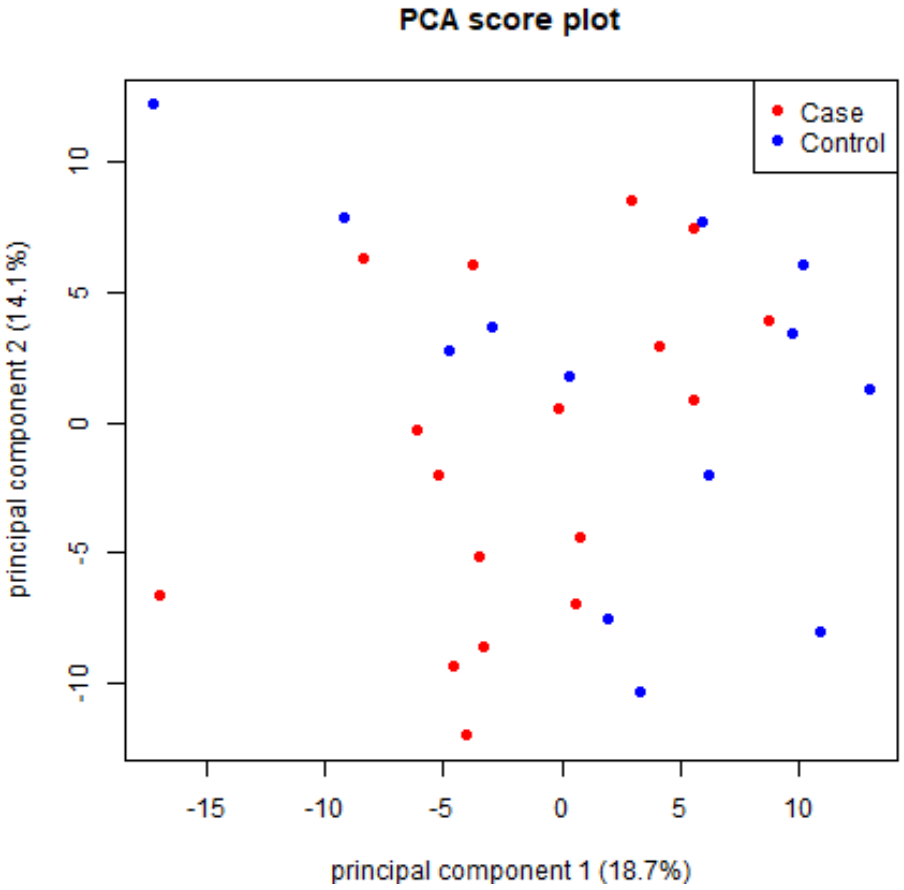


Figure 2. Step length was calculated by the distance and step count collected by the pedometer. The median step lengths were shorter in CIPN cases than controls (0.54 vs 0.78 m, *P*=.007).



Acceleration data from 7 patients (Figure 1) were removed from the PCA and PLS analyses due to holding the phone in their hands (n=6) or phone changing orientation in the pocket during the test (n=1). Two patients completed only 1 walk of the Gait and Balance test, so only the completed walk was used in the analysis. Using PCA to integrate mhealthtools gait features and hand features of 16 patients, the first 2 components explained 32.8% of the variance between cases and controls (Figure 3), and each of the remaining principal components explained ≤10.0% of the variance. However, no individual features showed high importance (all loadings between –0.2 and 0.2) in the first 2 principal components.

Figure 3. Score plot of PCA integrating mhealthtools gait features and hand features. The first 2 components explained 32.8% of the variance between CIPN cases and controls. However, no individual features showed high importance in the first 2 components (all loadings between -0.2 and 0.2). CIPN: chemotherapy-induced peripheral neuropathy; PCA: principal component analysis.



In PLSDA, the first 2 components achieved good separation between CIPN cases and controls (Figure 4). A total of 145 features had VIP scores ≥1 (Multimedia Appendix 1), and 12 features had VIP scores ≥2 (Figure 5). These 12 features were all significant in unpaired 2-sample *t* tests. They were 5 gait features in the swaying axis and 7 hand features. In terms of gait features of frequency decomposition in the swaying axis, CIPN cases had higher Shannon entropy (*P*=.01), higher median frequency (*P*=.009), lower maximum frequency (*P*=.007), lower skewness (*P*=.004), and lower kurtosis (*P*=.007). In terms of hand features, CIPN cases had lower mean (*P*=.008), median (*P*=.01), maximum (*P*=.001), and SD (*P*<.001) of the distance, and lower maximum (*P*=.02) and SD (*P*=.01) of the speed of the peg removing movement in the dominant hands, and lower peg removing speed SD in the subordinate hands (*P*=.006). The first 6 components explained 52.4% of the variance between groups, while each of the remaining components explained ≤6.1% of the variance. The leave-one-out validation of the first 6 components in the PLSDA model showed R^2=0.590 and RMSE=0.312 (Multimedia Appendix 2).

Figure 4. Score plot of partial least squares discriminant analysis integrating mhealthtools gait features and hand features. The first 2 components achieved good separation between CIPN cases and controls and explained 23.0% of the variance between groups. In the first 2 components, 145 features had variable importance in projection scores ≥1, including 12 with scores ≥2 (see Figure 5). All loadings between -0.2 and 0.2. CIPN: chemotherapy-induced peripheral neuropathy; PLS: partial least squares.

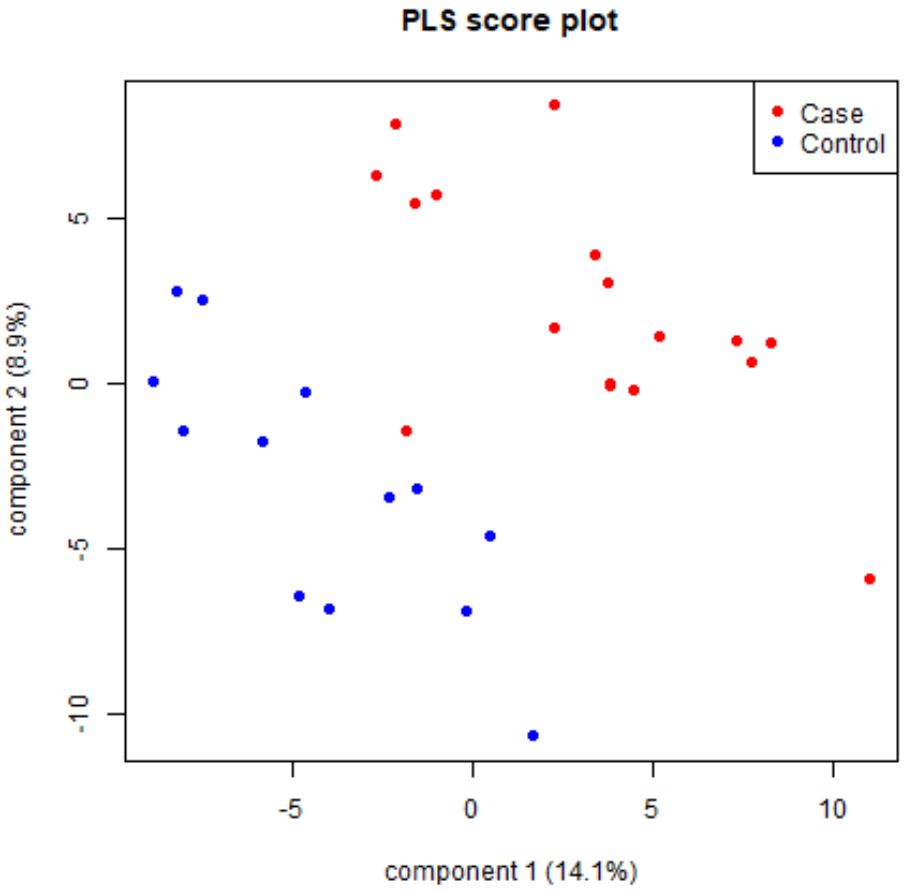
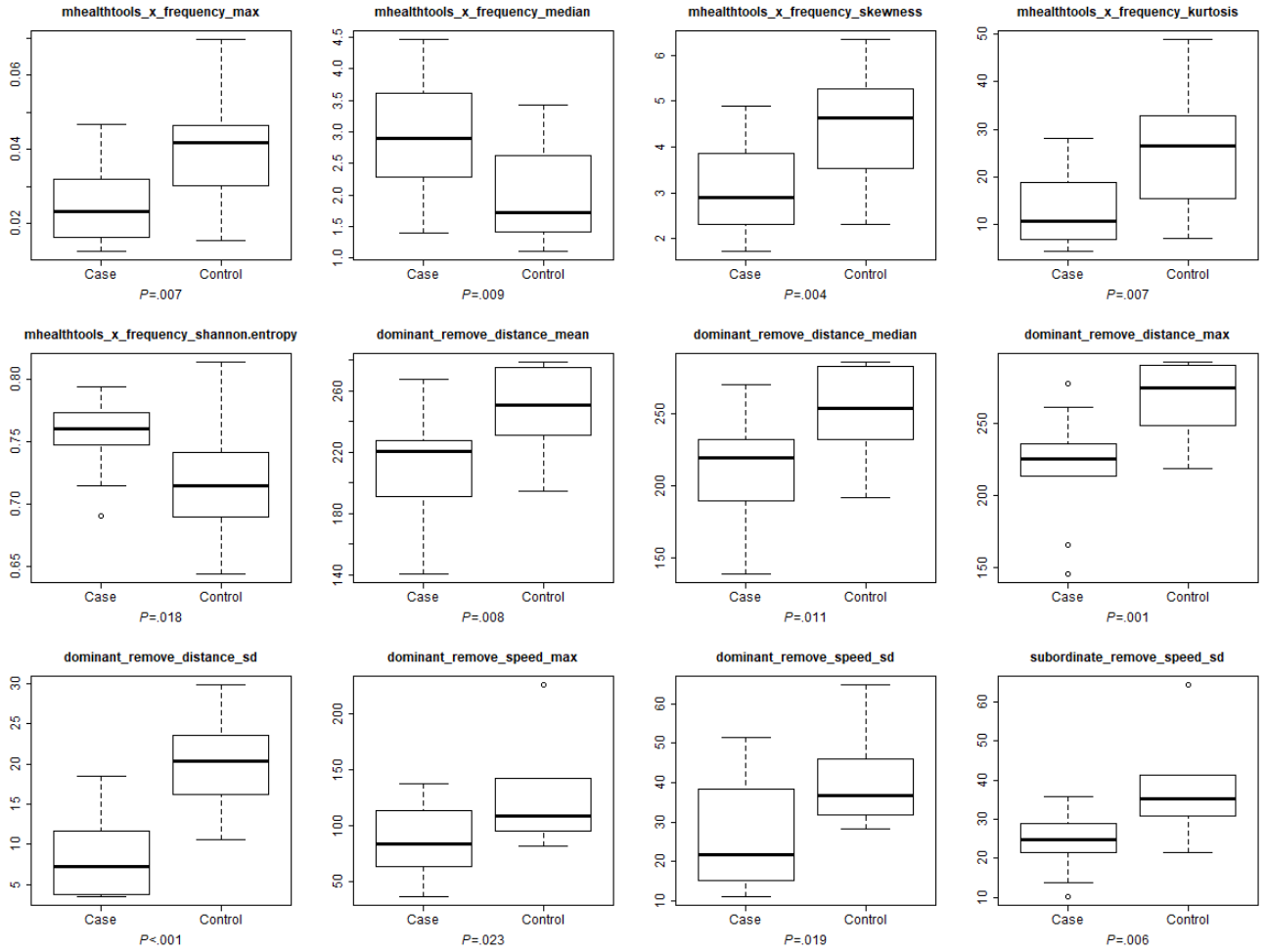


Figure 5. A total of 12 important features with variable importance in projection scores ≥2 in the first 2 components of partial least squares discriminant analysis. The *P* values annotated in each subplot were from unpaired two-sample t tests. The x-axis is the swaying axis. CIPN cases had higher Shannon entropy (*P*=.02), higher median frequency (*P*=.009), lower maximum frequency (*P*=.007), lower skewness (*P*=.004), and lower kurtosis (*P*=.007) in the swaying axis. CIPN cases also had lower mean (*P*=.008), median (*P*=.01), maximum (*P*=.001), and SD (*P*<.001) of the distance and lower maximum (*P*=.02) and SD (*P*=.02) of the speed of the peg removing movement in the dominant hands, and lower peg removing speed SD (*P*=.006) in the subordinate hands. CIPN: chemotherapy-induced peripheral neuropathy.



Exploratory PLSDA classifying patients as CIPN feet cases identified the same set of features, and the model of CIPN hands cases identified a similar set of the peg removing distance features in the dominant hands and additional gait features with high VIP scores (VIP>2; additional gait features are shown in Multimedia Appendix 3).

Exploratory PLS regression analyses of the secondary endpoints of cumulative CIPN20 score and PRO-CTCAE severity and interference identified several of the same features that were identified in the prior case–control analyses, including gait features in the swaying axis and hand features of the peg removing movement in the dominant hands (additional features are shown in Multimedia Appendix 4).

Structured Interviews

Among the 26 patients, 12 cases and 7 controls participated in the interviews. Of these, 5 patients (26%) experienced challenges during the enrollment due to the password requirement. Two (11%) had problems reading the consent form and preferred the format to be a continuous scroll, while it was a booklet format in the app where a user clicks to navigate to the next page. A total of 11 patients (58%) had issues with the questionnaires, and among them, 8 patients wished they had an option to go back and change their answers to the baseline demographic questions, and 3 patients found it difficult to measure their neuropathy symptoms in the scale from 1 (not at all) to 4 (very much). Four (21%) had issues with the Gait and Balance test because they were too tired or did not feel comfortable performing the test. Four patients (21%) had difficulty understanding the 9-Hole Peg Test.

Discussion

Principal Findings

CIPN is a common and debilitating toxicity of chemotherapy that has both subjective and objective components. Smartphone apps are being used to collect subjective data via PRO and may also be a convenient platform for integrating objective CIPN assessments. Our pilot study supported the feasibility of downloading and completing remote assessments within a smartphone app and explored whether these functional assessments could differentiate patients with and without CIPN. We found evidence that patients with CIPN have shorter step length, unique swaying acceleration patterns during a walking task, and shorter hand moving distance in the dominant hands

during a manual dexterity task. Our results indicate that functional assessments of feet and hands via a smartphone app may be helpful in detecting CIPN during chemotherapy treatment, but future studies are needed.

Feasibility and Adherence

As described above, among the patients enrolled in this study, 2 cases did not complete CIPN20, and 3 cases did not complete PRO-CTCAE (Table 2). Over half did not allow the collection of pedometer data, and 7 did not follow instructions to place their phone in their pocket during the walking task (Figure 1). These nonadherence issues led to some loss of data. The interview results showed that some patients had difficulties using the app due to the password requirements, demographics questionnaire, and functional assessments, which indicated a need for additional training or oversight during the initial app download.

Comparison With Prior Work

Shorter step length, slower gait speed, and loss of balance are the most common CIPN symptoms, and these symptoms progress throughout treatments [4,24,25]. In patients with CIPN, wearable sensors can detect impairment in stride length, stride duration, and gait speed during a natural walk, and an increase in sway when standing [10,11]. However, a prior study in patients that self-reported peripheral neuropathy symptoms caused by exposure to chemotherapy found no correlation between CIPN severity and a single posttreatment assessment of gait and balance, including stride time, cadence, speed, or sway [26]. Our exploratory study found that CIPN cases had shorter step lengths. Using the frequency decomposition methods in mhealthtools, we also found that the frequency spectrum of CIPN cases acceleration data had higher median frequency and lower maximum frequency and were more symmetrical and less predictable. These differences in frequency distribution could indicate CIPN symptoms, even if they do not have an intuitive interpretation. Our seemingly discordant findings may be due to advances in collecting and analyzing gait and balance data via the smartphone app. Additional work is needed to identify the specific gait features most strongly indicative of CIPN, similar to the development of algorithms for detecting Parkinson disease using mPower data [21,27,28]. Future work should also consider alternatives to a natural walk test, including balance assessments commonly used during clinical neurological assessments such as walking heel-to-toe (tandem walk) [29] or balancing with your eyes closed (Romberg test) [30].

In the 9-Hole Peg Test, we found that CIPN cases had shorter distance and lower distance SD of the peg removing movement in the dominant hands. We also found slower maximum peg removing speed in the dominant hands and lower peg removing speed SD in both hands. Motor weakness is a common manifestation of CIPN [2,3]. A hand kinematic analysis in patients that reported numbness with or without neuropathic pain due to chemotherapy used 3D recordings of the hand grip-release test and revealed CIPN cases had more jerks in grasp movement but not in reach movement [31]. It would be difficult to adapt this test into a smartphone app–based functional assessment, but there may be opportunities for other manual strength and dexterity tasks such as finger tapping [32,33] or line tracing [34].

Impact on Patient Care

Our pilot study indicates that a smartphone app–based functional assessment may be helpful in detecting CIPN and could possibly be useful for CIPN monitoring during neurotoxic chemotherapy treatment. Because CIPN can manifest in the feet or hands [2,3], our smartphone app includes both gait and hand assessments. CIPN monitoring via a smartphone would circumvent several challenges to objective CIPN assessment. Patients could complete assessments whenever and wherever is most comfortable and convenient for them. Collection via a smartphone does not require any trained personnel or specialized equipment other than a free app, eliminating most costs associated with CIPN monitoring. The seamless integration of subjective and objective CIPN data collection within a smartphone app may improve early CIPN detection, allowing for early initiation of treatment modification to prevent irreversible, life-altering toxicity [35].

Limitations

This is the first study that identified smartphone sensor gait features associated with CIPN, and this is also the first study using a smartphone app–based functional assessment for hand symptoms in CIPN. Despite these strengths, this study has several limitations that are worth considering. First, this was a small study of a heterogeneous cohort of patients and a large number of variables. Because of the exploratory nature of this study, we allowed a higher false-positive rate and did not correct for multiple comparisons in unpaired 2-sample *t* tests. The 2 smartphone app–based functional assessments and the open-source data decomposition tool have not been tested in patients with CIPN. Larger studies are needed to validate the smartphone app–based tests and features indicative of CIPN and adjust for confounders such as comorbidities. Second, another major limitation was the assessment at a single time point after completion of chemotherapy. There may be differences between patient's willingness or ability to complete the app-based assessment after treatment versus during treatment. Our subsequent study will collect assessments before and during treatment to determine the feasibility of assessment during treatment and to better understand the functional changes caused by CIPN. Third, the smartphone app–based functional assessments were affected by users' consent to data collection and adherence to the app instruction. In future studies, we will need more specific instructions during the initial app downloading to ensure that valid data are collected from all users. Finally, smartphone app–based gait analyses are more reliable if the phone is placed at the lumbar location [27]. Our data may have more variability caused by the placement of devices in a pocket or bag, but this is a trade-off we have made to develop an app that requires no external equipment, including a harness or wearable sensors.

Conclusions

Our findings suggest that smartphone app–based functional assessments may be useful to detect functional impairment indicative of CIPN. Future work will conduct longitudinal

assessments in patients undergoing neurotoxic chemotherapy using a second-generation app that includes additional functional assessments to identify features that are most strongly associated with CIPN and determine whether functional assessment detects CIPN prior to a patient's subjective assessment. Upon validation, we may be able to integrate convenient, low-cost smartphone-based CIPN monitoring into chemotherapy treatment to improve detection and prevent irreversible and life-altering CIPN.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variable importance in projection score distribution of the first 2 components of partial least squares discriminant analysis. A total of 145 features had variable importance in projection scores ≥1. Among them, 12 had variable importance in projection scores ≥2 (see <xref ref-type="fig" rid="6figure5">Figure 5</xref>).
[PNG File , 7 KB - mhealth_v9i7e27502_app1.png]

Multimedia Appendix 2
Leave-one-out validation of the first 6 components in partial least squares discriminant analysis. R2 = 0.590 and RMSE = 0.312.
[PNG File , 6 KB - mhealth_v9i7e27502_app2.png]

Multimedia Appendix 3
A total of 11 additional gait features with variable importance in projection scores ≥2 were identified in the first 2 components of the exploratory partial least squares discriminant analysis of CIPN hand cases. The <italic>P</italic> values annotated in each subplot were from unpaired two-sample t test. The x-axis is the swaying axis, and the z-axis is the vertical axis. In the vertical axis, CIPN hand cases had higher maximum frequency (<italic>P</italic>=.044), skewness (<italic>P</italic>=.028), kurtosis (<italic>P</italic>=.043), and lower energies in lower frequency bands (band 6 <italic>P</italic>=.014, band 6.5 <italic>P</italic>=.005, band 7 <italic>P</italic>=.007) of frequency decomposition and lower entropies of empirical wavelet transform (Shannon entropy <italic>P</italic>=.049, permutation entropy <italic>P</italic>=.049, Renyi entropy <italic>P</italic>=.029).
[PNG File , 18 KB - mhealth_v9i7e27502_app3.png]

Multimedia Appendix 4
A total of 11 additional features with variable importance in projection scores ≥2 were identified in the first 2 components of the exploratory partial least squares regression analysis of CIPN20 and PRO-CTCAE severity and interference. The <italic>P</italic> values annotated in each subplot were from unpaired two-sample t test. The x-axis is the swaying axis, and the z-axis is the vertical axis. Among the gait features of frequency decomposition, higher energy in a lower frequency band in the swaying axis (band 4.5 <italic>P</italic>=.049) and higher energies in medium frequency bands (band 12 <italic>P</italic>=.047 and band 12.5 <italic>P</italic>=.037) were associated with worse CIPN scores. Among the hand features, the maximum peg removing speed in the subordinate hands were associated with worse CIPN scores.
[PNG File , 19 KB - mhealth_v9i7e27502_app4.png]

References

1. Seretny M, Currie GL, Sena ES, Ramnarine S, Grant R, MacLeod MR, et al. Incidence, prevalence, and predictors of chemotherapy-induced peripheral neuropathy: A systematic review and meta-analysis. Pain 2014 Dec;155(12):2461-2470. [doi: 10.1016/j.pain.2014.09.020] [Medline: 25261162]
2. Miltenburg NC, Boogerd W. Chemotherapy-induced neuropathy: A comprehensive survey. Cancer Treat Rev 2014 Aug;40(7):872-882. [doi: 10.1016/j.ctrv.2014.04.004] [Medline: 24830939]
3. Quasthoff S, Hartung HP. Chemotherapy-induced peripheral neuropathy. J Neurol 2002 Jan;249(1):9-17. [doi: 10.1007/pl00007853] [Medline: 11954874]
4. Winters-Stone KM, Horak F, Jacobs PG, Trubowitz P, Dieckmann NF, Stoyles S, et al. Falls, Functioning, and Disability Among Women With Persistent Symptoms of Chemotherapy-Induced Peripheral Neuropathy. J Clin Oncol 2017 Aug 10;35(23):2604-2612 [FREE Full text] [doi: 10.1200/JCO.2016.71.3552] [Medline: 28586243]
5. Bandos H, Melnikow J, Rivera DR, Swain SM, Sturtz K, Fehrenbacher L, et al. Long-term Peripheral Neuropathy in Breast Cancer Patients Treated With Adjuvant Chemotherapy: NRG Oncology/NSABP B-30. J Natl Cancer Inst 2018 Feb 01;110(2):djx162 [FREE Full text] [doi: 10.1093/jnci/djx162] [Medline: 28954297]

6. Knoerl R, Smith EML, Han A, Doe A, Scott K, Berry DL. Characterizing patient-clinician chemotherapy-induced peripheral neuropathy assessment and management communication approaches. Patient Educ Couns 2019 Sep;102(9):1636-1643 [FREE Full text] [doi: 10.1016/j.pec.2019.04.012] [Medline: 31003878]
7. Park SB, Goldstein D, Krishnan AV, Lin CS, Friedlander ML, Cassidy J, et al. Chemotherapy-induced peripheral neurotoxicity: a critical analysis. CA Cancer J Clin 2013;63(6):419-437 [FREE Full text] [doi: 10.3322/caac.21204] [Medline: 24590861]
8. Cavaletti G, Cornblath DR, Merkies ISJ, Postma TJ, Rossi E, Frigeni B, CI-PeriNomS Group, et al. The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. Ann Oncol 2013 Feb;24(2):454-462 [FREE Full text] [doi: 10.1093/annonc/mds329] [Medline: 22910842]
9. Salgado TM, Quinn CS, Krumbach EK, Wenceslao I, Gonzalez M, Reed HL, et al. Reporting of paclitaxel-induced peripheral neuropathy symptoms to clinicians among women with breast cancer: a qualitative study. Support Care Cancer 2020 Sep;28(9):4163-4172. [doi: 10.1007/s00520-019-05254-6] [Medline: 31897779]
10. Zahiri M, Chen KM, Zhou H, Nguyen H, Workeneh BT, Yellapragada SV, et al. Using wearables to screen motor performance deterioration because of cancer and chemotherapy-induced peripheral neuropathy (CIPN) in adults - Toward an early diagnosis of CIPN. J Geriatr Oncol 2019 Nov;10(6):960-967 [FREE Full text] [doi: 10.1016/j.jgo.2019.01.010] [Medline: 30665876]
11. Fino PC, Horak FB, El-Gohary M, Guidarelli C, Medysky ME, Nagle SJ, et al. Postural sway, falls, and self-reported neuropathy in aging female cancer survivors. Gait Posture 2019 Mar;69:136-142 [FREE Full text] [doi: 10.1016/j.gaitpost.2019.01.025] [Medline: 30716669]
12. Kobesova A, Nyvltova M, Kraus J, Kolar P, Sardina A, Mazanec R, et al. Evaluation of muscle strength and manual dexterity in patients with Charcot-Marie-Tooth disease. J Hand Ther 2016;29(1):66-72; quiz 72. [doi: 10.1016/j.jht.2015.12.002] [Medline: 26847322]
13. Weaver A, Love SB, Larsen M, Shanyinde M, Waters R, Grainger L, et al. A pilot study: dose adaptation of capecitabine using mobile phone toxicity monitoring - supporting patients in their homes. Support Care Cancer 2014 Oct;22(10):2677-2685. [doi: 10.1007/s00520-014-2224-1] [Medline: 24771299]
14. Campbell G, Skubic MA. Balance and Gait Impairment: Sensor-Based Assessment for Patients With Peripheral Neuropathy. Clin J Oncol Nurs 2018 Jun 01;22(3):316-325. [doi: 10.1188/18.CJON.316-325] [Medline: 29781455]
15. Dorsey SG, Kleckner IR, Barton D, Mustian K, O'Mara A, St Germain D, et al. The National Cancer Institute Clinical Trials Planning Meeting for Prevention and Treatment of Chemotherapy-Induced Peripheral Neuropathy. J Natl Cancer Inst 2019 Jun 01;111(6):531-537 [FREE Full text] [doi: 10.1093/jnci/djz011] [Medline: 30715378]
16. Hertz DL, Kidwell KM, Vangipuram K, Li F, Pai MP, Burness M, et al. Paclitaxel Plasma Concentration after the First Infusion Predicts Treatment-Limiting Peripheral Neuropathy. Clin Cancer Res 2018 Apr 27;24(15):3602-3610. [doi: 10.1158/1078-0432.ccr-18-0656]
17. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, National Cancer Institute PRO-CTCAE Study Group. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol 2015 Nov;1(8):1051-1059 [FREE Full text] [doi: 10.1001/jamaoncol.2015.2639] [Medline: 26270597]
18. Hannan MT, Gagnon MM, Aneja J, Jones RN, Cupples LA, Lipsitz LA, et al. Optimizing the tracking of falls in studies of older participants: comparison of quarterly telephone recall with monthly falls calendars in the MOBILIZE Boston Study. Am J Epidemiol 2010 May 01;171(9):1031-1036 [FREE Full text] [doi: 10.1093/aje/kwq024] [Medline: 20360242]
19. Ashburn A, Stack E, Ballinger C, Fazakarley L, Fitton C. The circumstances of falls among people with Parkinson's disease and the use of Falls Diaries to facilitate reporting. Disabil Rehabil 2008;30(16):1205-1212. [doi: 10.1080/09638280701828930] [Medline: 18608387]
20. Apple Inc.. CMMotionManager. 2020. URL: https://developer.apple.com/documentation/coremotion/cmmotionmanager [accessed 2021-05-17]
21. Snyder P, Tummalacherla M, Perumal T, Omberg L. mhealthtools: A Modular R Package for Extracting Features from Mobile and Wearable Sensor Data. JOSS 2020 Mar;5(47):2106. [doi: 10.21105/joss.02106]
22. Huang HQ, Brady MF, Cella D, Fleming G. Validation and reduction of FACT/GOG-Ntx subscale for platinum/paclitaxel-induced neurologic symptoms: a gynecologic oncology group study. Int J Gynecol Cancer 2007;17(2):387-393. [doi: 10.1111/j.1525-1438.2007.00794.x] [Medline: 17362317]
23. Wolf SL, Barton DL, Qin R, Wos EJ, Sloan JA, Liu H, et al. The relationship between numbness, tingling, and shooting/burning pain in patients with chemotherapy-induced peripheral neuropathy (CIPN) as measured by the EORTC QLQ-CIPN20 instrument, N06CA. Support Care Cancer 2012 Mar;20(3):625-632 [FREE Full text] [doi: 10.1007/s00520-011-1141-9] [Medline: 21479990]
24. Varedi M, Lu L, Howell CR, Partin RE, Hudson MM, Pui C, et al. Peripheral Neuropathy, Sensory Processing, and Balance in Survivors of Acute Lymphoblastic Leukemia. J Clin Oncol 2018 Aug 01;36(22):2315-2322 [FREE Full text] [doi: 10.1200/JCO.2017.76.7871] [Medline: 29812998]

25. Monfort SM, Pan X, Patrick R, Ramaswamy B, Wesolowski R, Naughton MJ, et al. Gait, balance, and patient-reported outcomes during taxane-based chemotherapy in early-stage breast cancer patients. Breast Cancer Res Treat 2017 Jul;164(1):69-77 [FREE Full text] [doi: 10.1007/s10549-017-4230-8] [Medline: 28374323]
26. Jacobs P, Hanaway P, Leitschuh J, Condon J, Rajhbeharrysingh U, Mosquera-Lopez C, et al. Design and evaluation of a portable smart-phone based peripheral neuropathy test platform. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:1-4. [doi: 10.1109/EMBC.2018.8513100] [Medline: 30440300]
27. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
28. Cox SM, Lane A, Volchenboum SL. Use of Wearable, Mobile, and Sensor Technology in Cancer Clinical Trials. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.17.00147] [Medline: 30652590]
29. Cohen HS, Mulavara AP, Peters BT, Sangi-Haghpeykar H, Kung DH, Mosier DR, et al. Sharpening the tandem walking test for screening peripheral neuropathy. South Med J 2013 Oct;106(10):565-569 [FREE Full text] [doi: 10.1097/SMJ.0000000000000009] [Medline: 24096950]
30. Argyriou AA, Bruna J, Anastopoulou GG, Velasco R, Litsardopoulos P, Kalofonos HP. Assessing risk factors of falls in cancer patients with chemotherapy-induced peripheral neurotoxicity. Support Care Cancer 2020 Apr;28(4):1991-1995. [doi: 10.1007/s00520-019-05023-5] [Medline: 31378844]
31. Osumi M, Sumitani M, Abe H, Otake Y, Kumagaya S, Morioka S. Kinematic evaluation for impairment of skilled hand function in chemotherapy-induced peripheral neuropathy. J Hand Ther 2019;32(1):41-47. [doi: 10.1016/j.jht.2017.06.003] [Medline: 29042161]
32. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. Parkinsonism Relat Disord 2015 Jun;21(6):650-653. [doi: 10.1016/j.parkreldis.2015.02.026] [Medline: 25819808]
33. Barrios L, Oldrati P, Lindlbauer D, Hilty M, Hayward-Koennecke H, Holz C. A rapid tapping task on commodity smartphones to assess motor fatigability. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 2020; Honolulu, HI p. 1-10. [doi: 10.1145/3313831.3376588]
34. Bajaj JS, Thacker LR, Heuman DM, Fuchs M, Sterling RK, Sanyal AJ, et al. The Stroop smartphone application is a short and valid method to screen for minimal hepatic encephalopathy. Hepatology 2013 Sep;58(3):1122-1132 [FREE Full text] [doi: 10.1002/hep.26309] [Medline: 23389962]
35. Loprinzi CL, Lacchetti C, Bleeker J, Cavaletti G, Chauhan C, Hertz DL, et al. Prevention and Management of Chemotherapy-Induced Peripheral Neuropathy in Survivors of Adult Cancers: ASCO Guideline Update. J Clin Oncol 2020 Oct 01;38(28):3325-3348. [doi: 10.1200/JCO.20.01399] [Medline: 32663120]

Abbreviations

CIPN: chemotherapy-induced peripheral neuropathy
CIPN20: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-induced Peripheral Neuropathy 20-item scale
PCA: principal component analysis
PLSDA: partial least squares discriminant analysis
PRO: patient-reported outcome
PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
VIP: variable importance in projection

Edited by G Eysenbach; submitted 27.01.21; peer-reviewed by S Faithfull, K Trinkley, F Lanfranchi, R Knoerl; comments to author 11.03.21; revised version received 07.04.21; accepted 15.04.21; published 05.07.21.

Please cite as:
Chen CS, Kim J, Garg N, Guntupalli H, Jagsi R, Griggs JJ, Sabel M, Dorsch MP, Callaghan BC, Hertz DL
Chemotherapy-Induced Peripheral Neuropathy Detection via a Smartphone App: Cross-sectional Pilot Study
JMIR Mhealth Uhealth 2021;9(7):e27502
URL: https://mhealth.jmir.org/2021/7/e27502
doi:10.2196/27502
PMID:36260403

Original Paper

Patients' and Clinicians' Perceived Trust in Internet-of-Things Systems to Support Asthma Self-management: Qualitative Interview Study

Chi Yan Hui[1], BEng, CEng, MSc, PhD; Brian McKinstry[2], MD; Olivia Fulton[3], BN; Mark Buchner[4], BSc; Hilary Pinnock[1], MD

[1]Asthma UK Centre for Applied Research, Usher Institute, Edinburgh, United Kingdom

[2]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[3]Patient Advisory Group, Asthma UK Centre for Applied Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[4]Tactuum Ltd, Glasgow, United Kingdom

Corresponding Author:
Hilary Pinnock, MD
Asthma UK Centre for Applied Research, Usher Institute
The University of Edinburgh
Doorway 3, Medical School, Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 0131 650 8102
Email: hilary.pinnock@ed.ac.uk

Abstract

Background: Asthma affects 235 million people worldwide. Supported self-management, including an action plan agreed with clinicians, improves asthma outcomes. Internet-of-things (IoT) systems with artificial intelligence (AI) can provide customized support for a range of self-management functions, but trust is vital to encourage patients' adoption of such systems. Many models for understanding trust exist, some explicitly designed for eHealth, but no studies have used these models to explore trust in the context of using IoT systems to support asthma self-management.

Objective: In this study, we aim to use the McKnight model to explore the functionality, helpfulness, and reliability domains of patients' and clinicians' trust in IoT systems to deliver the 14 components of self-management support defined by the PRISMS (Practical Reviews in Self-Management Support) taxonomy.

Methods: We used *think-aloud* techniques in semistructured interviews to explore the views of patients and clinicians. Patients were recruited from research registers and social media and purposively sampled to include a range of ages, genders, action plan ownership, asthma duration, hospital admissions, and experience with mobile apps. Clinicians (primary, secondary, and community-based) were recruited from professional networks. Interviews were transcribed verbatim, and thematic analysis was used to explore perceptions of the functionality, helpfulness, and reliability of IoT features to support components of supported self-management.

Results: A total of 12 patients and 12 clinicians were interviewed. Regarding perceived functionality, most patients considered that an IoT system had functionality that could support a broad range of self-management tasks. They wanted a system to provide customized advice involving AI. With regard to perceived helpfulness, they considered that IoT systems could usefully provide integrated support for a number of recognized components of self-management support. In terms of perceived reliability, they believed they could rely on the system to log their asthma condition and provide preset action plan advice triggered by their logs. However, they were less confident that the system could operate continuously and without errors in providing advice. They were not confident that AI could generate new advice or reach diagnostic conclusions without the interpretation of their trusted clinicians. Clinicians wanted clinical evidence before trusting the system.

Conclusions: IoT systems including AI were regarded as offering potentially helpful functionality in mediating the action plans developed with a trusted clinician, although our technologically adept participants were not yet ready to trust AI to generate novel advice. Research is needed to ensure that technological capability does not outstrip the trust of individuals using it.

(JMIR Mhealth Uhealth 2021;9(7):e24127) doi:10.2196/24127

KEYWORDS

asthma; self-management; telehealth; internet-of-thing; trust

Introduction

Background

Asthma is a variable long-term condition that affects 235 million people worldwide [1]. In everyday life, patients decide how to maintain control of their asthma and what to do if their condition worsens. When they are unsure what to do, they contact their health care advisor; within the UK health care system, this would normally be the general practitioner (GP) or primary care asthma nurse. Supported self-management of asthma has many components [2] but specifically includes provision by the patients' usual health care professional of a personalized action plan summarizing agreed decisions (eg, medication adjustment or emergency strategies) [3-5].

With the increasing availability of sensors and improved coverage of wireless networks, an internet-of-things (IoT) system has the capability to observe patients' status and medication use and support self-management. Devices have the intelligence to perform a task independently or connect to other sensory networks, platforms, and mobile phones to perform multiple tasks. Artificial intelligence (AI) may be *narrow* (artificial narrow intelligence: systems that interact with users based on a set of planned rules) or can *mimic*, *equal*, or ultimately *surpass* human intelligence to create new ways to interact with users (described as artificial general intelligence or artificial superintelligence, respectively) [6].

The IoT has been used to support clinical management in a range of contexts (eg, asthma, diabetes, and hypertension), with examples including diagnosis, remote monitoring, remote consultation, self-management, emergency care, and home rehabilitation [7-13]. Traditional trust between patients and their clinicians is associated with improved medication adherence and health outcomes [14-18] and can be harnessed to encourage adoption and continued use of digital health systems [19-21]. Underpinning this is a gradual shift in trust from the clinician to technology.

The Concept of Trust

The concept of trust is *elusive* [22] but is typically illustrated as a relationship between 2 agents (a trustor and a trustee) [23]. Terms such as *confidence*, *have faith in*, and *believe in* are commonly used, and in the health care context, multiple attributes have been summarized broadly as "The belief that a doctor is working in the patient's best interest" [14]. The term *e-trust* has been used to describe the trust between a human agent (eg, patient, clinician, or health carer) and a digital artifact agent (eg, whether it can achieve a given goal) [24]. However, in an IoT system there may also be *trust* among artifact agents; for example, an AI system may rely on (or *trust*) the technical specifications of a smart device and system to collect and transfer accurate data on which to base advice to the user.

In the context of supported self-management, patients are the core users of digital health services such as health information websites, web-based consultations, or online support groups. Patients adopt telehealth for many reasons, such as personal, technological, institutional, and legislative, but in the decision of delegating a specific task to an intelligent system, a fundamental factor is whether the patient trusts that the system can fulfill their expectations. Models of e-trust have defined multiple factors required for the trustor to decide to trust a digital health system broadly classified as follows [25-30]:

- Personal factors such as altruism, ease of use, self-efficacy, sociodemographic characteristics, usefulness, recommendation by others, fair use of data, and cost.
- Technological factors such as customization, interoperability, and data privacy.
- Institutional factors such as ability (or not) to improve communication with their clinician, professional training, the accuracy of the information provided to the clinical service, service provider's reputation, the organization's nature, or business model.
- These might be added to *legislative factors* in the context of health care, as (for example) medical device registration requires evidence of technological performance, effectiveness, and safety, which demonstrates that a product is worthy of trust [31-34].

The McKnight model [35], in comparison with e-trust models, is based on an interpersonal trust model between the human agent and the digital artifact agents and sees trust in technology as task specific (aligned with the Castelfranchi and Falcone [23] cognitive trust model for human agents [36]). This model conceptualizes three dimensions of task-specific trust in technology: functionality, helpfulness, and reliability. In the context of supporting asthma self-management, *functionality* is how patients and clinicians believe an IoT system has the features and capability to accomplish a range of self-management tasks. *Helpfulness* is the degree to which patients and clinicians believe an IoT system can provide an adequate, responsive, and useful aid to support their asthma self-management tasks and decisions. *Reliability* is whether patients and clinicians believe an IoT system can operate continuously and properly to support tasks.

Trust in the Context of Digital Support for Self-management

Although there are many trust models [20,37], including some in eHealth [29,30,38], no studies have used existing models explicitly to explore trust in using IoT systems to support asthma self-management. The McKnight trust model is task specific, enabling a comprehensive investigation of the app features and various device combinations of the IoT system as opposed to examining the digital health system as a *black box*. Therefore, using asthma as an example, we aim to use the McKnight trust model to explore the domains of trust beliefs between patients or clinicians and IoT systems in the context of the PRISMS (Practical Reviews in Self-Management Support) taxonomy, a framework defining components of self-management support in long-term conditions [2].

Methods

Ethical Considerations

The study was conducted between May 2019 and January 2020 with the approval of the London Fulham Research Ethics Committee (ref: 19/LO/0703), sponsored by the University of Edinburgh and the National Health Service (NHS) Lothian (Academic and Clinical Central Office for Research and Development) and funded by the Chief Scientist Office/Asthma UK Innovation Grant (ref: CSO-AUK-2018-03). All participants provided informed consent before the interviews.

Design

We used semistructured interviews to explore patients' and clinicians' trust in using IoT systems to support asthma self-management. Purposive sampling (see *Purposive Sampling* section) continued until we achieved data saturation with regard to our aim; we estimated from previous studies that this would be 12 patients and 12 clinicians [39].

Patient Recruitment

We recruited people (aged ≥16 years) with *active asthma* (defined as a physician diagnosis of asthma and at least one asthma treatment prescribed in the previous year [40]) in the United Kingdom. We wanted to explore the perceived trust between patients and technology, so we excluded children and adolescents as the involvement of a parent or guardian would have added an additional person to the interactions. We recruited patients through volunteer databases (Scottish Health Research Register [41], Register for Asthma Research [42], Asthma UK volunteer database, and social media of Asthma UK and Asthma UK for Applied Research).

Potential participants were invited to register their interest on our recruitment webpage, which provided an information sheet. They were asked to confirm their eligibility (diagnosed with asthma by their GP, ≥16 years, and living in the United Kingdom), provide their demographics, and give us consent to contact them to complete registration.

Purposive Sampling

From the information provided, we purposively sampled patients to achieve maximum diversity of perceptions about the use of technology to support self-management. Sampling was based on the following criteria:

- Age range (16-25 years, 26-45 years, 46-65 years, and ≥65 years)
- Ownership of action plan (or not)
- Duration of asthma (diagnosed within <6 months, 6-12 months, 1-10 years, or >10 years)
- Admission to hospital in the past 12 months (or not)
- App download experience (ie, can download apps by themselves, have asked someone to download apps for them, or have never downloaded an app)

Clinicians' Recruitment

We recruited primary, secondary, and tertiary care clinicians who provided routine care for children or adults with asthma. We posted advertisements in the newsletter and social media of the NHS Research Scotland Primary Care Network and professional bodies such as the Primary Care Respiratory Society [43]. We also invited individual clinicians known to have an interest in asthma and technology.

Data Collection

We conducted in-depth semistructured interviews with patients to understand their perceived use of self-management support features and specifically explore their perceived trust in using IoT systems to support their self-management. The potential features we explored were from our previous work and the scope of commercially available devices. We provided images of smart devices and data (Multimedia Appendix 1) and asked patients to design a personalized IoT system incorporating the features they thought would help them live with asthma. We used *think-aloud* techniques to explore their trust (or not) in using the IoT system they had created to support their self-management. Clinicians were asked to formulate IoT systems that would support self-management and the care they provide for people with asthma and explored their trust in the features and the IoT system (see Multimedia Appendix 2 for the topic guide).

Data Synthesis and Analysis

Interviews were digitally recorded, transcribed, and analyzed using NVivo version 12 (QSR International) [44]. We used the McKnight trust model [35] to categorize patients' and clinicians' perceptions of their trust in the functionality, helpfulness, and reliability of using IoT systems to support asthma self-management.

We used a framework analysis [45], creating a matrix of self-management support features against perceptions of the McKnight trust model (or not) expressed by patients and clinicians. All interview data related to trust were extracted into the matrix and aligned with the features to which they referred. To increase applicability to other long-term conditions and because the perceived trust domains (functionality, helpfulness, and reliability) in the McKnight model are task specific, we mapped the perceptions of trust to components used to support self-management in long-term conditions, as described in the Practical Systematic Review of Self-Management Support taxonomy for long-term conditions [2]. We were alert to other trust-related themes that did not fit the matrix either because they did not reflect the domains of *functionality, helpfulness, and reliability* or because they were overarching rather than task-related.

A research team member (CYH) coded 1 patient and 1 clinician interview, which was then reviewed by another researcher (HP). The 2 researchers discussed their decisions and standardized the coding for the rest of the transcriptions. CYH coded all the data related to perceived trust (or not). HP reviewed the matrix for quality control.

Reflexivity and Interpretation

CYH has research expertise in exploring user preferences for asthma apps and academic interest in developing IoT systems to support asthma self-management. She discussed the coding and interpretation of results with the study team members from

different backgrounds and experiences, including GPs, a patient, and a technology developer, to ensure a broad interpretation.

Results

Participants

Patients

From 362 expressions of interest (268, 74% women), we purposively sampled 12 (3.3%) patients with a range of ages, gender, and action plan ownership for interviews (Textbox 1). The resultant maximum variation sample included more women (8 women and 4 men). None had been diagnosed with asthma for less than a year, and all were confident in their ability to download an app without asking for help.

Textbox 1. Patients' and professionals' demographics.

Patients (N=12)

- Age (years): spread across 4 age groups from teenage or young adults to ≥65 years: 16-25 years (n=3), 26-45 years (n=2), 46-65 years (n=3), and ≥65 years (n=4).
- Gender: 8 women and 4 men.
- Ownership of an asthma action plan: only 4 had been given a written action plan. Of the 8 who did not have an action plan, 5 had been "told what to do." Of the 5 participants who had been "told what to do," 2 were aged 46-65 years, and 3 were ≥65 years.
- Duration of asthma: 8 (4 men) had had asthma for more than 10 years; none were newly diagnosed.
- Hospital admissions in the previous 12 months: only 4 had had a hospital admission in the previous year, 3 of whom were still under a specialist clinic. None of the male participants had had an admission.
- Experience in using apps: all the participants were confident to download an app by themselves.

Clinicians (N=12)

- Primary care clinicians, n=4 (2 general practitioners [GPs] and 2 asthma nurses).
  - Gender: 1 man and 3 women.
  - Practice experience: GPs with 8 years' experience; asthma nurses with 20 years' experience. GPs had research experience in digital health for patients with asthma.
  - Technology experience: asthma nurses had experience in using remote telemonitoring for hypertension.
- Secondary care clinicians, n=4 (1 respiratory consultant and 3 respiratory pediatricians).
  - Gender: 1 man and 3 women.
  - Practice experience: respiratory consultant: diagnostics, asthma management, and severe asthma care; respiratory pediatricians: asthma management in a range of asthma severities.
  - Technology experience: 1 had used an asthma app, 1 uses smart inhalers in their service and research, and 2 had research experience in asthma technology.
- Pharmacists, n=4 (1 hospital pharmacist and 3 primary care support pharmacists or prescribing advisors).
  - Gender: 1 man and 3 women.
  - Practice experience: 1-14 years' experience in reviewing asthma medications.
  - Technology experience: all used web-based repeat prescriptions services; 1 developed an asthma app.

Clinicians

We recruited 12 UK clinicians (GPs, asthma nurses, pharmacists, consultant chest physicians, and respiratory pediatricians) who provided care for people with asthma. Most had experience using technologies such as smart inhalers, mobile apps, and SMS text messages to support respiratory patients in their practices or hospitals.

Overview of Results

Perceptions related to the 3 domains of the McKnight model of task-specific trust in technology (functionality, helpfulness, and reliability [35]) are synthesized below. Multimedia Appendix 3 lists the perceived trust in functionality, helpfulness, and reliability in IoT features related to generic long-term conditions or asthma self-management tasks. Finally, we considered the overarching domain of trust in data security, which was clearly important to our participants, reflecting not only the properties of the technology but also the context within which it was implemented.

Perceived Functionality of IoT Components to Support Self-management

Most patients perceived that the IoT had functionality that could well support a range of self-management tasks (Multimedia Appendix 3 lists examples of tasks that participants trusted the IoT to deliver). This belief was often based on past technological experiences:

> *I do use technology to control my asthma, so I keep copies of my peak flows and I can do charts on my laptop so, when I am deteriorating and I end up in hospital, I can take this with me and show them that obviously it's happened over a period of days. And I use alarms on my phone as well so I can wake up and have my medication because I have to have four-hourly nebulisers as well at the minute to control it.* [P6, 16-25 years, female]

> *I tend to put a reminder on my phone so I can have the one (asthma review) in a year's time, but it is a bit of effort.* [P10, 16-25 years, female]

Some patients perceived that the IoT system could have the functionality to support how they lived with asthma, although these features were not yet available in the market:

> *I think if there was something similar [to energy saving tips in a smart home] on the app where you're using the app but it gives you a tip each day that you know, "air pollution could be a trigger for your asthma" or "washing can be a trigger for asthma", then that might give you some additional information.* [P7, 46-65 years, male]

> *I think kind of mindfulness breathing exercises you can find on, like, YouTube. If it was, like, breathing exercises to assess the asthma, it might be the sort of thing I might try once and see what I thought of it and if I thought it was useful I might try it again.* [P10, 16-25 years, female]

Some clinicians perceived that the IoT had the functionality to engage patients to look after their asthma and support self-management. They believed (in the future) systems could transfer patients' manual or auto logs to a health care professional for review or flag up when inhaler medication needed to be replenished. In contrast, others doubted whether technology could change patients' behavior:

> *There isn't an app that I'm aware of that can link with the GP systems. So if that is possible from a technology perspective, inputting how much they're using and there's a log then of when they have their new prescription, and then that app then talks to the GP system it can flag when they get to a certain level and order a repeat, I think that's perfectly feasible.* [HCP2, pharmacist in hospital, female]

> *I think patients either are physically active or they're not, and the app's not going to make them physically active if they're not.* [HCP5, consultant chest physician, female]

> *In the last few months in my pharmacy we've introduced...well, we always had online ordering but there wasn't huge engagement with it but we've introduced an app-based system for ordering. A younger population who is ordering things like asthma inhalers and contraceptive pills and so on have really engaged with that actually quite well.* [HCP7, pharmacist in practice, male]

Perceived Helpfulness of Supporting Components of Self-management

Most patients had a perception that IoT systems could provide a useful service to provide integrated support for a number of recognized components of self-management support [39] (Multimedia Appendix 3 lists some examples of tasks that participants thought would be helpful for the IoT to provide).

They wanted IoT systems to log data about their asthma symptoms, peak flow, medication use, inhaler technique, indoor or outdoor environmental data, activity intensity, and weight, and perceived it would be helpful if these data could be collected effortlessly, such as a voice assistant asking about their asthma (eg, "Good morning! Did your asthma disturb your sleep last night?") or automatic data collection from wearable devices or environmental sensors in their living areas. Some specific ways in which they thought an IoT system would be useful if they could help them look after their asthma by providing customized alerts and advice were the following:

- Identifying unusual asthma symptoms or peak flows and automatically providing customized information about their asthma and advice on medication adjustment and follow-up actions (suggesting and counting the number of rescue puffs to be taken in an emergency, calling medical help)
- Alerting them if their inhaler technique was incorrect
- Detecting unusual use of rescue inhalers to help them identify what triggered their asthma
- Reminding them to comply with their preventer inhaler

In addition, they thought that an IoT system would be helpful in supporting their communication with clinicians. Most participants believed it would be helpful to be able to ask quick questions or arrange follow-up consultations with clinicians via text, WhatsApp, or email and then be able to share their data with clinicians to assess their asthma status. Some patients thought objective evidence from logs would help explain their asthma to their friends or senior colleagues at work:

> *I've missed so many events in my life because of my asthma and I think it's difficult to say to someone. I think if you had this medical evidence behind you, they'd understand without you having to explain it.* [P6, 16-25 years, female]

> *I think particularly my parents. I live in a flat on my own and if for whatever reason during the night I was suddenly puffing my blue inhaler multiple times, I'd almost want a warning siren to be sent to my parents just in case I'm really struggling.* [P12, 26-45 years, female]

A patient who had had a recent hospital admission thought it would be helpful to automatically share their asthma logs with the emergency department and share test reports between different hospitals to prevent treatment delay. Some patients

described how they panicked when they were very short of breath and could lose track of how many puffs of their reliever inhaler they had taken in a short period. A system that counted the doses of reliever inhaler they had taken and warned them in real time about overdose would be a helpful safety net. Patients with multimorbidity wanted the system to integrate information from different health care specialists about all their treatments and provide medication advice to reduce the side effects of different drugs.

Most clinicians agreed that receiving data about peak flow and symptoms would help them assess asthma status in reviews, but also highlighted the benefits of an IoT system that could transfer objective data on incorrect or correct inhaler technique and medication use to help assess the adherence and suitability of the inhaler device:

> *Because very often patients don't remember to bring their inhaler with them so it's difficult to always test when they're in the clinic. So if you're being alerted to that, when they're doing it at home, then that's perfect, because if you ask a patient are you doing it right, they always say "yes".* [HCP1, GP, female]

> *If I've got some hard data on their peak flow and their symptoms over the last couple of months, and their adherence, that gives me a much better idea of what I need to do with them, so that's incredibly helpful for me.* [HCP5, consultant chest physician, female]

Perceived Reliability

Patients and clinicians discussed reliability—whether they trusted the IoT system would operate continuously and without error—in two contexts: logging data and providing advice.

Logging Data

Some patients observed that a system that logged data (such as coughing, sleep disturbance, and medication use) automatically *in the background* would reduce missing data. They believed that smart peak flow meters and smart inhalers could reliably capture data, although there were caveats. Some patients did not always carry these devices with them or had more than one reliever inhaler in use (at home, at work, and in the car), and a reliable system would need to accommodate these behaviors. Some patients suggested that a voice assistant was easier to use, but others raised concerns about its accuracy. Most clinicians agreed that automatic logs are more accurate, as they reduce human error:

> *If it can capture most things, like obviously in the air it's cold or there's pollen or there's pollution, I could probably trust it quite a bit, that, because it's solid data that's already captured in other places.* [P1, 46-65 years, female]

> *I think it (an IoT system) might be more accurate as well than say if I did it (logging) myself.* [P11, 16-25 years, female]

> *I know some people say that sometimes they [patients] come in and they sit in the waiting room and they're filling in the results.* [HCP4, prescribing support pharmacist in practice, female]

> *I suppose adding technology in to it might make it more accurate and take out the human error and that.* [HCP10, GP, male]

Providing Advice

Most patients believed that the system could accurately highlight the advice on an agreed action plan when their condition was getting worse but were skeptical that the system could safely generate new advice. They would trust the system to reliably prompt an alarm when their condition worsened or if they took their inhaler incorrectly, identify environmental triggers and recommend avoidance, and remind them of the actions suggested on an agreed action plan. In contrast, all patients preferred their clinician to interpret the data and decide on new advice. Similarly, patients did not believe that the system could take *human factors* (such as the impact of psychological or emotional context) into account when reaching a diagnosis. Clinicians were also comfortable with an IoT-based early warning system to alert patients to seek further assistance when their condition worsened but considered the automatic generation of new advice as an *unproven route*. They also cited the importance of personal relationships. They accepted that AI may be used to generate new intelligent advice to patients in the future but would need evidence to prove clinical accuracy before trusting in its reliability:

> *Well, again it goes through two stages, so if I'm really bad, maybe a message to say I'm really bad and...probably notice my wife first (to make decisions) and then the health care professional (to arrange follow up actions)...I wouldn't want that to trigger an appointment with a health care professional.* [P4, 26-45 years, male]

> *Like if it was suggesting changes to me, that feels more like the time that I would actually have to have a conversation with the GP or nurse rather than my phone triggering stuff like that.* [P12, 26-45 years, female]

> *So if there's some sort of really intelligent system that can work out how to do an asthma action plan for somebody based on intelligent peak flow monitoring and intelligent looking at the symptoms and all of that, great, but until we've got that we need a human being I think to sit with the patient and make an asthma action plan. Because even as an experienced clinician it can be quite challenging sometimes because you have to know quite a lot about asthma to do them.* [HCP5, consultant chest physician, female]

> *I don't think we're at a stage where a system can advise patients. I would be a little bit nervous about it. I'd have to have proof that that actually works because I think I would recognize that in my practice I establish a relationship with a patient. A machine advice, automated advice doesn't necessarily understand that.* [HCP11, pediatrician, male]

Trust in Data Security

Privacy of personal data was a strong overarching theme that emerged in the interviews. Although clearly relevant to trust but not task specific (as in the McKnight trust model), patients were found to accept the health services to implement IoT systems if they knew how their data would be used. Attitudes varied, with a patient suggesting they were not concerned about data security, whereas another explained that they were not happy to use a voice assistant because it was connected to the cloud service. Most patients and clinicians wanted to use SMS text messages or emails for follow-up questions, although both suggested that the General Data Protection Regulation was a barrier to adopting these services in the NHS. Clinicians balanced the data privacy risk and the helpfulness of the services and thought that explaining to patients about the use of their data and having their consent was a pragmatic approach, as opposed to blocking the service completely:

> *That (Email communication) would be useful sometimes, but they (health care professionals) wouldn't do it, so I don't really know...(the clinicians) they'd be worried about that (spam in email), same with text messages and WhatsApp...It might work from my side, but I don't think it would work from their side.* [P5, >65 years, male]

> *The NHS contract can be difficult, with regards to GDPR and so on, so nearly everything is done via phone, and if you can't speak to an actual person, we don't routinely leave messages and so on.* [HCP7, pharmacist in practice, male]

> *This is personal data but it's only about your health condition. So in one way I wouldn't be that worried about that so much because actually that's just about one condition that actually you want to make sure that people know about so that you actually get treated properly.* [HCP4, prescribing support pharmacist in practice, female]

Discussion

Principal Findings

Most patients believed IoT systems to be functional and helpful in supporting a broad range of self-management tasks, but they raised some concerns about reliability. They believed IoT systems could collect their data accurately from devices, check for incorrect inhaler techniques, and advise them on treatment options based on the thresholds and actions agreed with clinicians (eg, in an action plan) and customized to their situation. However, they doubted whether the system could interpret their data to generate novel advice or reach diagnostic conclusions. They would want to check with a health care professional for reassurance and *human* advice before acting on AI-determined actions. Most patients' beliefs resonated with those of the clinicians. Before trusting and adopting AI-developed advice, clinicians wanted evidence to reassure them about accuracy. Pragmatic approaches are required to deliver services based on the requirements of the General Data Protection Regulation. Our study did not find a diversity of views among different ages and genders, possibly because all participants had experience with technology and were the end users of similar NHS asthma care services in the United Kingdom. Racial biases, sociocultural norms, and an understanding of AI are other potential factors that need to be considered when developing IoT-supported services applicable to diverse communities.

Strength and Limitations

We explored perceived trust in IoT systems from the perspective of patients and clinicians; however, there are some limitations. First, patients and clinicians based their opinions on their past experience of existing technologies and arrangements within current health care services. Our clinician participants were interested in technology and asthma, which will have influenced their opinions that were based on their experience, personal interest, age, and gender. The views from these groups of participants may not apply to users with limited access to technology or lack of experience with digital options. Real-life experiences with an IoT system may have generated new themes. However, our findings represent perceived expectations from patients and clinicians and can therefore inform future IoT system design and underpin further investigation. Second, because of time and resource limitations, we did not interview children (patients aged ≤16 years) and their parents, although we included experienced pediatricians to explore some of the issues from their perspective. Third, we could not recruit patients who were newly diagnosed (0-1 year) with asthma who may have had specific requirements, although our experienced asthma patients provided some feedback on their needs or expectations when they were newly diagnosed. Fourth, all of our participants were confident in using technology such as social media, web information, voice assistants, and activity trackers. Hence, participants may be biased in assessing perceived functionality, helpfulness, and reliability because of their past use of technologies. However, their real-life experience enabled them to give examples of IoT features that they considered trustworthy (or not) from personal experience. Finally, the McKnight domains used in this study were limited to perceived functionality, helpfulness, and reliability; other domains such as perceived ease of use, perceived value, and the source of the recommendation (eg, an app recommended by a clinician that the patient believes understands their asthma may engender more trust in technology than an app recommended by a clinician that the patient does not know or trust) [21,46] may also be important to the perceived trust in the asthma self-management IoT system.

Interpretation in Relation to the Published Literature

Our findings show that patients and clinicians both recognized the potential of IoT systems to provide a range of customized support for self-management, which they believed would help them look after their asthma. They trusted smart devices to observe their status accurately and had confidence that the system could trigger advice previously agreed with the clinician when they experienced unusual asthma symptoms or reduced peak flows. They found it acceptable for systems to detect errors, correct inhaler techniques, and address noncompliance to medication. These functions imply that IoT systems can be trusted to include AI that can learn about an individual's asthma

throughout time and provide advice based on a set of rules. However, neither patients nor clinicians trusted the IoT system to mimic clinicians' intelligence and create new self-management advice, preferring a human check to reassure that the AI advice was applicable to the individual before deciding what to do. This resonates with the findings of a recent review on AI clinical interventions in the context of other long-term conditions, such as depression, weight, nutrition, limb pain, and smoking cessation management [47]. People trusted a customized system including elements that imitated human-human (patient-clinician) interactions and provided easy communication channels between the patient and clinician. Furthermore, the involvement of clinicians was pivotal to encourage patients' adoption and adherence to digitized self-management [21].

Technically, patients and clinicians are reluctant to move from using *narrow* intelligence that follows preset rules (artificial narrow intelligence) to general (artificial general intelligence) or superintelligence (artificial superintelligence) in which the system initiates rules. Although there are high-level guiding principles [48,49] and governance recommendations [50] to ensure that future AI designs are ethically and technically trustworthy [51] (eg, to ensure the use of AI is fair, is transparent, and meets universal human values), they focus on the trust between AI and the community. Few have explicitly considered trust between AI and individual patients in the context of supported self-management.

Patients are not yet ready to transfer their trust from the clinician (a human) they know to an IoT system (a machine) generating self-management advice through AI. In the traditional self-management model, the GP or asthma nurse assesses the patient's condition and agrees with self-management advice in a face-to-face consultation. In the new IoT self-management model, the app interface, smart devices, or lifelike robots (in the near future) have the responsibility to sense the patients' condition, which replaces the clinicians' intelligence in giving self-management advice to patients. The decision process is an impenetrable *black box* for patients and clinicians. In contrast, clinicians in the traditional model can discuss options with the patient and consider aspects such as patients' mood, personality, self-management habits, and experiences so that the final decision is (relatively) transparent. This may be a reason many patients trust that AI-based IoT systems can record their asthma condition better than themselves, but none have shifted their trust from the clinician to the AI in terms of issuing new advice.

From e-commerce literature, we know that it is possible to shift people's trust from a known person, organization, or shop to an electronic service related to the known entity [30,52,53] or recommended by the known person, organization, or shop [54]. Iterative interaction with an automated system or lifelike robot can build up trust for first-time users who are curious about new systems and robots but struggle to use them in their daily lives [55,56]. In the health care context, studies of apps and e-consultations have suggested the potential to transfer trust from a physical health care service (eg, appointment booking and monitoring physiological parameters or activity after discharge from hospital) to an app [57-59]. However, to encourage clinicians to recommend an AI system to patients, strong clinical evidence is required to earn their trust. Currently, there is little evidence in the context of asthma self-management to reassure clinicians or patients.

Conclusions

Introducing IoT systems involving advice from AI to support self-management requires more than just functionality that can deliver tasks users regard as helpful. There is a need to increase the trust of users in the reliability of systems as AI moves from the currently acceptable *narrow* intelligence directed by clinician-determined action plans to a future in which advice is generated by the IoT system. Our technologically adept participants were not yet ready for this step; research is needed to ensure that technological capability does not outstrip the trust of the individuals using it.

Acknowledgments

The authors would like to thank the patient and professional participants for taking part in this study. They also thank Asthma UK and the Asthma UK Centre for Applied Research, the Scottish Health Research Register, and Register for Asthma Research for help with patient recruitment; the NHS Research Scotland Primary Care Network, the Primary Care Respiratory Society, and the NHS Lothian Respiratory Managed Clinical Network for help with professional recruitment; and Asthma UK and the Asthma UK Center for Applied Research Patient Public Involvement group members for advice on the interview materials.

This work was funded by the Chief Scientist Office/Asthma UK Innovation grant (ref: CSO-AUK-2018-03). The views expressed in this publication are those of the authors and not necessarily those of the Chief Scientist Office (Scotland) or Asthma UK.

Authors' Contributions

CYH and HP designed the study. CYH performed the data extraction and synthesized the data with HP. HP is the study guarantor. CYH and HP wrote the initial draft and final version of the manuscript. BM reviewed the final manuscript. OF commented on the findings from the patient's perspective, and MB commented on implications from a technology perspective. All the authors approved the final version of the manuscript.

Conflicts of Interest

CYH has received grant funding from the Chief Scientist Office/Asthma UK Innovation grant (ref: CSO-AUK-2018-03) to plan and carry out the study. BM and HP received grant funding from Philips NV. BM is paid as a consultant to the Scottish Government

to advise on clinical issues of remote health monitoring. He has received consultancy payments from Pharmatics, a company developing remote health monitoring for monitoring COPD. MB is the Managing Director of Tactuum Ltd. OF contributes to the lay capacity of Teva Pharmaceuticals, AstraZeneca, and WEGOHealth.

Multimedia Appendix 1
Images of smart devices and data.
[PDF File (Adobe PDF File), 483 KB - mhealth_v9i7e24127_app1.pdf]

Multimedia Appendix 2
Topic guide.
[PDF File (Adobe PDF File), 176 KB - mhealth_v9i7e24127_app2.pdf]

Multimedia Appendix 3
Perceived trust table.
[PDF File (Adobe PDF File), 247 KB - mhealth_v9i7e24127_app3.pdf]

References

1. Asthma. World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/asthma [accessed 2021-02-28]
2. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CL, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
3. Pinnock H, Parke HL, Panagioti M, Daines L, Pearce G, Epiphaniou E, PRISMSRECURSIVE groups. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64 [FREE Full text] [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
4. British Guideline on the Management of Asthma. British Thoracic Society/Scottish Intercollegiate Guideline Network. 2019. URL: https://brit-thoracic.org.uk/quality-improvement/guidelines/asthma/ [accessed 2021-02-28]
5. The Global Strategy for Asthma Management and Prevention (GINA). Global IFA. 2020. URL: https://ginasthma.org/reports/ [accessed 2021-02-28]
6. Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons 2019 Jan;62(1):15-25. [doi: 10.1016/j.bushor.2018.08.004]
7. Saheb T, Izadi L. Paradigm of IoT big data analytics in the healthcare industry: a review of scientific literature and mapping of research trends. Telematics Informat 2019 Aug;41:70-85. [doi: 10.1016/j.tele.2019.03.005]
8. Thilakarathne NN, Kagita MK, Gadekallu DT. The role of the internet of things in health care: a systematic and comprehensive study. Int J Eng Manag Res 2020 Aug 31;10(4):145-159. [doi: 10.31033/ijemr.10.4.22]
9. Deshkar S, Thanseeh R, Menon V. A review on IoT based m-Health systems for diabetes. Int J Comput Sci Telecommun 2017 Jan;8(1):13-18 [FREE Full text]
10. Baali H, Djelouat H, Amira A, Bensaali F. Empowering Technology Enabled Care Using <italic>IoT</italic> and Smart Devices: a Review. IEEE Sensors J 2018 Mar 1;18(5):1790-1809. [doi: 10.1109/jsen.2017.2786301]
11. MyAirCoach. 2015. URL: https://www.myaircoach.eu/content/what-myaircoach-project [accessed 2021-02-28]
12. Bui AA, Hosseini A, Rocchio R, Jacobs N, Ross MK, Okelo S, et al. Biomedical REAl-Time Health Evaluation (BREATHE): toward an mHealth informatics platform. JAMIA Open 2020 Jul;3(2):190-200 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa011] [Medline: 32734159]
13. Hui C, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 1;24(3):619-632 [FREE Full text] [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
14. Rolfe A, Cash-Gibson L, Car J, Sheikh A, McKinstry B. Interventions for improving patients' trust in doctors and groups of doctors. Cochrane Database Syst Rev 2014 Mar 4(3):CD004134 [FREE Full text] [doi: 10.1002/14651858.CD004134.pub3] [Medline: 24590693]
15. Bending ZJ. Reconceptualising the doctor-patient relationship: recognising the role of trust in contemporary health care. J Bioeth Inq 2015 Jun 16;12(2):189-202. [doi: 10.1007/s11673-014-9570-z] [Medline: 25124984]
16. O'neill O. Autonomy and Trust in Bioethics. Cambridge University Press. URL: http://locatorplus.gov/cgi-bin/Pwebrecon.cgi?DB=local&v1=1&ti=1,1&Search_Arg=101149509&Search_Code=0359&CNT=20&SID=1 [accessed 2021-02-28]
17. Birkhäuer J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: a meta-analysis. PLoS One 2017 Feb 7;12(2):e0170988 [FREE Full text] [doi: 10.1371/journal.pone.0170988] [Medline: 28170443]
18. Gille F, Smith S, Mays N. Why public trust in health care systems matters and deserves greater research attention. J Health Serv Res Policy 2015 Jan 17;20(1):62-64. [doi: 10.1177/1355819614543161] [Medline: 25038059]

19. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
20. Sabater J, Sierra C. Review on computational trust and reputation models. Artif Intell Rev 2005 Sep;24(1):33-60. [doi: 10.1007/s10462-004-0041-5]
21. Hui CY, McKinstry B, Walton R, Pinnock H. A mixed method observational study of strategies to promote adoption and usage of an application to support asthma self-management. J Innov Health Inform 2019 Jan 9;25(4):243-253 [FREE Full text] [doi: 10.14236/jhi.v25i4.1056] [Medline: 30672405]
22. Corritore CL, Kracher B, Wiedenbeck S. On-line trust: concepts, evolving themes, a model. Int J Hum-Comput Stud 2003 Jun;58(6):737-758. [doi: 10.1016/s1071-5819(03)00041-7]
23. Castelfranchi C, Falcone R. Trust Theory: A Socio-cognitive and Computational Model. United States: John Wiley & Sons; 2010.
24. Taddeo M. The Role of E-trust in Distributed Artificial Systems. Trust and Virtual Worlds. 2011. URL: https://tinyurl.com/72up66fv [accessed 2021-06-29]
25. Wang YD, Emurian HH. An overview of online trust: concepts, elements, and implications. Comput Hum Behav 2005 Jan;21(1):105-125. [doi: 10.1016/j.chb.2003.11.008]
26. Corritore CL, Kracher B, Wiedenbeck S. On-line trust: concepts, evolving themes, a model. Int J Hum-Comput Behav 2003 Jun;58(6):737-758. [doi: 10.1016/s1071-5819(03)00041-7]
27. Sillence E, Briggs P, Harris P, Fishwick L. A framework for understanding trust factors in web-based health advice. Int J Hum-Comput Behav 2006 Aug;64(8):697-713. [doi: 10.1016/j.ijhcs.2006.02.007]
28. Briggs P, Burford B, De Angeli A, Lynch P. Trust in online advice. Soc Sci Comput Rev 2016 Aug 18;20(3):321-332. [doi: 10.1177/089443930202000309]
29. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
30. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
31. Digital, data and technology standards. Digital, Data and Technology Standards. 2020. URL: https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework [accessed 2021-02-28]
32. Application to Risk Management. ISO. URL: https://www.iso.org/obp/ui/#iso:std:iso:14971:ed-3:v1:en [accessed 2021-02-28]
33. Quality Management Systems. International Organization for Standardization. 2016. URL: https://www.iso.org/standard/59752.html [accessed 2021-02-28]
34. Medical Device Software. International Organization for Standardization. 2006. URL: https://www.iso.org/standard/38421.html [accessed 2021-02-28]
35. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manage Inf Syst 2011 Jun;2(2):1-25. [doi: 10.1145/1985347.1985353]
36. Castelfranchi C, Falcone R. Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification. In: Proceedings International Conference on Multi Agent Systems. 1998 Presented at: IEEE'98; July 3, 1998; Paris, France. [doi: 10.1109/icmas.1998.699034]
37. Ferrario A, Loi M, Viganò E. In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. Philosophy & Technology. URL: https://doi.org/10.1007/s13347-019-00378-3 [accessed 2021-02-28]
38. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mhealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
39. Hui CY, Walton R, McKinstry B, Pinnock H. Time to change the paradigm? A mixed method study of the preferred and potential features of an asthma self-management app. Health Informatics J 2020 Jun;26(2):862-879 [FREE Full text] [doi: 10.1177/1460458219853381] [Medline: 31195877]
40. Quality and Outcomes Framework Guideline. NHS. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/05/gms-contract-qof-guidance-april-2019.pdf [accessed 2021-02-28]
41. Scottish Health Research Register. NHS. URL: https://www.registerforshare.org/ [accessed 2021-02-28]
42. Register for Asthma Research. AUKCAR. URL: https://www.aukcar.ac.uk/reach [accessed 2021-02-28]
43. Primary Care Respiratory Update. URL: https://www.pcrs-uk.org/ [accessed 2021-02-28]
44. NVivo Version 12. QSR International: Technology & Software Solutions. URL: https://www.qsrinternational.com/nvivo/support-overview/downloads [accessed 2021-02-28]
45. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117-118 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
46. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020 Feb 3;3(1):14-16 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]

47. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. J Med Internet Res 2019 Apr 5;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
48. Trusted AI. IBM. 2018. URL: https://www.research.ibm.com/artificial-intelligence/trusted-ai/ [accessed 2021-02-28]
49. Ethically Aligned Design, A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. IEEE. 2019. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined [accessed 2021-02-28]
50. Good AI Governance. AI4Peoples Scientific Committee. 2019. URL: https://www.eismd.eu/wp-content/uploads/2019/11/AI4Peoples-Report-on-Good-AI-Governance_compressed.pdf [accessed 2021-02-28]
51. Sethumadhavan A. Trust in artificial intelligence. Ergo Design 2018 Dec 10;27(2):34. [doi: 10.1177/1064804618818592]
52. Belanche D, Casaló LV, Flavián C, Schepers J. Trust transfer in the continued usage of public e-services. Info Manag 2014 Sep;51(6):627-640. [doi: 10.1016/j.im.2014.05.016]
53. Dong C, Russello G, Dulay N. Trust Transfer in Distributed Systems. In: Conference proceeding: Joint iTrust and PST Conferences on Privacy, Trust Management and Security. 2007 Presented at: IFIP'07; July 30, 2007; Boston, MA. [doi: 10.1007/978-0-387-73655-6_2]
54. Stewart KJ. Trust transfer on the world wide web. Organ Sci 2003 Feb;14(1):5-17. [doi: 10.1287/orsc.14.1.5.12810]
55. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
56. Haring K, Matsumoto Y, Watanabe K. How Do People Perceive and Trust a Lifelike Robot. In: Conference Proceeding: World Congress on Engineering and Computer Science. 2013 Presented at: CECS'13; May 6-9, 2013; San Francisco, USA.
57. 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 8;7(1):e12269 [FREE Full text] [doi: 10.2196/12269] [Medline: 30622092]
58. Ackerman SL, Gleason N, Shipman SA. Comparing patients' experiences with electronic and traditional consultation: results from a multisite survey. J Gen Intern Med 2020 Apr;35(4):1135-1142 [FREE Full text] [doi: 10.1007/s11606-020-05703-7] [Medline: 32076987]
59. Lin Y, Huang G, Ho Y, Lou M. Patient willingness to undergo a two-week free trial of a telemedicine service for coronary artery disease after coronary intervention: A mixed-methods study. J Nurs Manag 2020 Mar 21;28(2):407-416. [doi: 10.1111/jonm.12942] [Medline: 31891424]

Abbreviations

AI: artificial intelligence
GP: general practitioner
IoT: internet-of-things
NHS: National Health Service
PRISMS: Practical Reviews in Self-Management Support

Edited by L Buis; submitted 04.09.20; peer-reviewed by G Sorwar, L Seuren, L Kaye; comments to author 31.10.20; revised version received 28.02.21; accepted 28.04.21; published 16.07.21.

Please cite as:
Hui CY, McKinstry B, Fulton O, Buchner M, Pinnock H
Patients' and Clinicians' Perceived Trust in Internet-of-Things Systems to Support Asthma Self-management: Qualitative Interview Study
JMIR Mhealth Uhealth 2021;9(7):e24127
URL: https://mhealth.jmir.org/2021/7/e24127
doi:10.2196/24127
PMID:34269684

Original Paper

Wound Image Quality From a Mobile Health Tool for Home-Based Chronic Wound Management With Real-Time Quality Feedback: Randomized Feasibility Study

Jia Zhang[1], MSc; Carina Mihai[2], MD; Laura Tüshaus[1], PhD; Gaetano Scebba[1], MSc; Oliver Distler[2], MD; Walter Karlen[1], PhD

[1]Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[2]Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

Corresponding Author:
Jia Zhang, MSc
Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems
Department of Health Sciences and Technology
ETH Zurich
BAA, Lengghalde 5
Zurich, 8008
Switzerland
Phone: 41 44 510 72 28
Email: jia.zhang@hest.ethz.ch

Abstract

Background: Travel to clinics for chronic wound management is burdensome to patients. Remote assessment and management of wounds using mobile and telehealth approaches can reduce this burden and improve patient outcomes. An essential step in wound documentation is the capture of wound images, but poor image quality can have a negative influence on the reliability of the assessment. To date, no study has investigated the quality of remotely acquired wound images and whether these are suitable for wound self-management and telemedical interpretation of wound status.

Objective: Our goal was to develop a mobile health (mHealth) tool for the remote self-assessment of digital ulcers (DUs) in patients with systemic sclerosis (SSc). We aimed to define and validate objective measures for assessing the image quality, evaluate whether an automated feedback feature based on real-time assessment of image quality improves the overall quality of acquired wound images, and evaluate the feasibility of deploying the mHealth tool for home-based chronic wound self-monitoring by patients with SSc.

Methods: We developed an mHealth tool composed of a wound imaging and management app, a custom color reference sticker, and a smartphone holder. We introduced 2 objective image quality parameters based on the sharpness and presence of the color checker to assess the quality of the image during acquisition and enable a quality feedback mechanism in an advanced version of the app. We randomly assigned patients with SSc and DU to the 2 device groups (basic and feedback) to self-document their DU at home over 8 weeks. The color checker detection ratio (CCDR) and color checker sharpness (CCS) were compared between the 2 groups. We evaluated the feasibility of the mHealth tool by analyzing the usability feedback from questionnaires, user behavior and timings, and the overall quality of the wound images.

Results: A total of 21 patients were enrolled, of which 15 patients were included in the image quality analysis. The average CCDR was 0.96 (191/199) in the feedback group and 0.86 (158/183) in the basic group. The feedback group showed significantly higher (P<.001) CCS compared to the basic group. The usability questionnaire results showed that the majority of patients were satisfied with the tool, but could benefit from disease-specific adaptations. The median assessment duration was <50 seconds in all patients, indicating the mHealth tool was efficient to use and could be integrated into the daily routine of patients.

Conclusions: We developed an mHealth tool that enables patients with SSc to acquire good-quality DU images and demonstrated that it is feasible to deploy such an app in this patient group. The feedback mechanism improved the overall image quality. The introduced technical solutions consist of a further step towards reliable and trustworthy digital health for home-based self-management of wounds.

(JMIR Mhealth Uhealth 2021;9(7):e26149) doi:10.2196/26149

KEYWORDS

data quality; remote assessment; digital ulcers; scleroderma; mobile app; digital health; ehealth; mhealth; telemedicine; teledermatology

Introduction

Background

Chronic wounds do not heal within the expected time and can lead to severe complications if not treated appropriately. Therefore, chronic wounds require stringent management, including regular observation, assessment, documentation, and care of the wound by a medical professional. This management is tedious for patients as they need to visit the clinic frequently for specialist consultation, wound documentation, and adjunct therapies [1]. Furthermore, it strains the health care system, as it requires considerable clinical as well as financial resources. Repetitive wound documentation is tedious but essential to chronic wound care [2]. Treatment plans and scheduling of follow-up assessments are based on the information found in the wound documentation. Wound information such as wound location, depth, size, edges, and surrounding skin conditions is often documented, but can vary in detail among different clinical practices [2]. Therefore, the introduction of digitalization of chronic wound management to reduce the burden for patients, clinicians, and the health care system is desirable.

As part of the digitalization of chronic wound care, remote assessment of chronic wounds using mobile technology is raising great interest. Telemedicine approaches could reduce the constraints of time and geographical location and therefore reduce trips to clinics. It has been shown that connecting home-care nurses to hospital-based wound experts can significantly improve the likelihood of wound healing [3] and patient outcomes [4]. Patients sending wound images and symptom questionnaires enables remote follow-up monitoring for post-surgical wounds [5]. Smartphone apps with visual wound analytics and feedback to engage patients in self-care of diabetic foot ulcers resonated positively with the users in a usability study [6]. This study did not evaluate the accuracy of the automated wound image analysis. However, it highlighted the need for user-friendly image acquisition methods that assist the user in controlling the factors influencing the wound image quality.

Wound images constitute an essential part of chronic wound documentation during routine clinical assessment. Clinicians consult the images to assess wound changes [7]. Wound images are also used to determine wound dimensions and tissue conditions by measuring the wound area and color change over time [8]. However, current documentation approaches depend largely on routines established in the clinics, which are rarely standardized across institutions. Wound images are primarily taken by clinical staff with digital cameras and manually uploaded to the clinical information system to perform basic assessments, such as checking the wound history [7] and measuring the wound area [9].

The quality of wound images is essential for further analysis and processing. In clinical practice, high-quality images, characterized by attributes such as proper lighting condition and wound positioning, are prerequisites for wound experts to perform the wound assessment accurately and reliably, including visually inspecting the development of wound status and measuring the wound area from an image [10]. For automated image analysis, high-quality images, with characteristics such as proper lighting condition, corrected color, adequate sharpness, clear wound boundary, and often, the use of an associated color and size reference, are important for algorithms to perform specific tasks such as wound segmentation [11] or wound classification [12] with good performance. Specifications of cameras on modern smartphones are sufficient to be used in clinical practice [13]. However, image quality is largely based on environmental conditions and the person who captures the image. In mobile health (mHealth) applications, remote sensing and documentation by nonexperts are unsupervised and often prone to noise and artifacts [14]. For example, the reliability and accuracy of a teledermatoscopy-based diagnosis increased when the image quality improved [15]. Specifically, it is important to reduce blurriness and keep environmental conditions such as image angle and lighting consistent without over- or underexposure when capturing wound images, so that the wound size can be reliably calculated and the colors of wound areas can be correctly defined [16]. It is challenging to control the aforementioned conditions in a remote setting and with smartphones, especially for patients with little prior knowledge of technical and clinical requirements for wound images. Neither current clinical practices nor existing apps for remote wound monitoring have standardized procedures implemented for taking high-quality wound images. To our knowledge, no quality assurance measures for image quality have been proposed. In studies, low-quality images that are not useful are simply discarded and excluded from the analysis [16]. Therefore, it is crucial for mHealth systems that support remote documentation of chronic wounds to facilitate the high-quality acquisition of images that is sufficient for both visual clinical interpretation and automated processing.

Our goal was to develop an mHealth tool to facilitate the remote assessment of digital ulcers (DUs) and support patients with systemic sclerosis (SSc) in the self-management of their DUs. The tool should provide functions to self-document their wounds at home, including an image acquisition system that can ensure clinical-grade data quality. Such a system would enable telemedical functions for clinicians and reduce the burden of travel to the clinics for the patients.

Systemic Sclerosis and Digital Ulcers

SSc is a systemic autoimmune disease characterized by fibrosis and microangiopathy. DUs are common in SSc, with an occurrence probability of up to 70% at a 10-year observation period [17]. Between one-third and two-thirds of patients with

SSc develop recurrent DUs [18]. DUs are slow to heal and may complicate with bone infection and amputation, which significantly influence the patient's quality of life and hand function [19]. The management of SSc-associated DUs often requires repeated presentation in a specialized clinic for assessment and adapted treatment, which may include wound debridation and specific topical measures. For this, patients face considerable burden from traveling to the clinics at a fixed time interval, and a timely assessment is not always achieved. There is a great unmet need for patients with SSc and DUs to be able to document their wounds remotely, in order to facilitate close follow-up by clinicians, enabling early detection of complications and reducing the burden induced by unnecessary travel to the clinic.

Digital Health Tools for Wound Assessment

Many studies have investigated digital health tools for the assessment of wounds. Overall, these studies highlight the feasibility of remote wound care and self-documentation. However, only one study has evaluated the remote assessment of DUs so far [20]. Patients with SSc were using their own smartphone camera to take images of the DUs for a maximum of 35 days at home. While the study demonstrated the feasibility of home-based documentation of DUs, it was limited to 4 patients. No information on image quality was obtained, and images were manually collected and transferred to a database.

A large proportion of existing wound imaging and documentation apps focus on the assessment of diabetic foot ulcers. Yap et al [21] presented a mobile app aiming to standardize diabetic foot images by creating an outline of a precaptured foot image and aligning the foot when taking a new image. The images were taken in a clinical environment and it was up to the medical professional to decide whether to save the image, which might be challenging to translate to patient self-assessment. Patients are confronted with high uncertainty about the status of their own wound and might not be aware of the important parameters to control during wound image acquisition. Several apps integrate semi-automated wound measurement algorithms providing wound size and diameter [6,9]. However, the wound measurements require manual localization of the wound. Little evidence exists on the accuracy of such systems, due to the absence of validation studies on real wounds. To test interrater variability, plastic wound models have been used [9].

Besides an image-capturing app, supporting infrastructure is needed to enable ease of use and good-quality imaging results. The use of reference markers for color and size normalization is common [9]. Wang et al [22] developed a capturing box to capture wounds on feet that are otherwise not easily accessible. The box contains a mirror, and patients place their foot on a glass plate next to the phone that captures the mirror reflection inside the box. Alternatively, the front-facing camera can be used together with voice commands [6]. Neither of these studies investigated objective image quality of the obtained wound images, in particular when obtained by a nonmedically trained user.

We developed an mHealth tool to enable wound self-documentation by patients with SSc and DUs. Our specific aims of this study were to (1) define objective measures for assessing the quality of wound images taken by patients using this tool unsupervised at home, (2) compare whether an automated real-time feedback feature on wound image quality improves the quality of the transmitted wound images, and (3) evaluate the feasibility of implementing this mHealth tool for home-based chronic wound monitoring and self-management by patients with SSc.

Methods

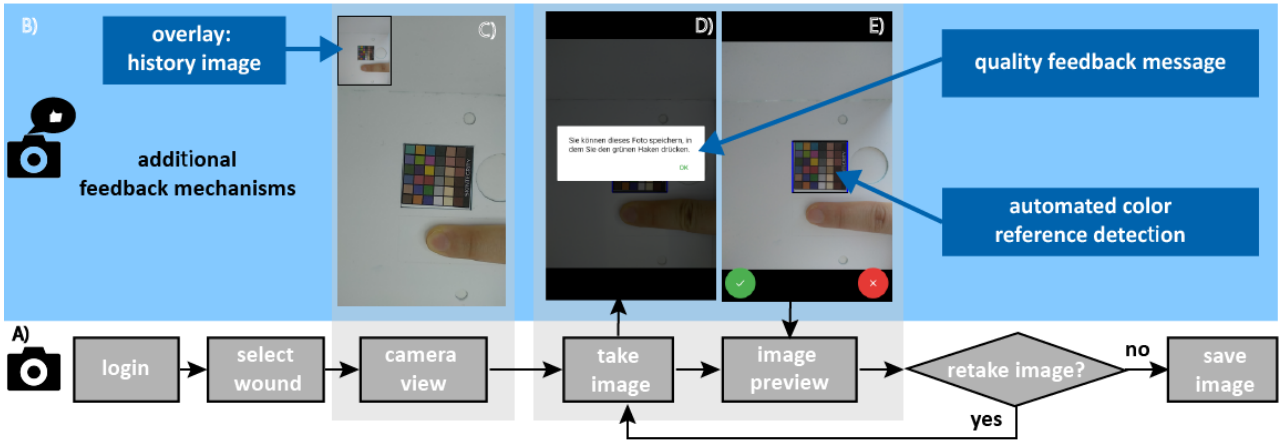
The newly developed mHealth tool was composed of a smartphone with a custom wound imaging and management app, a custom color reference sticker, and a smartphone holder. All components were designed to facilitate standardized wound image acquisition that is consistent over time and does not require supervision or advanced training of the user. With a randomized study design in a home-based setting, we tested and compared 2 versions of the smartphone app (basic and with feedback) that differed in the way the app interacts with the patient during image acquisition. The version with feedback relied on real-time algorithms to assess image quality.

mHealth Tool Development

Wound Management App

The smartphone app implemented a simple workflow to take wound pictures (Figure 1). It essentially consisted of a first-time patient login; first-time metadata entry (ie, wound locations), which was replaced with a selection in subsequent use; image acquisition; and background process that included secure upload to a REDCap study database [23] via an application programming interface (API) when an internet connection was available. The app was developed for the Android operating system. Two versions of the app were developed with the goal of improving the overall image quality. The basic app version did not provide any feedback on image quality other than showing the freshly acquired image for subjective review, which is native to the Android operating system. The version with feedback included specifically developed mechanisms to interact with the user and communicate quality improvement opportunities at the time of image acquisition (Figure 1).

Figure 1. Flowchart of the interaction steps for a user for (A) the basic app and (B) additional feedback mechanisms for the version with feedback, which is illustrated with screenshots showing (C) a history image in the left corner of the screen to guide patients to create consistent fields of view across assessments, (D) feedback messages based on the automated detection of image quality, and (E) automated color reference detection to determine the image quality.



Quality Feedback Mechanisms

For developing the feedback mechanisms, an objective measure for image quality was needed. We focused on the evaluation of image sharpness as it is a parameter that also measures how well the object of interest is in focus and has traditionally been an important parameter in image quality evaluation in other areas of application. Sharper images contain finer details and lead to more reliable object recognition and classification [24]. Blurry images negatively influence the performance of automated image processing algorithms [25]. In addition, it is recommended to discard blurry wound images and not use them for wound assessment, as they could mislead diagnosis or alter derived conclusions [16]. Therefore, it was desirable to identify unsuitable images early in the process and encourage repetition of the image acquisition at the instance where the user is already engaged with the process. For this purpose, we defined the sharpness as the variance of the Laplacian [26]. The Laplacian is a second derivative operator to high-pass spatial frequencies, which are associated with sharp edges, and it is sensitive to the rapid intensity changes in an image [26]. The variance reflects the spread of intensity changes. Thus, sharper images have higher Laplacian variance.

An additional objective measure for quality available in our system was the desired presence of the color reference sticker in the wound image. Therefore, we implemented an automatic color reference sticker detection routine into the imaging process. First, the color checker position was detected in the field of view by analyzing the rectangular morphology and exceeding a predefined threshold of its size, and then the sharpness of the sticker was analyzed. When the color reference sticker was detected and the corresponding region of interest was deemed sharp enough by exceeding a predefined threshold, the app provided the patient with a feedback message indicating that the image is deemed to be of good quality and can be saved. Otherwise, the user was prompted to retake the image.

As a third feedback mechanism, we provided a subjective comparison between the current image preview and a previously taken image of the same wound. The historical image was displayed in the left corner of the image preview screen. The intention of this was to encourage the patient to maintain a consistent field of view across multiple capturing sessions.

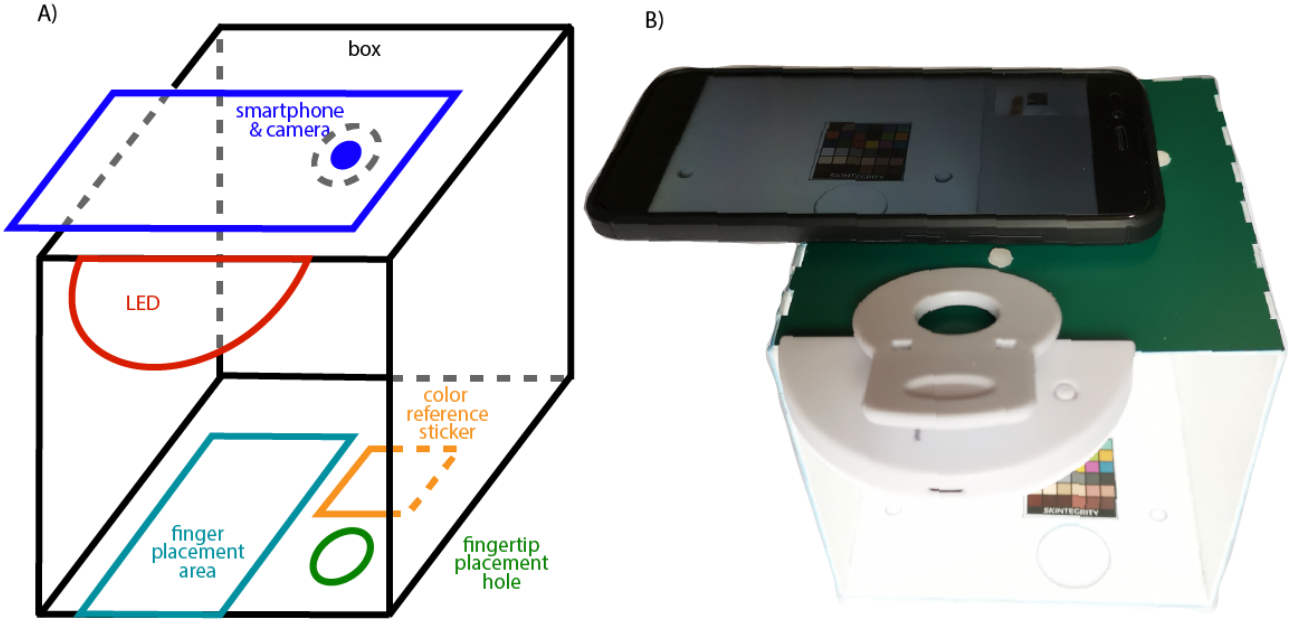
Color Reference Sticker

We developed a customized color reference sticker as part of a larger project named SwissWOU, which is a subproject of the SKINTEGRITY Flagship of the Universities and Hospitals of Zurich, with the aim of developing a large wound image database for wound research. The color reference sticker contained 36 color patches that included all the colors for regular photography color calibration [27] and an additional set of skin and wound-specific color shades (total size 30 x 30 mm).

Smartphone Holder

Patients with SSc often have severely reduced hand and finger function [19], limiting the range of motion and possibilities to interact with a smartphone. To ease the image-taking procedure and address the patient's needs, we designed a smartphone holder (Figure 2). Patients placed the smartphone on top of the smartphone holder. Their finger was placed on the bottom plate or, in the case of fingertip wounds, through a hole in the bottom plate. The color reference sticker was also placed on the bottom plate adjacent to the finger placement area. An additional LED light ring assured a homogeneous lighting condition and prevented strong shadows. The smartphone holder provided a constant distance between the camera and the wound and consistent illumination, thereby also contributing to the consistency of image quality across time series.

Figure 2. (A) 3D schematic of the mobile health tool and (B) experimental setup.

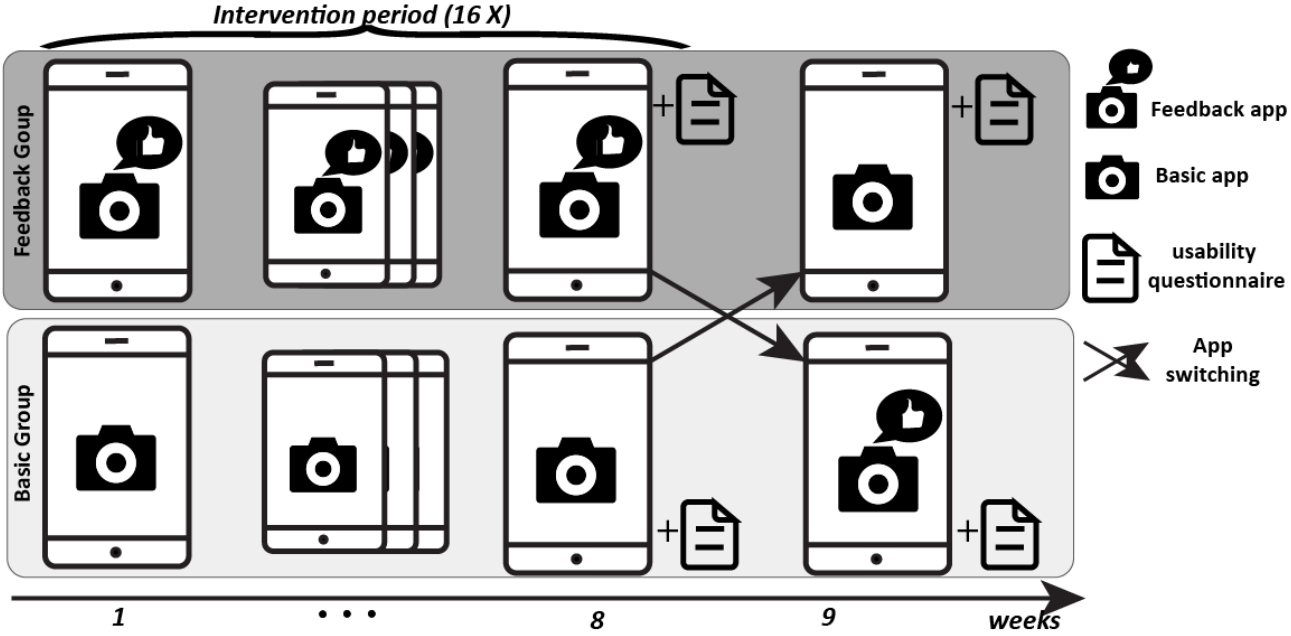


Experimental Protocol

The study protocol followed the ethical principles for research involving humans according to the Declaration of Helsinki and was approved by the institutional ethics committee (ETH EK 2019-N-22). We recruited consecutive patients with SSc who attended the Department of Rheumatology at the University Hospital Zurich, Switzerland, for routine or emergency consultations from May 2019 to May 2020. Adult patients fulfilling classification criteria for SSc [28] and having at least one DU on the fingers were included after written informed consent. First, patients performed a practical baseline test and filled in a questionnaire on mobile literacy and familiarity with smartphone camera usage. We identified patients who were familiar with using basic smartphone functions and also patients who were accustomed to photography. Based on this test, we stratified patients into experienced users and nonexperienced users, based on the combined criteria of whether they used the smartphone multiple times a day or used it at least once a day and were familiar with taking photos or videos. We randomly assigned the patients to 2 device groups (feedback and basic groups), while balancing the user experience between the groups. After this randomization, all patients were given an Android smartphone with all required apps pre-installed and the smartphone holder to keep for the duration of the study. We then instructed them on how to use the mHealth tool, explained the study protocol, and walked them step by step through the image-taking process of the corresponding app version (feedback or basic) once. Questions from patients were then answered. Under supervision, the patients then used the tool by themselves to take the first set of images of their wounds.

The data acquisition at home followed a regular protocol (Figure 3). Both groups were asked to perform the wound assessments with the app every third day for the duration of 8 weeks (16 assessments). The assessment consisted of taking wound images and completing pain level questionnaires. The patient had to first login to unlock the app, place the smartphone on the smartphone holder, turn on the LED light ring, select the wound location, take an image, and if it was deemed of good quality, save the image or otherwise retake an image. After completing the wound image capture, patients evaluated their pain levels. At the end of the 8th week, data collection for the randomized app and image quality evaluation was completed, and a usability questionnaire was delivered. For an additional week on every third day (2-3 assessments), patients got to explore and test the alternate app version from the other group and evaluate the usability of this version as well. At the end of the study, patients returned the mHealth tool by mail.

Figure 3. Overview of the wound assessment protocol performed at the patient's home for both the feedback and basic groups, which involved wound assessments every third day (16 times in total) for the first 8 weeks, after which, patients completed the usability questionnaires. In week 9, patients switched to the alternative app version and performed image assessments every third day (2-3 times), after which they evaluated the usability.



Analysis

Image Quality Comparison

To evaluate and compare the image quality between the randomized groups, we calculated the following image quality parameters.

We defined the color checker detection ratio (CCDR) as the number of images with clearly visible color reference stickers detected over the total number of images collected (Equation 1). We calculated the CCDR across each group and patient

We compared the sharpness of images between groups, where the variance of the Laplacian was selected as a measure for sharpness [26]. Since the variance of the Laplacian is dependent on the image content and the variability of wounds across subjects was not negligible, we restricted the region of interest to the color reference sticker, which was expected to be available and constant across all wound images. We evaluated the color checker sharpness (CCS) of each image where the color reference sticker area was manually labeled by a researcher after the data acquisition. The median CCS for each patient was evaluated to compare between groups.

In addition to the above-calculated quality measurements, subjective image quality was obtained. A research assistant was trained to evaluate the quality of wound images. She labeled all the images blinded to the randomized groups and applied a binary label. The binary label indicated whether the image was usable for the unambiguous identification of the finger and wound area. The ratios of usable images across each group and patient were calculated.

Statistical Analysis of Image Quality Comparison

We performed a Wilcoxon rank-sum test to compare whether there were statistically significant differences in the image quality between the 2 groups for the images that were collected from the first 8-week documentation period. The analysis was carried out with the Python SciPy library (version 1.4.1) [29]. A P value <.01 was considered statistically significant.

Feasibility Evaluation

The feasibility of the mHealth intervention was determined by analyzing the usability feedback from the questionnaires, the user behavior throughout the study, and the overall quality of wound images as described in the previous section.

The patients rated the usability of the mHealth tool with a questionnaire after each type of use (basic or feedback app) had ended. The questionnaires consisted of questions on the overall mHealth tool experience and the subcomponents like the smartphone holder and the versions of the app. The usability questions originated from the Post-Study System Usability Questionnaire [30], which is widely used to measure users' perceived satisfaction, such as software, system, or product at the end of a study, and were adapted to the specific functions of our application. Additionally, we asked questions about whether the patients were willing to continue using the tool and which app version they preferred. Each question was answered on a 7-point Likert scale, where 1 indicated "strongly agree" and 7 indicated "strongly disagree." For analysis, we aggregated the answers in bar plots.

The monitoring of user behavior was focused on whether the patients (1) were able to self-document their wound images and follow the study protocol with good adherence, (2) were efficient in using the app to capture images, and (3) dropped out of the study. The self-documentation was evaluated based on the patients' compliance with the image documentation routine

given by the protocol. We calculated the number of image capturing assessment days actually performed divided by the assessment days that were assigned to be performed during the 8 weeks (16 assessments). When a patient took more days of assessments than expected (eg, ratio higher than 1), we then assigned the ratio as 1. We evaluated the efficiency by calculating the duration of each assessment by measuring the time between starting the app and the end of the assessment. For patients who documented multiple wounds, we only measured the time until saving the first wound image. We eliminated duration outliers from the analysis as they may have resulted from other distractions unrelated to the imaging task. Outliers were defined as a measurement duration of more than 3 scaled median absolute deviations away from the median duration over all assessments from each patient. For each group, we then calculated median durations and fitted a linear trend over time to evaluate the assessment efficiency.

Results

A total of 21 patients were recruited and randomized into feedback (10 patients) and basic (11 patients) groups. During the course of the study, 6 patients returned the mHealth tool before completing the protocol and were subsequently excluded from the image quality analysis (Table 1). All dropouts happened in the first 2 weeks (Table 2) and were by female participants (Table 1). They had balanced strata (Table 1), and 2 belonged to the feedback group and 4 to the basic group (Table 2). The reasons for not continuing with the study were severe illness (3 patients), overwhelmed after instructions (2 patients), and overwhelmed after subsequent usage (1 patient; Table 2). Of the 15 patients who completed the study, 8 were in the feedback group, and 7 were in the basic group. All of the included patients were in the experienced smartphone user group strata (Table 1). A total of 382 wound images were collected during the intervention period, and 24 images were collected during week 9 after switching apps for the second usability evaluation (Table 1).

Table 1. Metadata for the patient randomization, dropouts, and number of images from each group.

Patient characteristics	Intervention period (Weeks 1-8)		Week 9		Dropouts (n=6)
	Feedback (n=8)	Basic (n=7)	Feedback (n=4)	Basic (n=4)	
Age (years), mean (SD)	47 (11)	52 (9)	N/A[a]	N/A	48 (13)
Sex, n					
Male	4	1	N/A	N/A	0
Female	4	6	N/A	N/A	6
Strata ratio[b], n					
Not familiar	0	0	N/A	N/A	3
Familiar	8	7	N/A	N/A	3
Total number of images	199	183	15	9	N/A
Distribution of images per patient, mean (SD)	24.88 (11.21)	26.14 (10.59)	1.88 (1.76)	1.29 (1.39)	N/A
Number of images with color checker detected	191	158	N/A	N/A	N/A
Subjective quality[c], n	180	155	15	7	N/A

[a]N/A: not applicable.

[b]Ratio of the number of patients that were not familiar to those who were familiar with smartphone camera usage.

[c]Number of images manually labeled as good quality by the research assistant.

Table 2. Study dropouts and time since recruitment, by dropout reason.

Patient characteristics	Overwhelmed at instruction (n=2)	Severe illness (n=3)	Overwhelmed during usage (n=1)
Time since recruitment for each patient (days)	0, 0	3, 6, 9	13
Strata ratio[a], n			
Not familiar	1	2	0
Familiar	1	1	1
Group, n			
Feedback	1	0	1
Basic	1	3	0

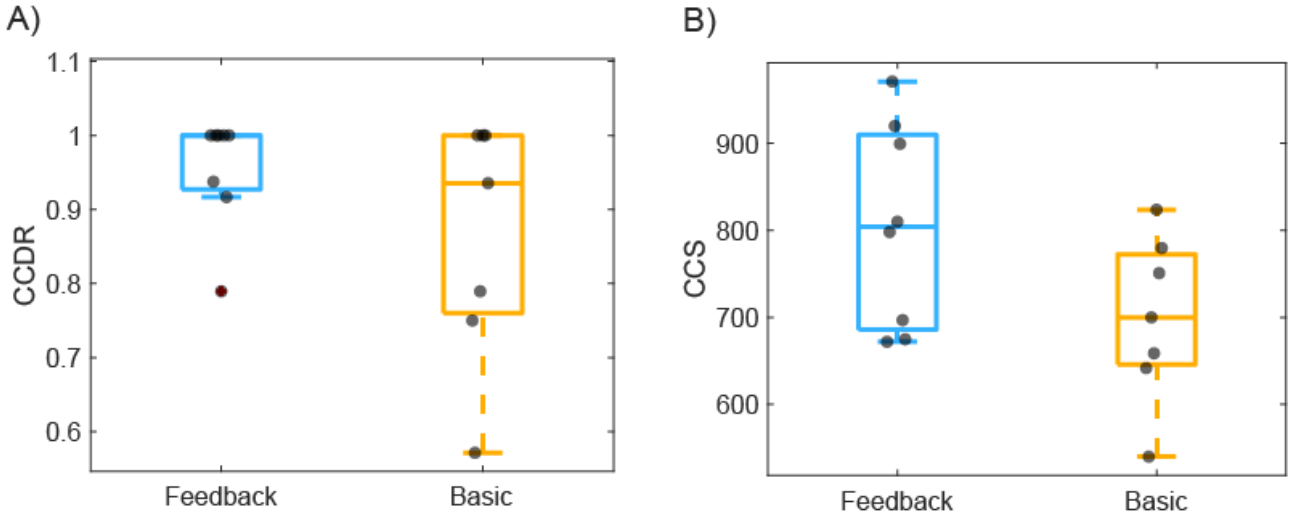
[a]Ratio of the number of patients that were not familiar to those who were familiar with smartphone camera usage.

Image Quality Comparison

Color Checker Detection Ratio

The CCDR was 0.96 (191/199) in the feedback group compared to 0.86 (158/183) in the basic group. As depicted in the boxplot in Figure 4, the feedback group faced less variance (median 1; 25th percentile=0.93; 75th percentile=1) compared to the basic group (median 0.94; 25th percentile=0.76; 75th percentile=1).

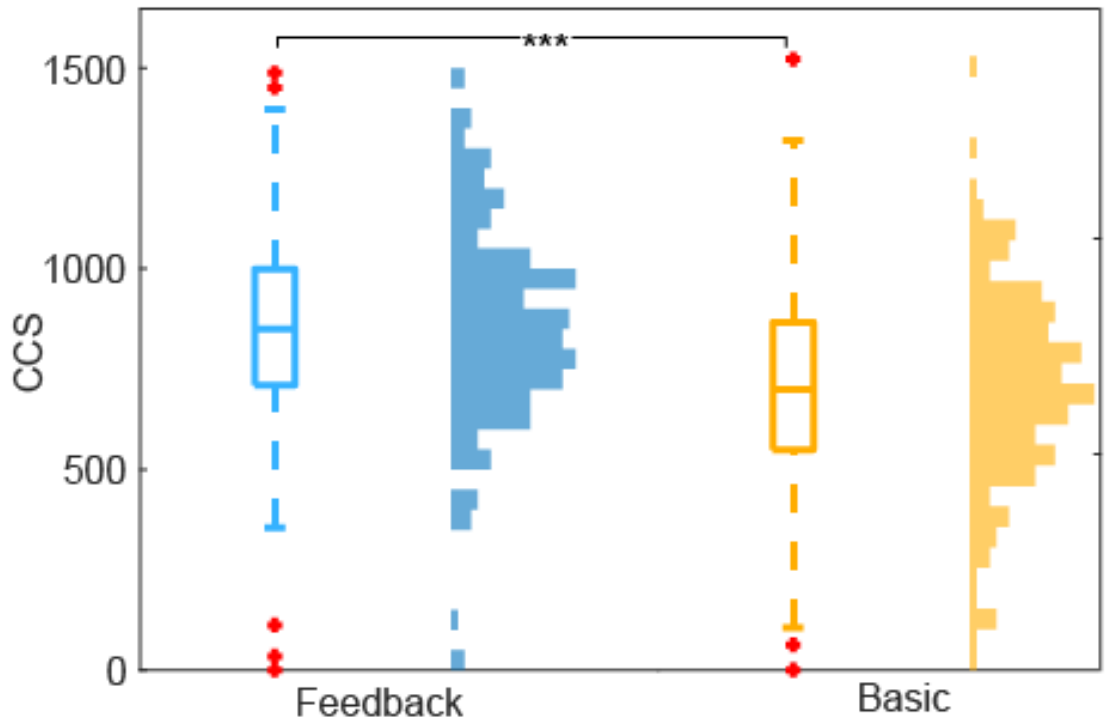
Figure 4. (A) Color checker detection ratio (CCDR) and (B) color checker sharpness (CCS) per patient for the feedback and basic groups. The grey dots indicate each individual patient. The central, bottom, and top edges of the boxes indicate the median, 25th percentile, and 75th percentile, respectively; the whiskers indicate the outliers.



Color Checker Sharpness

The feedback group achieved higher median CCS across patients (median 804) compared to the basic group (median 700; Figure 4). The feedback group also showed overall higher CCS distribution across all images (median 894; 25th percentile=710; 75th percentile=999) compared to the basic group (median 700; 25th percentile=549; 75th percentile=867) as depicted in Figure 5. The Wilcoxon rank-sum test showed a significant difference (P<.001) between the 2 groups.

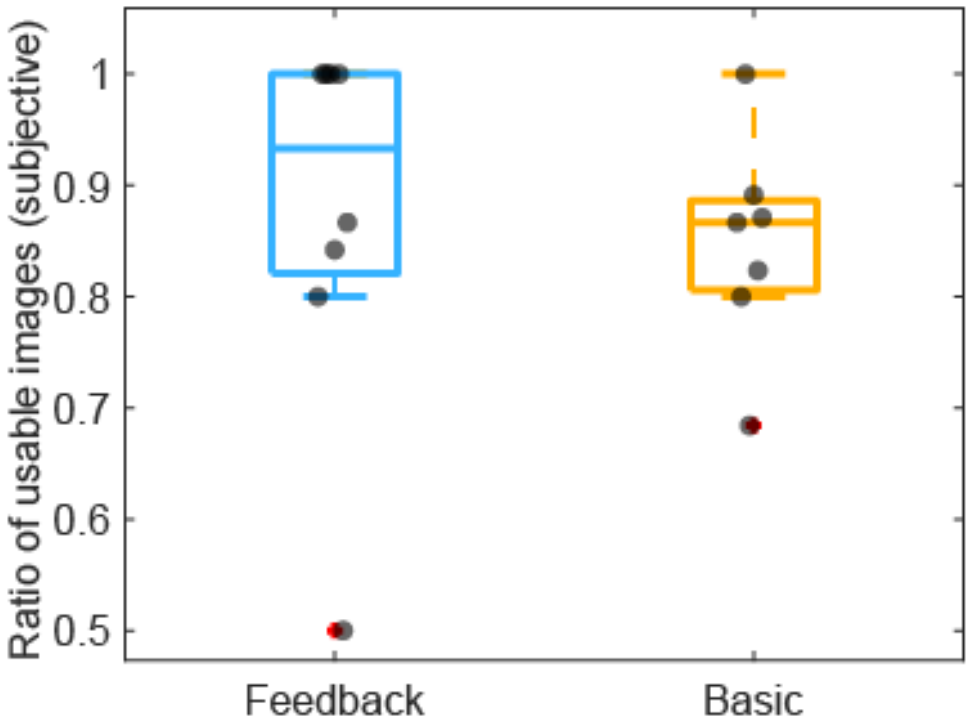
Figure 5. Comparison of the color checker sharpness (CCS) distribution for all images between the feedback and basic groups. The central, bottom, and top edges of the boxes indicate the median, 25th percentile, and 75th percentile. ***P<.001.



Manually Labeled Image Quality

In the feedback group, 90% (180/199) of the images were subjectively labeled as usable compared to 85% (155/183) in the basic group (Table 1). The feedback group had a higher ratio of images per patient that were manually labeled as usable (median 0.93; 25th percentile=0.82; 75th percentile=1) than the basic group (median 0.87; 25th percentile=0.81; 75th percentile=0.88) as depicted in Figure 6.

Figure 6. Distribution of subjective image quality, which was measured as the ratio of images that were labeled as usable, for the manually labeled images for the feedback and basic groups. The grey dots indicate each individual patient. The central, bottom, and top edges of the boxes indicate the median, 25th percentile, and 75th percentile, respectively.



Feasibility Evaluation

The median ratios of the patients' compliance to the image documentation routine were 0.77 (25th percentile=0.59; 75th percentile=0.91) for the feedback group and 0.94 (25th percentile=0.59, 75th percentile=0.98) for the basic group (Figure 7). The overall median compliance ratio for both groups was 0.88 (25th percentile=0.58; 75th percentile=0.94)

The median durations for one image assessment were 42 seconds (overall), 47 seconds (feedback group), and 42 seconds (basic group). The assessment durations showed a decreasing trend for both the feedback and basic groups (Figure 8).

Figure 7. Ratio of patients' compliance to the documentation routine for the feedback and basic groups. The grey dots indicate each individual patient.

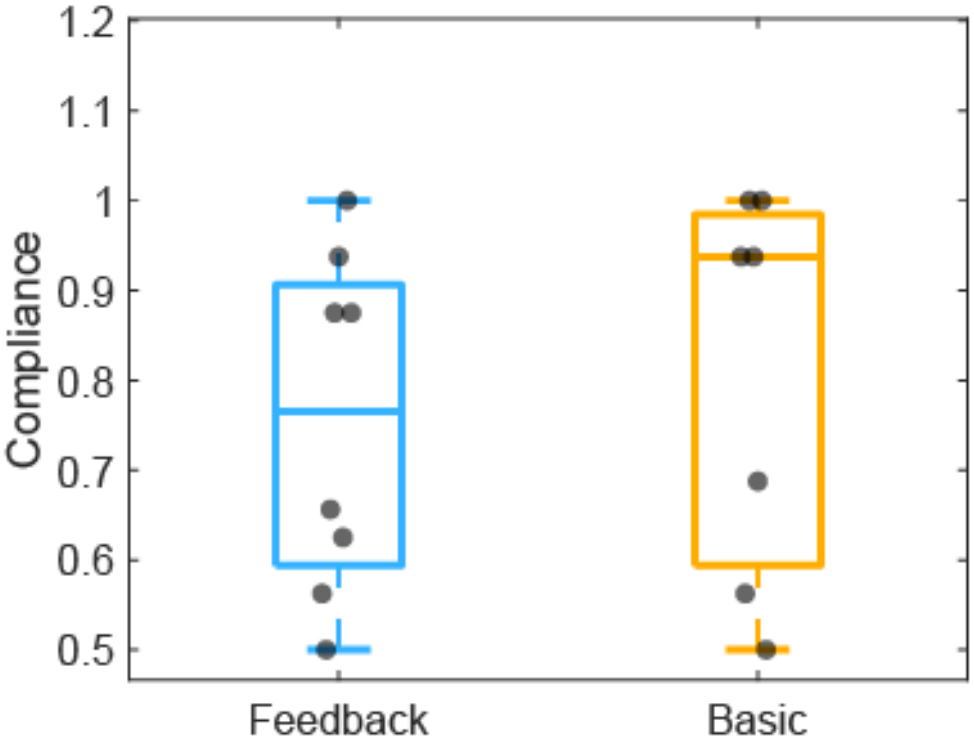
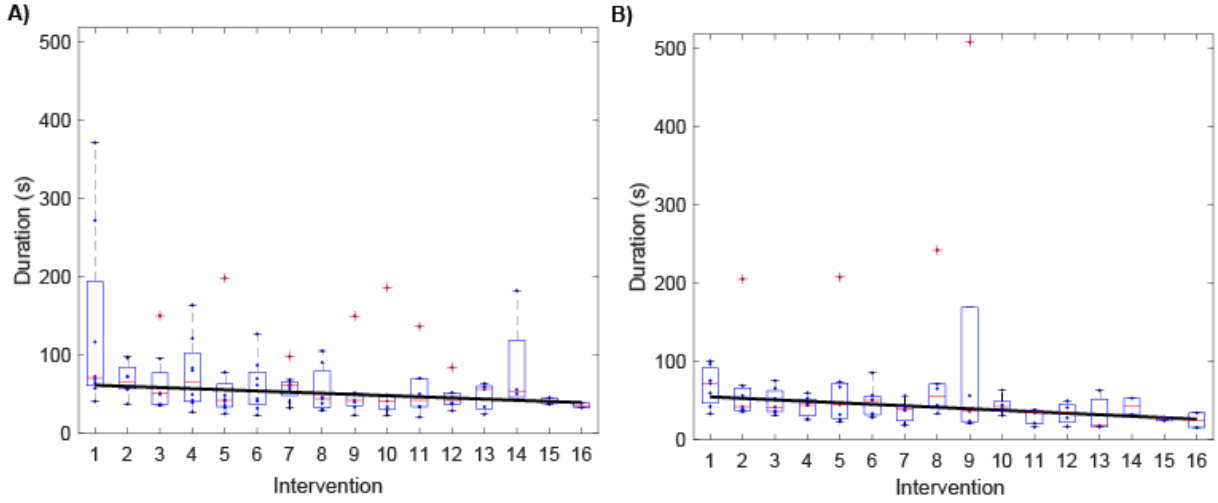
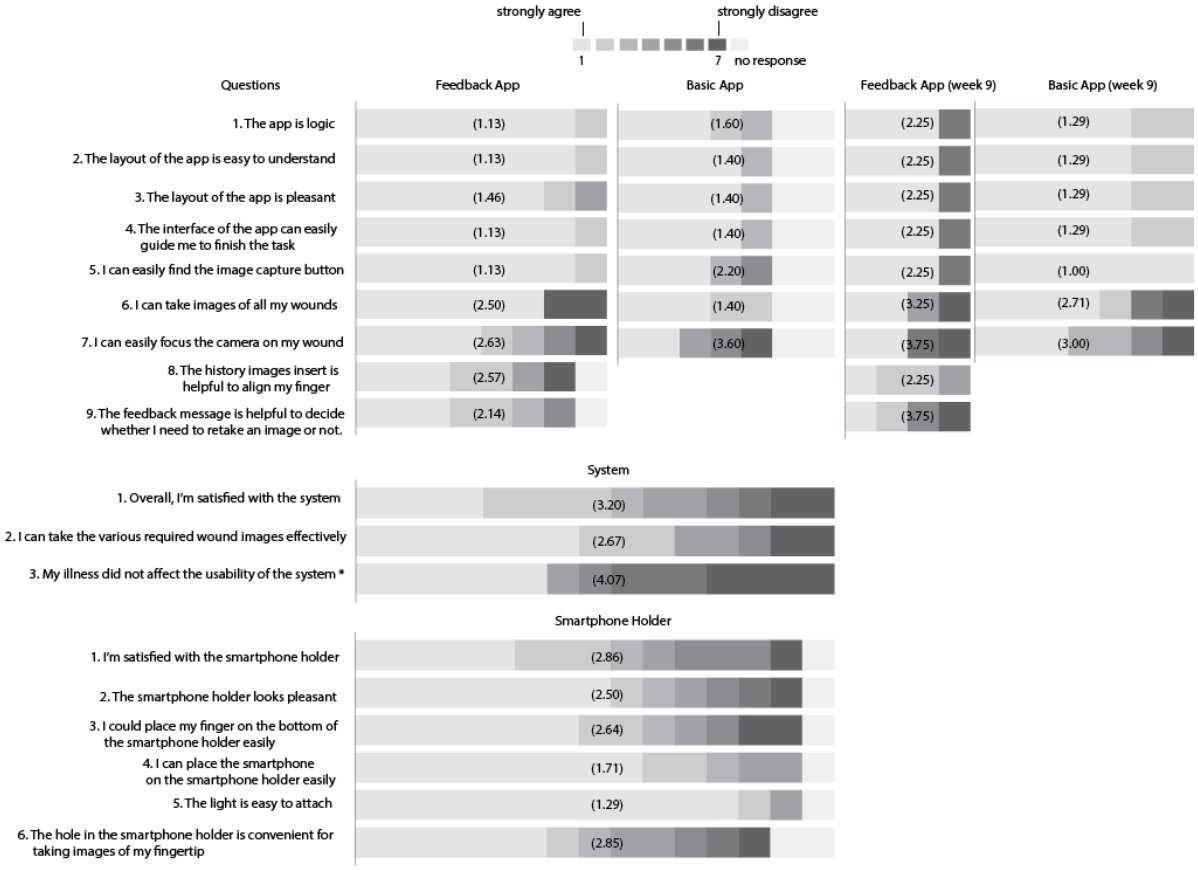


Figure 8. Durations of image assessments over the 8-week study period, consisting of 16 interventions for the (A) feedback and (B) basic groups. The solid lines are the linearly fitted trends per group.



The usability rating showed an overall higher user agreement for the feedback app (range 1.13-2.63) compared to the basic app (range 1.40-3.60) after the 8th week (Figure 9). The smartphone holder obtained an agreement between 1.29 and 2.86 for both groups, and the mHealth tool as a whole was rated between 2.67 and 4.07. At week 9, after switching the app version, the feedback group rated the basic version app (range 1.00-3.00) similarly to the previously used feedback app (Figure 9). The basic group rated the feedback version app lower (range 2.25-3.75) than their scores for the basic version app (Figure 9).

Figure 9. Results of the usability evaluation, in which each question was answered on a 7-point Likert scale (1=strongly agree, 7=strongly disagree). The average rating is shown in brackets on the barplot. *The 3rd question from "System" and its rating were reversed from its original form (Does my illness affect the usability of the system?).



A total of 7 patients indicated an interest in using the system in the future (Figure 10). They provided several reasons:

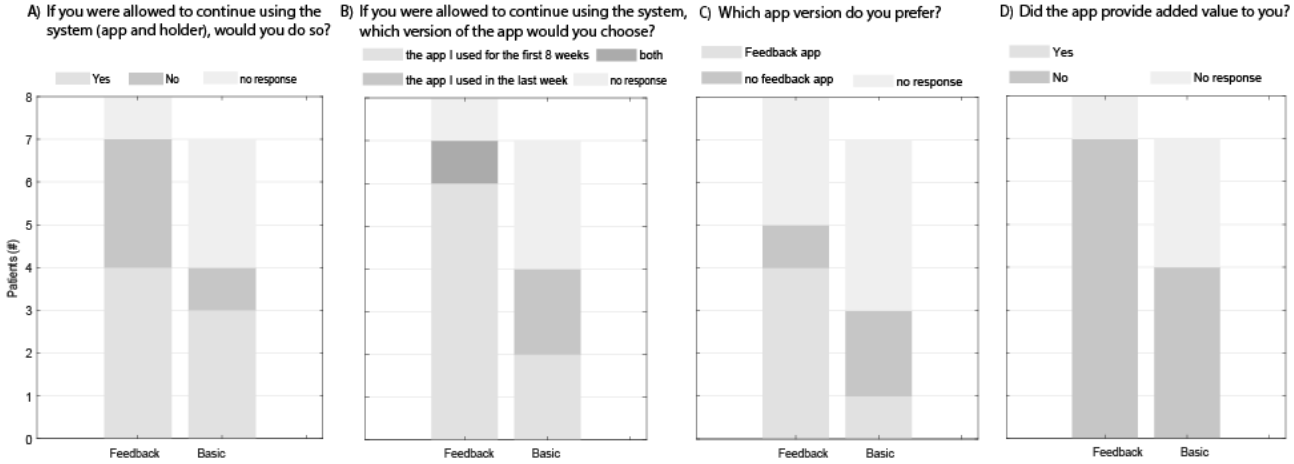
> *…be able to communicate changes without going to the hospital to check.*
>
> *…provide information to support wound healing.*
>
> *…monitor how the wound is changing and to be able to intervene more quickly…*

The responses from the 4 patients who answered "no" were:

> *…only when I feel that my condition is getting worse.*
>
> *I don't see what to use it for.*
>
> *I would find a solution that is easier for me.*
>
> *Because not all wounds can be photographed.*

Of the 8 patients in the feedback group, 6 preferred the feedback app over the basic app (Figure 10), while 2 of 4 patients from the basic group preferred the basic app.

Figure 10. Additional questionnaire with 4 questions about (A) the patients' willingness to continue using the system, (B,C) their preference on the app version, and (D) whether the app added value for them.



Discussion

Principal Findings

We developed an mHealth tool for SSc patients with DUs to remotely self-document wound images in a standardized way and introduced an image quality–based feedback mechanism to encourage higher data integrity through targeted human-machine interaction. To our knowledge, this was the first time that a research study systematically evaluated the quality of patient self-documented wound images with a randomized protocol. We introduced objective image quality parameters such as CCDR and CCS that use the standardized region of interest of a color reference sticker to enable a reproducible quality estimation. These objective image quality metrics, along with subjective expert image assessments, showed that using the feedback mechanism yielded a higher image quality compared to using the basic app without real-time feedback. More specifically, the CCDR, CCS per patient, and manually labeled image quality were higher for the feedback group compared to the basic group. The feedback group achieved significantly higher CCS over all image distributions compared to the basic group. Overall, the high number of good-quality images obtained and the usability evaluation indicated that home-based DU assessment by SSc patients who are familiar with digital tools is feasible.

The importance of taking high-quality wound images for diagnostic and treatment monitoring purposes cannot be emphasized enough. From a technical point of view, modern smartphones are designed to capture high-resolution images with color representations comparable to professional digital cameras [13]. However, capturing images of wounds for clinical applications is more critical and has more stringent requirements, because bad image quality could negatively influence the diagnosis and management plan [15]. With our automatic feedback mechanism that processed images in real-time at the smartphone frontend, we improved the image quality during the process of image collection. This approach increased the reliability and consistency of images and avoided postprocessing or discarding of poor images, therefore also increasing the integrity of the data. Such an increase could positively impact further processing of the images, such as automated classification and segmentation algorithms or assessments of abnormalities during telemedical consultations. Whether the increase in quality has an impact on such applications and can influence patient health outcomes need to be investigated in a separate study.

We expected to observe a benefit from the real-time feedback mechanism. Other digital health applications previously reported such benefits. For example, respiratory rate counters that provide audio and visual feedback enable direct comparison of measurements with the breathing subject. Such comparison leads to improved respiratory rate estimations and repeated measurements where no initial agreement between feedback and subjects can be observed [31]. Furthermore, computing a signal quality index that is displayed in real time with a color coding in the background of a vital sign trace assists the user in recording good-quality pulse oximetry recordings [32]. In fact, specifications in standards for medical pulse oximeters (ISO 80601-2-61) enforce a real-time indication for signal inadequacy [33]. However, standards leave it up to the manufacturer to define the quality indicator metric and its display method. Therefore, it is important to test its performance, usability, and feasibility in a user study as part of the medical device validation and certification process. The proposed randomization of users in 2 groups that use 2 different mHealth tools is a direct way to evaluate which tool can provide better quality data.

The usability questionnaires were another way to evaluate the feasibility and showed that the majority of the patients were satisfied with the system. All questions except one received a rating ≤3.75 on a 7-point Likert scale. The question of whether the usability of the system was influenced by the patient's illness received a rating of 4.07. A closer investigation of this result revealed 2 reasons that should be considered in further designs. First, hand function of some patients was severely impacted by the disease. Two patients brought forward that it was not easy to place their fingers correctly in the smartphone holder box to capture all angles of the wounds. Second, one patient stated that it was more difficult to get a sharp image from specific wound locations. This indicates that there is still room for improvement of the mHealth tool design. For patients whose hand function is severely impaired by the disease or whose fingers are affected on both hands, an obvious solution to avoid additional painful interaction with the system could be assistance from a second person (eg, family member or home care staff) or the integration of voice commands, which has already been proposed for diabetic foot ulcer management [6]. Additional technical developments might be required for this (ie, an algorithm for the automated election of the focus point).

We re-evaluated the usability after switching the app version and using the alternative for another week. Interestingly, the feedback group did provide a similar usability rating for the basic app version, while the basic group had lower usability ratings for the feedback app version. An explanation could be that the basic group did not receive training on the feedback mechanisms, which might have led to confusion and lower usability. This would suggest that adequate training is needed before using the mHealth tools and could lead to better acceptance of the intervention.

We reported 6 patients who dropped out during the course of the study. The dropouts occurred immediately after recruiting or at an early stage of the study. The main reason for dropout was due to severe illness and the need for hospitalization, so that the assessment had to be stopped. It is interesting to note that all 3 subjects that were placed in the "not familiar with smartphone use" strata left the study prematurely. While this is too small a number to draw conclusions, it may indicate that familiarity with smartphone use could be a prerequisite for patients to engage with an mHealth approach to wound documentation. This important parameter, together with low recruiting rates, should be closely monitored in subsequent studies and analyzed to determine whether they indicate that the use of technology might not be suitable for all patient groups or demographics.

We demonstrated that remote self-management of wounds by patients with SSc is feasible. An overall rate of routine documentation of 88% was achieved by patients, indicating good adherence to the suggested protocol. Independent of the app version, the obtained image quality was sufficient for further use (>84% of images got a usable label from the subjective expert assessment). The median image assessment duration was within the range of 50 seconds, indicating that the mHealth tool is efficient to use and can be integrated into the daily routine of a patient. While our analysis was limited to 15 patients, this number was sufficient to highlight the preferences of participants and demonstrate the benefits of real-time feedback on improving wound image quality. The usability results indicated that, with small adaptations and optimizations of the design, satisfaction of the patients could be further improved. With 64% of the patients willing to continue using the current version of the mHealth tool, these adaptations will be well justified.

The technical features of the mHealth tool could also be improved in further design iterations. Currently, the image-taking process was based on the standard camera sensor of the smartphone, which produced 2D RGB images that can be used for documenting the wound and assessing basic wound characteristics such as area and color. More advanced measuring approaches could involve the recording of 3D wound images for wound depth assessment [34] or using additional sensor modalities, such as a thermal or infrared sensor to visualize the perfusion changes in and around a wound [35]. Such modalities would be even more difficult to subjectively assess for quality by a lay user because they are less common and require more advanced user interaction. Therefore, integrating automated quality mechanisms would be of utmost importance also in these approaches. However, different objective metrics will need to be developed.

The impact of a home-based mHealth tool for self-documentation of wounds and telemedical application for patients with SSc was evident. First of all, patients identified that such tools could reduce unnecessary travel to the clinics. In addition, during the COVID-19 pandemic, patients with SSc were considered more vulnerable to infection due to immunosuppressive treatment that dampens the immune system [36]. Therefore, patients with autoimmune diseases were recommended to avoid crowds and unnecessary travel or hospital visits [36]. In such a context, telemedicine consultations could serve as a valid tool for enhancing disease management without additionally endangering the patients. No less important, the transmission of wound images with the mHealth tool could enable clinicians or decision support algorithms to detect abnormal wound conditions, such as new or worsening ulcers, earlier [37], leading to more timely referral and earlier treatment.

Conclusions

There is a great need for mobile solutions to support self-documentation of chronic wounds, as well as to enable clinical-grade wound image acquisition at the patient's home. We developed an mHealth tool that provides such health services and enables telemedical support for patients with SSc and DUs. We demonstrated that it is feasible to deploy such an app in this patient group and high-quality wound images can be consistently acquired. System usability had positive ratings, with room for improvement to address the disease-specific needs of patients. Introducing a feedback mechanism identifying image quality deficiencies and encouraging repetition of the imaging process improved overall image quality when compared to a solution without such a feedback mechanism. The mHealth tool can be further investigated to evaluate the clinical efficacy and effectiveness and establish whether patient outcomes could be improved with this sensor-based telemedicine intervention.

Acknowledgments

We would like to thank all patients who participated in the study and provided valuable feedback. We thank Dr. Rucsandra Cristina Dobrota for helping with patient recruitment, Mayank Patwari and Steffen Schmidt for designing app prototypes, and Ilan Misano for designing prototypes of the smartphone holder. We especially thank Amilen Souto Cortes who performed the manual expert labeling of all wound images.

This project is part of the Hochschulmedizin Zurich Flagship SKINTEGRITY. It was supported by the SCS Swiss Child Support Foundation, ETH Zürich Foundation, the Novartis Foundation (Freenovation), and the Swiss National Science Foundation (150640).

Conflicts of Interest

A patent application has been filed based on this manuscript under EP21175108.6: "A system for recording a high quality wound image and for real-time image quality assessment."

References

1. Othman D. Negative pressure wound therapy literature review of efficacy, cost effectiveness, and impact on patients' quality of life in chronic wound management and its implementation in the United kingdom. Plast Surg Int 2012 May 30;2012:374398-374396 [FREE Full text] [doi: 10.1155/2012/374398] [Medline: 22701169]
2. Khalil H, Cullen M, Chambers H, Steers N, Walker J. Implementation of a successful electronic wound documentation system in rural Victoria, Australia: a subject of collaboration and community engagement. Int Wound J 2014 Jun 19;11(3):314-318 [FREE Full text] [doi: 10.1111/iwj.12041] [Medline: 23418740]

3. Zarchi K, Haugaard VB, Dufour DN, Jemec GBE. Expert advice provided through telemedicine improves healing of chronic wounds: prospective cluster controlled study. J Invest Dermatol 2015 Mar;135(3):895-900 [FREE Full text] [doi: 10.1038/jid.2014.441] [Medline: 25290685]
4. Vowden K, Vowden P. A pilot study on the potential of remote support to enhance wound care for nursing-home patients. J Wound Care 2013 Sep;22(9):481-488. [doi: 10.12968/jowc.2013.22.9.481] [Medline: 24005782]
5. Segura-Sampedro JJ, Rivero-Belenchón I, Pino-Díaz V, Rodríguez Sánchez MC, Pareja-Ciuró F, Padillo-Ruiz J, et al. Feasibility and safety of surgical wound remote follow-up by smart phone in appendectomy: A pilot study. Ann Med Surg (Lond) 2017 Sep;21:58-62 [FREE Full text] [doi: 10.1016/j.amsu.2017.07.040] [Medline: 28794868]
6. Ploderer B, Brown R, Seng LSD, Lazzarini PA, van Netten JJ. Promoting Self-Care of Diabetic Foot Ulcers Through a Mobile Phone App: User-Centered Design and Evaluation. JMIR Diabetes 2018 Oct 10;3(4):e10105 [FREE Full text] [doi: 10.2196/10105] [Medline: 30305266]
7. Hess CT. The art of skin and wound care documentation. Home Healthc Nurse 2005 Aug;23(8):502-513. [doi: 10.1097/00004045-200508000-00010] [Medline: 16103800]
8. Queen D, Harding K. Is wound photography becoming sloppy? Int Wound J 2020 Feb;17(1):5-6 [FREE Full text] [doi: 10.1111/iwj.13302] [Medline: 31930710]
9. Wang SC, Anderson JAE, Evans R, Woo K, Beland B, Sasseville D, et al. Point-of-care wound visioning technology: Reproducibility and accuracy of a wound measurement app. PLoS One 2017;12(8):e0183139 [FREE Full text] [doi: 10.1371/journal.pone.0183139] [Medline: 28817649]
10. Rennert R, Golinko M, Kaplan D, Flattau A, Brem H. Standardization of wound photography using the Wound Electronic Medical Record. Adv Skin Wound Care 2009 Jan;22(1):32-38. [doi: 10.1097/01.ASW.0000343718.30567.cb] [Medline: 19096283]
11. Goyal M, Reeves ND, Rajbhandari S, Yap MH. Robust Methods for Real-Time Diabetic Foot Ulcer Detection and Localization on Mobile Devices. IEEE J. Biomed. Health Inform 2019 Jul;23(4):1730-1741. [doi: 10.1109/jbhi.2018.2868656]
12. Shenoy V, Foster E, Aalami L, Majeed B, Aalami O. Deepwound: Automated Postoperative Wound Assessment and Surgical Site Surveillance through Convolutional Neural Networks. Cornell University. 2018. URL: https://arxiv.org/abs/1807.04355 [accessed 2021-07-05]
13. Boissin C, Fleming J, Wallis L, Hasselberg M, Laflamme L. Can We Trust the Use of Smartphone Cameras in Clinical Practice? Laypeople Assessment of Their Image Quality. Telemed J E Health 2015 Nov;21(11):887-892 [FREE Full text] [doi: 10.1089/tmj.2014.0221] [Medline: 26076033]
14. Zhang J, Tüshaus L, Nuño Martínez N, Moreo M, Verastegui H, Hartinger SM, et al. Data Integrity-Based Methodology and Checklist for Identifying Implementation Risks of Physiological Sensing in Mobile Health Projects: Quantitative and Qualitative Analysis. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11896 [FREE Full text] [doi: 10.2196/11896] [Medline: 30552079]
15. van der Heijden JP, Thijssing L, Witkamp L, Spuls PI, de Keizer NF. Accuracy and reliability of teledermatoscopy with images taken by general practitioners during everyday practice. J Telemed Telecare 2013 Sep;19(6):320-325. [doi: 10.1177/1357633X13503437] [Medline: 24163296]
16. Sperring B, Baker R. Ten top tips for taking high-quality digital images of wounds. Wound Essentials 2014;9(2):62-64 [FREE Full text]
17. Wirz EG, Jaeger VK, Allanore Y, Riemekasten G, Hachulla E, Distler O, EUSTAR coauthors. Incidence and predictors of cutaneous manifestations during the early course of systemic sclerosis: a 10-year longitudinal study from the EUSTAR database. Ann Rheum Dis 2016 Jul;75(7):1285-1292. [doi: 10.1136/annrheumdis-2015-207271] [Medline: 26232495]
18. Mihai C, Landewé R, van der Heijde D, Walker UA, Constantin PI, Gherghe AM, EUSTAR co-authors. Digital ulcers predict a worse disease course in patients with systemic sclerosis. Ann Rheum Dis 2016 Apr;75(4):681-686. [doi: 10.1136/annrheumdis-2014-205897] [Medline: 25688073]
19. Hughes M, Herrick AL. Digital ulcers in systemic sclerosis. Rheumatology (Oxford) 2017 Jan;56(1):14-25. [doi: 10.1093/rheumatology/kew047] [Medline: 27094599]
20. Dinsdale G, Moore TL, Manning JB, Murray AK, Atkinson R, Ousey K, et al. Tracking digital ulcers in systemic sclerosis: a feasibility study assessing lesion area in patient-recorded smartphone photographs. Ann Rheum Dis 2018 Sep;77(9):1382-1384. [doi: 10.1136/annrheumdis-2017-212829] [Medline: 29431120]
21. Yap MH, Chatwin KE, Ng C, Abbott CA, Bowling FL, Rajbhandari S, et al. A New Mobile Application for Standardizing Diabetic Foot Images. J Diabetes Sci Technol 2018 Jan;12(1):169-173 [FREE Full text] [doi: 10.1177/1932296817713761] [Medline: 28637356]
22. Wang L, Pedersen PC, Strong DM, Tulu B, Agu E, Ignotz R. Smartphone-based wound assessment system for patients with diabetes. IEEE Trans Biomed Eng 2015 Feb;62(2):477-488. [doi: 10.1109/TBME.2014.2358632] [Medline: 25248175]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

24. de Villiers JP. A comparison of image sharpness metrics and real-time sharpening methods with GPU implementations. 2010 Presented at: 7th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa; June 21-23, 2010; Franschhoek, South Africa. [doi: 10.1145/1811158.1811168]
25. Dodge S, Karam L. Understanding how image quality affects deep neural networks. 2016 Presented at: 8th International Conference on Quality of Multimedia Experience, QoMEX 2016; June 6-8, 2016; Lisbon, Portugal.
26. Pech-Pacheco JL, Cristöbal G, Chamorro-Martínez J, Fernândez-Valdivia J. Diatom autofocusing in brightfield microscopy: a comparative study. Diatom autofocusing in brightfield microscopy; 2000 Presented at: 15th International Conference on Pattern Recognition. ICPR-2000; September 3-7, 2000; Barcelona, Spain. [doi: 10.1109/icpr.2000.903548]
27. McCamy CS, Marcus H, Davidson JG. Color-Rendition Chart. Journal of Applied Photographic Engineering 1976;2(3):95-99.
28. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum 2013 Nov;65(11):2737-2747 [FREE Full text] [doi: 10.1002/art.38098] [Medline: 24122180]
29. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Polat, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
30. Lewis JR. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. International Journal of Human-Computer Interaction 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
31. Karlen W, Gan H, Chiu M, Dunsmuir D, Zhou G, Dumont GA, et al. Improving the accuracy and efficiency of respiratory rate measurements in children using mobile devices. PLoS One 2014;9(6):e99266 [FREE Full text] [doi: 10.1371/journal.pone.0099266] [Medline: 24919062]
32. Garde A, Zhou G, Raihana S, Dunsmuir D, Karlen W, Dekhordi P, et al. Respiratory rate and pulse oximetry derived information as predictors of hospital admission in young children in Bangladesh: a prospective observational study. BMJ Open 2016 Aug 17;6(8):e011094 [FREE Full text] [doi: 10.1136/bmjopen-2016-011094] [Medline: 27534987]
33. ISO 80601-2-61:2017 Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment. International Standards Organization. 2017. URL: https://www.iso.org/standard/67963.html [accessed 2021-07-05]
34. Kuang B, Pena G, Szpak Z, Edwards S, Battersby R, Cowled P, et al. Assessment of a smartphone-based application for diabetic foot ulcer measurement. Wound Repair Regen 2021 May 03;29(3):460-465. [doi: 10.1111/wrr.12905] [Medline: 33657252]
35. Maddah E, Beigzadeh B. Use of a smartphone thermometer to monitor thermal conductivity changes in diabetic foot ulcers: a pilot study. J Wound Care 2020 Jan 02;29(1):61-66. [doi: 10.12968/jowc.2020.29.1.61] [Medline: 31930943]
36. Orlandi M, Lepri G, Bruni C, Wang Y, Bartoloni A, Zammarchi L, et al. The systemic sclerosis patient in the COVID-19 era: the challenging crossroad between immunosuppression, differential diagnosis and long-term psychological distress. Clin Rheumatol 2020 Jul;39(7):2043-2047 [FREE Full text] [doi: 10.1007/s10067-020-05193-2] [Medline: 32514674]
37. Hughes M, Allanore Y, Chung L, Pauling JD, Denton CP, Matucci-Cerinic M. Raynaud phenomenon and digital ulcers in systemic sclerosis. Nat Rev Rheumatol 2020 Apr;16(4):208-221. [doi: 10.1038/s41584-020-0386-4] [Medline: 32099191]

Abbreviations

API: application prorgramming interface
CCDR: color checker detection ratio
CCS: color checker sharpness
DU: digital ulcer
mHealth: mobile health
SSc: systemic sclerosis

Edited by C Dias; submitted 20.01.21; peer-reviewed by R Halkes, D Pförringer, H Mehdizadeh, S Pillon; comments to author 17.02.21; revised version received 30.04.21; accepted 19.05.21; published 30.07.21.

Please cite as:
Zhang J, Mihai C, Tüshaus L, Scebba G, Distler O, Karlen W
Wound Image Quality From a Mobile Health Tool for Home-Based Chronic Wound Management With Real-Time Quality Feedback: Randomized Feasibility Study
JMIR Mhealth Uhealth 2021;9(7):e26149
URL: https://mhealth.jmir.org/2021/7/e26149
doi:10.2196/26149
PMID:34328440

Original Paper

Supportive Accountability and Mobile App Use in a Tobacco Control Intervention Targeting Low-Income Minority Mothers Who Smoke: Observational Study

Stephen J Lepore[1], BA, MEd, PhD; Bradley N Collins[1], BA, MA, PhD; Howard W Killam[2], BSc, MA; Barbara Barry[2], BM

[1]Department of Social and Behavioral Sciences, Temple University, Philadelphia, PA, United States

[2]User Centered Design Inc, Ashburn, VA, United States

Corresponding Author:
Stephen J Lepore, BA, MEd, PhD
Department of Social and Behavioral Sciences
Temple University
975 Ritter Annex
1301 Cecil B Moore Ave
Philadelphia, PA, 19122
United States
Phone: 1 215 204 9422
Email: slepore@temple.edu

Abstract

Background: Smartphone mobile apps are frequently used in standalone or multimodal smoking cessation interventions. However, factors that impede or improve app usage are poorly understood.

Objective: This study used the supportive accountability model to investigate factors that influence app usage in the context of a trial designed to reduce maternal smoking in low-income and predominantly minority communities.

Methods: We conducted a secondary analysis of data (N=181) from a randomized controlled trial that included a smoking cessation app (QuitPal-m). Supportive accountability was measured by the number of times a participant was advised by their cessation counselor to use QuitPal-m. Participants reported app use helpfulness and barriers. Investigators tracked reported phone and technical problems that impeded app use.

Results: Most participants rated the app as very helpful (103/155, 66.5%), but daily use declined rapidly over time. App use was positively related to the level of perceived app helpfulness (P=.02) and education (P=.002) and inversely related to perceived barriers (P=.003), phone technical problems (P<.001), and cigarettes smoked per day at the end of treatment (P<.001). Participants used the app a greater proportion of the days following app advice than those preceding app advice (0.45 versus 0.34; P<.001). The positive relation between counselor app advice and app usage 24 hours after receiving advice was stronger among smokers with no plan to quit than in those planning to quit (P=.03), independent of education and phone or app problems.

Conclusions: Findings show the utility of supportive accountability for increasing smoking cessation app use in a predominantly low-income, minority population, particularly if quit motivation is low. Results also highlight the importance of addressing personal and phone/technical barriers in addition to adding supportive accountability.

Trial Registration: ClinicalTrials.gov NCT02602288; https://clinicaltrials.gov/ct2/show/NCT02602288

(JMIR Mhealth Uhealth 2021;9(7):e28175) doi:10.2196/28175

KEYWORDS

tobacco cessation; smoking; mHealth; mobile apps; smartphone; mobile phone; adherence; engagement; minority health

Introduction

Mobile apps can offer convenient, low-cost, and on-demand support and intervention for smoking cessation. According to estimates, approximately 63%-76% of smokers own a smartphone [1,2], and hundreds of thousands of smokers download cessation apps monthly [3]. Mobile apps could serve as a standalone intervention or an adjunct to other behavioral

interventions, such as telephone quitlines. Unfortunately, users' low engagement with smoking cessation apps makes it difficult to evaluate their effectiveness [4]. In general, app abandonment is problematic: approximately one-fifth of apps are abandoned after one use [5] and over half within a month [6]. Presumably, greater utilization of cessation apps would increase their effectiveness. One study showed that fully adherent users of smoking cessation apps were more than four times as likely to abstain as nonadherent users, but only 24% were fully adherent [7]. This study aims to identify factors that increase smoking cessation app use to inform theory and future interventions.

The investigation was guided by the supportive accountability model, which maintains that adherence to eHealth interventions, including mobile health apps, can be increased through accountability to a supportive, trustworthy person with relevant expertise, such as a health coach or medical provider [8]. The model aligns with prior health behavior research and theory on treatment adherence. For example, social support from interventionists is positively related to treatment adherence across various medical treatments and health-related behaviors [9]. In one smoking cessation treatment study, social support was associated with higher nicotine patch adherence [10]. Clinical practice guidelines also underscore the importance of intratreatment support in professionally delivered cessation interventions [11]. The quality of support is important. Accountability born out of a drive to please a respected coach or health care provider is likely to be more effective than accountability born out of duress (eg, shame, fear, perceived penalties) [12].

Drawing upon self-determination theory [13], the supportive accountability model predicts that motivation to change a behavior can moderate the effect of supportive accountability on health behavior change [8]. Specifically, the more intrinsically motivated a person is to change a behavior, the less social support (ie, extrinsic motivation) they may require. Intrinsic motivation, which is reflected in behavior change intentions [14], has been linked to behavior change efforts and success. For example, higher intention to quit smoking has been linked positively to smoking abstinence [15], quit attempts, and use of electronic nicotine devices to reduce smoking [16]. Another corollary based on self-determination theory is that supportive accountability will become less necessary as individuals progress from being extrinsically motivated to being internally motivated to reach their goals [8]. Indeed, under these latter conditions, ongoing supportive messaging could be construed as controlling or signal that the support provider doubts the support recipient's ability.

This observational study investigates the relations between supportive accountability, motivation, and smoking cessation app use in the context of a clinical trial aimed to promote smoking cessation among low-income maternal smokers. The trial, Babies Living Safe and Smokefree (BLiSS) [17], targeted mothers who smoke and live in predominantly low-income and minority neighborhoods in a major US city. This population was targeted because children in these communities have an excess burden of environmental tobacco smoke exposure (TSE) [18]. We were especially interested in evaluating the uptake and usage of a mobile smoking cessation app in this population because compared to non-Hispanic White smokers, non-Hispanic Black and Hispanic smokers are less likely to use tobacco-cessation aids during a quit attempt [19]. Identifying correlates of app utilization in this high-risk population could inform future smoking cessation interventions that incorporate mobile apps. Further, this analysis provides a theoretical test of the supportive accountability model in an understudied population.

We tested two hypotheses based on the supportive accountability model:

1. A higher percentage of smokers will use a cessation app on their phone in the 24 hours after receiving prompts about app usage from a cessation counselor (ie, supportive accountability) than in the 24 hours preceding such prompts.
2. The relation between prompting and app usage will be stronger among participants not planning to quit in the next three months than among participants planning to quit in the next three months (ie, motivation as a moderator).

We also explored correlates and potential barriers to app usage. As the target population is low income, we anticipated some potential technical and phone-related barriers (eg, service disruptions due to late payments, phone sharing) as well as practical barriers (eg, no time, lack of interest). Finally, we explored whether app usage correlated with amount of smoking at end of treatment.

Methods

Study Overview

This investigation used secondary data collected as part of the BLiSS trial [17]. BLiSS used a randomized two-group design with three measurement points: baseline, 3-month follow-up, and 12-month follow-up. Outcomes include bioverified child TSE and bioverified maternal quit status. Maternal smokers with children <6 years old were recruited from government-subsidized clinics that deliver the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). All study participants received a WIC system-level intervention based on the Ask, Advise, Refer (AAR) clinical best practice guidelines established by the American Academy of Pediatrics [20]. WIC nutrition counselors delivered AAR. After a WIC referral, the trial's project manager randomized eligible and consented mothers to either a 3-month multimodal behavioral intervention (AAR + MBI) targeting parental smoking, or a 3-month attention control intervention (AAR + Control) targeting family nutrition. The appropriate Institutional Review Board approved all study procedures, and all participants provided informed consent to participate. This observational study is limited to the participants in the AAR + MBI arm of the BLiSS trial and app usage patterns and correlates, not trial outcomes.

Participants

Trial eligibility criteria included the following: received WIC clinic AAR intervention; English-speaking; at least 18 years old; report smoking; own a smartphone; and report their child aged <6 years is exposed to tobacco smoke. Exclusion criteria included the following: currently pregnant; presenting issues

that could interfere with their ability to provide informed consent or follow study procedures, such as psychosis, inadequate health literacy, or non-nicotine drug dependence. All BLiSS participants randomized to the AAR + MBI treatment arm (N=199) were potentially eligible for inclusion in this observation study. However, 18 participants were excluded from all analyses, leaving a sample of 181. Reasons for exclusion included technical problems downloading the mobile app (n=6), issues with the back-end software that tracked participants' app usage (n=6), and participant withdrawal from the trial before receiving intervention or app advice (n=6). Comparisons of excluded and nonexcluded participants revealed no statistically significant differences in age, race, marital status, employment status, education level, or phone operating system.

Procedures

After WIC staff referred mothers to the trial, trained research assistants screened for eligibility, administered informed consent, and collected baseline self-report data using computer-assisted telephone interviews. Participants were then randomized to either the AAR + MBI or AAR + Control condition. The AAR + MBI intervention included messaging about child TSE harms as well as support and guidance with skills training and problem-solving delivered via multiple channels: the project quitline, providing up to 5 telephone counseling sessions over 3 months; cessation mobile app; print materials for the participant and their family; intersession text follow-up, reminders, and support, as well as educational video clips that reinforced telephone session and written materials content; and 8 weeks of nicotine replacement therapy and instructional support.

After randomization, participants in the AAR + MBI condition had an orientation home visit that provided a review of intervention objectives, a binder of intervention print materials, and an illustrated guide to using the mobile app. Research staff also assisted participants in downloading the mobile app and showed them a brief video tutorial on how to use the app.

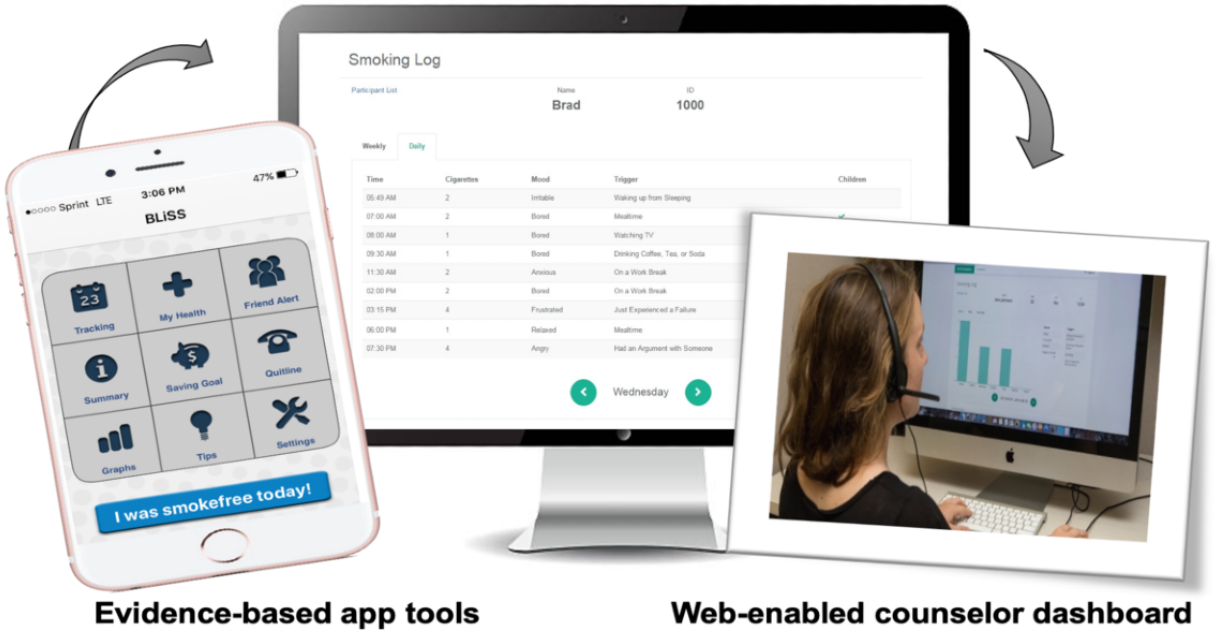
Telephone counselors delivering the skills training and support intervention received intensive training in tobacco treatment enhanced with support, advice, and problem solving around protecting children from tobacco smoke exposure and creating a smoke-free home and car. Importantly, they were trained in how to use a telephone counseling process that would promote participants' mobile app usage to complement and extend treatment beyond the phone sessions. For example, telephone counseling included guidance on goal setting, building social support, and improving skills (eg, self-monitoring) to reduce child TSE, identify smoking triggers, and manage urges to smoke. The mobile app has tools that support all these processes.

The BLiSS mobile app was a modified version of the National Cancer Institute's QuitPal app [21]. The content and tools provided by the QuitPal app are grounded in evidence-based research and US clinical practice guidelines for treating tobacco use and dependence, which makes it stand out among smoking cessation apps [3,22]. Originally designed as a standalone intervention for iOS-based phones, the modified QuitPal app (QuitPal-m) works on both iOS and Android platforms. Key features are shown in Figure 1. QuitPal-m includes features that promote goal setting (eg, quit date, financial goals), real-time self-monitoring of number of cigarettes smoked and cigarettes smoked with children in the same room, and monitoring of mood and context associated with smoking episodes. The app has algorithms that use tracking data and goals to send personalized notifications with tips about smoking triggers and managing cravings, as well as motivational reminders that coincide with progress (eg, health milestones and money saved by reducing or quitting smoking). Other features include goal progress summary and connectivity to the BLiSS quitline and to social media to alert friends of progress and build support for quitting. A video recording tool from the original QuitPal app was excluded from QuitPal-m to facilitate ease of use and to emphasize content and processes covered in the telephone counseling sessions.

An innovative feature of QuitPal-m is a web-linked portal that connects telephone counselors to a dashboard that displays participant app usage (Figure 1). Counselors were trained to review dashboard data before counseling phone sessions to guide their supportive feedback about app usage and the behavior change progress during phone sessions. Counselors aimed to drive participants' early adoption of tracking and responding to app reminders. Counselors also could provide positive reinforcement about tracking efforts and progress, review behavioral patterns emerging over time, and offer to troubleshoot challenges to app usage and behavior change efforts with nonusers. For those participants who readily engaged with the self-monitoring functions of the app, counselors could shift their attention to suggesting how the app could be used to address specific cessation challenges raised during counseling calls. Typically, app advice was provided in each counseling phone session, unless time was limited and other topics took precedence based on a participant’s progress. For example, if a participant missed a phone session, a counselor might have to cover topics from two sessions and not have a chance to address app usage. Counselors' session notes included a field for recording if app usage was discussed as part of the counseling session.

During the home visit orientation to the app, participants were told that the smoking self-monitoring, or tracking, features were the most important and should be used daily. Home visitors demonstrated how to enter data into the app. Participants also were informed about the counselor dashboard and how telephone counselors would routinely monitor app entries to learn more about participants' smoking habits and guide their advice about the participants' behavior change efforts. Finally, they were told that the counselor would remind them or initiate troubleshooting when app use adherence was low. Thus, participants were aware of the expectations, monitoring, and accountability related to app usage from the beginning of the intervention.

Figure 1. QuitPal-m app features and illustration of web-enabled counselor dashboard.



Measures

QuitPal-m App Adherence

Adherence is defined as the active use of the app as recommended [23]. In this study, BLiSS AAR + MBI participants were encouraged to use the app daily. Thus, the primary outcome of interest was the number of days the app was used (range 0 to 91). Each participant's phone was linked to a back-end software program that recorded the days when the app was launched, as well as specific features that were accessed and any inputted data. To count as use, a participant had to launch the app and use one of the eight features (eg, input smoking data, view savings, request a tip).

Adherence to Usage Advice

Another important outcome to test the supportive accountability model was the proportion of days that participants used the app 24 hours before and 24 hours after receiving advice from their counselor to do so. This usage was calculated by dividing the total number of days the app was used 24 hours after (or before) advice was offered divided by the total number of days advice was offered. For example, if a person received advice on three days and used the app within 24 hours on each of those three days, they would score 100 (3/3). If they only used the app within 24 hours on two of the days, they would score 66 (2/3). Thus, scores could range from 0 to 100.

App Helpfulness and Barriers to Use

Participants rated the app helpfulness on a 4-point scale (1=not at all, 2=a little helpful, 3=somewhat helpful, 4=very helpful) postintervention. They also reported (no/yes) whether they experienced any of the following barriers to app usage during the intervention: forgetting, lack of interest, lack of time, and confusion/difficulty using the app. In addition, we tracked phone and technical problems reported throughout the study that interfered with app use (eg, phone not in service, app freezing).

Motivation and Smoking Behavior

To assess participants' motivation/determination to quit smoking, we included a baseline question about whether they planned to quit smoking in the next 3 months (no/yes). We also included a self-report measure of average cigarettes smoked per day in the past week at the 3-month end-of-treatment period.

Results

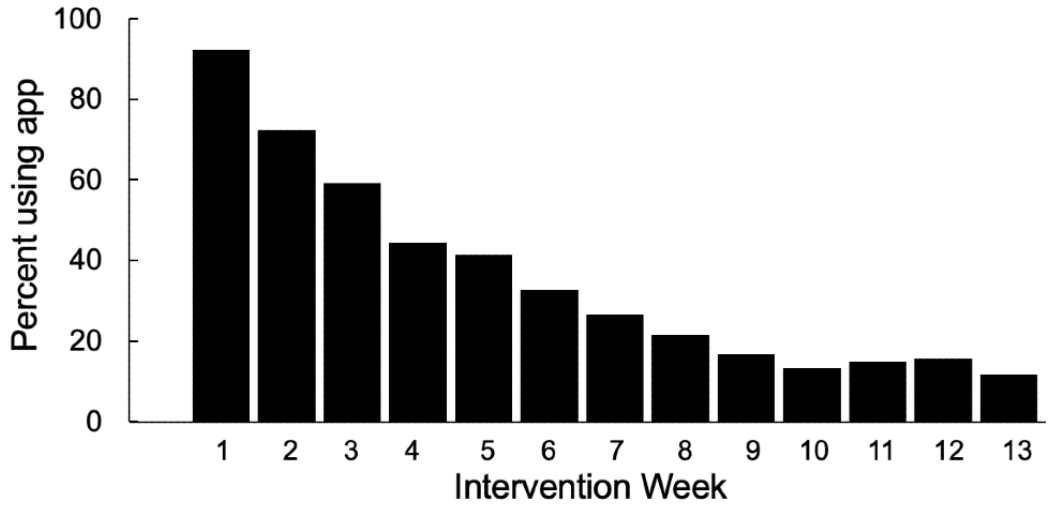
The sample (N=181) was comprised of mothers who were mostly single (113/181, 62.4%), Black (123/181, 68.0%), and unemployed (105/181, 58.0%). The highest level of education completed for the majority was high school or less (111/181, 61.3%). The most common phone operating system was Android (144/181, 79.6%), followed by iOS (37/181, 20.4%). At baseline, over three-fourths (138/181, 76.2%) of the participants reported that they were planning to quit smoking in the next three months.

On average, participants received advice to use the app 3 times (median 3; mean 2.98, SD 1.58) over the course of the intervention. A total of 10 of the 181 participants (5.5%) received no advice: 9 because they could not be reached for phone intervention sessions and 1 because the interventionist did not have an opportunity to bring it up during the single phone session the participant completed. Patterns of app usage are shown in Table 1. The most frequently used feature was the tracking of cigarettes smoked, with all other features used rarely. As shown in Figure 2, app usage was greatest during the first week of treatment and declined rapidly over time. On average, participants used the app on 16 days over the entire intervention period, and fewer than 50% (80/181) used the app after week 4. No participants used the app daily, as recommended, although one person used the app 87/91 days.

Table 1. Patterns of app usage (N=181).

Variable	Mean (SD)	Range
Days app used (out of 91)	16.48 (17.52)	0-87
Times tracking feature used	63.64 (81.68)	0-443
Times savings feature used	1.72 (3.09)	0-26
Times graph feature used	1.36 (2.43)	0-16
Times tips feature used	0.92 (1.85)	0-16
Times summary feature used	1.23 (2.46)	0-15
Times friend alert feature used	0.87 (1.69)	0-14
Times my health feature used	1.48 (2.17)	0-12
Times quitline phone number used	0.39 (0.87)	0-14

Figure 2. Percent of participants using QuitPal-m app by intervention week (N=181).



Examining the relation of days of app usage to sociodemographic factors revealed a single significant association, with education: participants used the app on significantly more days if they had more than a high school education (mean 21.49, SD 21.37) than if their highest education was high school or less (mean 13.32, SD 13.77; t_{179}=3.13, *P*=.002). Frequency of app usage was unrelated to participants' age, number of years smoking, level of dependence on cigarettes, marital/partnered status, and employment status. Consistent with the study's premise that greater adherence to app use can improve outcomes, more days of app use was negatively correlated with number of cigarettes smoked per day at the end of treatment (*r*=–.25, *P*<.001).

Of the 155 participants who completed the postintervention survey, most rated the app as very helpful (103/155, 66.5%) or somewhat helpful (26/155, 16.8%), a minority (23/155, 14.8%) rated it as not at all/a little helpful, and a few did not answer (3/155, 2%). Participants who rated the app as very helpful tended to open the app on significantly more days (mean 20.56, SD 19.44) than their counterparts who rated it less helpful (mean 13.56, SD 13.26; t_{153}=2.34, *P*=.02). A large minority (71/171, 41.5%) reported experiencing barriers to app usage during the intervention. Ordered from most to least common, barriers included the following: forgetfulness (28/171, 16.4%), lack of time (28/171, 16.4%), lack of interest (26/171, 15.2%), and confusing/difficult to use (26/171, 2.3%). Participants who reported any barrier during the intervention opened the app on significantly fewer days (mean 12.51, SD 14.08) than their counterparts who reported no barriers (mean 20.58, SD 19.29; t_{169}=3.00, *P*=.003).

A sizeable proportion (68/181, 37.6%) of the participants reported phone and other technical problems that occasionally interfered with using the app at some time during the intervention. This is in addition to the 12 who were excluded from the study due to app download and back-end issues from the start. The most frequently reported problem was intermittent service disruptions due to running out of minutes or being unable to pay bills on time (33/181, 18.2%), followed by deleting the app due to software problems (eg, app freezing) or insufficient memory for the app (23/181, 12.7%), followed by getting a new phone and having difficulty downloading the app again (12/181, 6.6%). Certain problems, such as phone service disruption, affect some app features and functions, such as communication between the app and the counselor dashboard, but not other features, such as tracking, savings, and tips. Participants who experienced phone/technical problems used the app on significantly fewer days (mean 11.40, SD 13.04) than their

counterparts who did not experience these problems (mean 21.49, SD 19.86; t_{179}=4.04, P<.001).

Finally, we analyzed patterns of usage following advice from the counselor and whether motivation/determination to quit moderated the effects of counselor advice on app usage. These analyses excluded the 10 participants who received no advice. Paired *t* tests showed that, on average, participants used the app a greater proportion of the days following app advice (mean 0.45, SD .37) than days preceding app advice (mean 0.34, SD .35; t_{170}=4.35, P<.001). To investigate if supportive accountability increased app usage, particularly among those who were not planning to quit (low motivation), we used multiple regression. The outcome was proportion of days using the app within 24 hours after receiving counselor advice. The predictors included total number of times advice was given, plan to quit in next three months (yes/no), and the interaction between amount of advice and plan to quit. Variables were centered around zero prior to creating the cross-products for the interaction term. Other covariates included factors known to predict the number of days participants used the app: education, app ever unavailable due to phone/technical problems, and perceived barriers to app usage.

As shown in Table 2, there was a significant main effect of advice, no main effect of planning to quit, and a significant advice × planning to quit interaction on likelihood of app usage within 24 hours after receiving advice. Figure 3 plots the interaction. Simple effects reveal that counselor advice was positively and significantly related to more app usage among those participants who were not planning to quit at the beginning of the study (ie, they were least motivated to change behavior). The simple slope relating advice to app usage was positive but not statistically significant among those who were planning to quit at the beginning of the study.

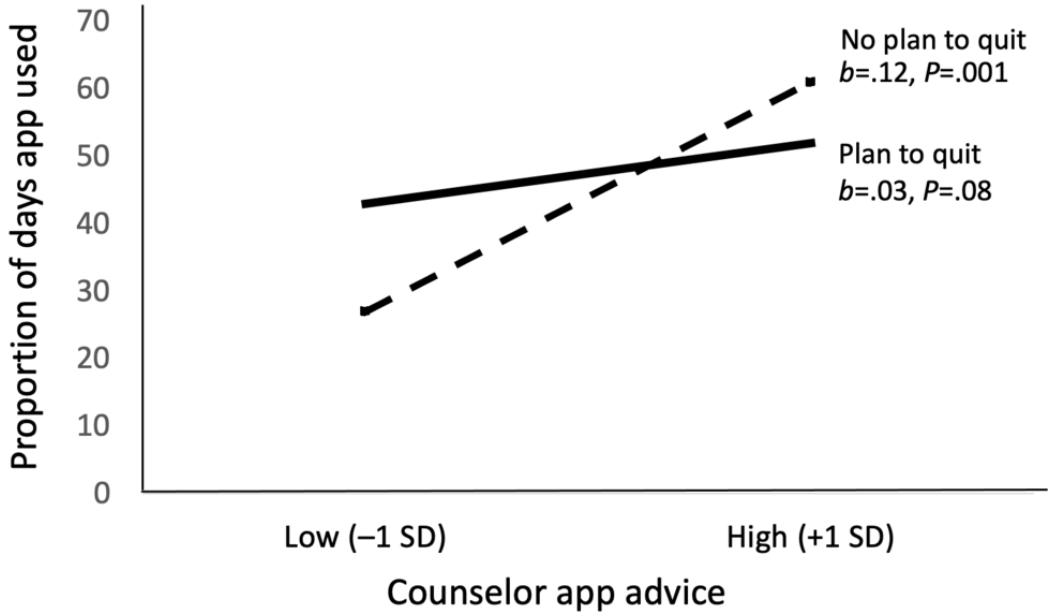
Table 2. Regression models predicting proportion of days app was used within 24 hours after receiving app advice from a counselor (N=171).

Predictor	*B* (SE)	*T* value	*P* value
Constant	.56 (.05)	12.10	<.001
Highest education level[a]	.06 (.05)	1.12	.27
App ever unavailable due to phone/technical problems[b]	–.16 (.06)	–2.82	.005
Reported any barriers to app use[b]	–.16 (.05)	–3.00	.003
Total counselor advice to use app	.05 (.02)	3.28	.001
Plan to quit in next 3 months[b]	.05 (.06)	.79	.43
Total advice × plan to quit	–.08 (.04)	–2.25	.03

[a]0=high school or less, 1=more than high school.

[b]0=no, 1=yes.

Figure 3. Relation between counselor app advice and proportion of days app was used within 24 hours after advice as a function of motivation to quit smoking (N=171).



Discussion

Principal Results

Overall adherence to daily app use goals was low and declined steadily over the course of the intervention. Consistent with the supportive accountability model [8], counselor monitoring and supportive advice about app usage was positively associated with app usage. Also consistent with the model, accountability to a counselor appeared to increase app engagement more among smokers who were not planning to quit than among their counterparts who were already planning to quit. Thus, adding supportive accountability to tobacco cessation interventions that deploy mobile apps can improve adherence, which we hypothesize would enhance intervention effectiveness in research or practice contexts. The observed negative correlation between days of app use and cigarettes smoked per day at the end of treatment is consistent with this expectation. Personal and nuisance factors also were linked to lower app use, including lower educational achievement, perceived barriers (eg, forgetfulness, no time or interest), and phone/technical problems. Altogether, these findings suggest that accountability to a trusted, supportive expert can increase adherence to mobile app treatment elements, but other factors also play a role.

The current findings suggest that supportive accountability is a promising method to improve adherence to mobile app use, especially among users with low levels of motivation to change. The interaction between accountability and motivation suggests that increased motivation is a primary mechanism of supportive accountability in app adherence. Intervention models, such as BLiSS, that incorporate mobile apps and interaction with counselors can build accountability directly into the counseling protocol. Folding in supportive accountability processes and messaging into treatment-goal setting and skills-training elements of behavioral interventions can maximize the chance that participants benefit as much as possible from app engagement. For example, getting participants to engage in self-monitoring/tracking as “homework” and taking stock of their smoking patterns, including triggers and consequences of smoking that sustain the behavior, may boost intervention efficacy. Emphasizing accountability messaging with nonadherent users could have the twin virtues of supporting users with the greatest need and of not alienating users who are already highly engaged.

A final noteworthy finding was the relatively high participant ratings of the helpfulness of the app. Further, the more satisfied participants were with the app, the more they tended to use it. Thus, increasing user satisfaction could help to promote more engagement. These findings appear to conflict with the overall low engagement and rapid drop-off in app use during the intervention. It is possible that participants were able to extract the primary value from the app in less time than we originally predicted. In retrospect, we realize that it would not take more than a week or two of tracking for users and counselors to recognize smoking behavior patterns and, on that basis, make smoking avoidance/cessation strategies. In addition, after several days of using the tracking feature, the app generates automated notifications about health improvements and personalized tips for avoiding smoking without the user having to launch the app daily and manually input data. Apps that use other strategies to address smoking, such as stress management, may be helpful over longer periods of time when relapse prevention is important. However, for initial quit attempts, when it is important to identify and address factors that trigger and sustain smoking, an app with a more circumscribed time of use may be sufficient.

An interesting question for future research is whether the medium and source of supportive accountability influences adherence levels. For example, is an SMS text message from a counselor as effective as direct contact and collaborative problem solving that can occur during a live counseling session? Is a supportive accountability approach as effective at improving app adherence when it is automated or delivered by an embodied conversational agent (eg, a computer-generated avatar of a counselor) versus a live human? Unlike common reminder software, embodied agents can be verbally expressive and mimic human gestures, which could elicit social responses from people that parallel human-human social interaction. If effective, such an approach would reduce some of the burden on the interventionist.

A related question is how much encouragement or nudging from a coach is beneficial for promoting app usage? The supportive accountability model suggests that once an individual has internal motivation to change a behavior, ongoing accounting may backfire, or at least show diminished returns. In this study, the amount of direct advice from the health counselor was deliberately modest and it was front-loaded to the early weeks of the intervention. However, as we did not experimentally manipulate frequency and intensity of supportive accountability-driven app advice and feedback, it raises the question of whether more advice might have resulted in greater adherence.

In thinking broadly about the challenges of app engagement, it is important to consider the social and economic contexts that influence user engagement. In this study, participants were drawn from predominantly minority and low-income communities. As reflected in some of the observed barriers (eg, phone service disruptions, phone sharing, time constraints), participants' life circumstances would have undermined adherence even among those motivated and otherwise engaged in the overarching multimodal intervention. Linking the QuitPal-m app to wearable smoking sensors [24] might overcome some contextual factors (eg, time constraints) that disrupt app usage and reduce the burden of tracking smoking. The introduction and eventual widespread availability of 5G will potentially overcome other barriers, such as app connectivity problems. Another noted barrier was lack of interest, which might be improved by amplifying some of the game-like elements of QuitPal-m, such as social connectivity, financial savings graphs, and praise for achieving goals [25].

Limitations

This study has some noteworthy limitations. The primary one is the lack of experimental data. Instead of manipulating levels of supportive accountability, we measured it and observed how it related to app use behaviors. This correlational design is

subject to internal validity threats, including the possibility that something other than accountability increased participants' app use after receiving counselor advice. For example, maybe the call itself served as a reminder and did not require the counselor to specifically discuss the app. Future research will need to add randomization and control conditions to rule out alternative explanations of findings. Another limitation is that participants used their own cell phones and service plans. This introduced extraneous factors that interfered with app use independent of users' intentions/desires to use the app. A controlled study would be able to isolate the effects of supportive accountability by providing a device and service, ensuring equitable access for all participants.

Conclusions

The findings show the potential utility of supportive accountability for increasing use of a smoking cessation app in a low-income, predominantly minority population. Consistent with the model advanced by Mohr and colleagues [8], supportive accountability-driven app advice was most helpful for smokers with low motivation to change their smoking behavior. This finding suggests that it might be possible to target messaging based on individuals' stage of change, or progress in treatment (eg, preparing to quit, initial quit phase, or efforts to maintain longer-term abstinence). Finally, we found that participants' social and economic life contexts influenced app use. Addressing these factors, including time constraints, interest level, and access to affordable high-quality phones/devices and service will also help improve app use.

Acknowledgments

The authors thank the research assistants and staff and administrators from North Inc who assisted with the study. This project was supported by a grant from the National Institutes of Health/National Cancer Institute (R01CA188813).

Conflicts of Interest

None declared.

References

1. 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]
2. Heffner JL, Mull KE. Smartphone Ownership Among US Adult Cigarette Smokers: 2014 Health Information National Trends Survey (HINTS) Data. J Med Internet Res 2017 Dec 31;19(8):e305. [doi: 10.2196/jmir.7953] [Medline: 28860108]
3. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
4. 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;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
5. Rodde T. 21% of users abandon an app after one use. Localytics Blog. 2018 Feb 28. URL: https://info.localytics.com/blog/21-percent-of-users-abandon-apps-after-one-use [accessed 2020-02-15]
6. Perro J. Mobile apps: what's a good retention rate. Technopreneurph. 2016 Mar 29. URL: https://technopreneurph.wordpress.com/2016/03/29/mobile-apps-whats-a-good-retention-rate-by-justina-perro/ [accessed 2020-02-15]
7. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: Adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124. [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
8. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
9. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
10. de Dios MA, Stanton CA, Cano, Lloyd-Richardson E, Niaura R. The Influence of Social Support on Smoking Cessation Treatment Adherence Among HIV+ Smokers. Nicotine Tob Res 2016 May;18(5):1126-1133 [FREE Full text] [doi: 10.1093/ntr/ntv144] [Medline: 26116086]
11. Fiore M, Jaen C, Baker T, Bailey W, Benowitz N, Curry S. Clinical Guidelines and Recommendations | Treating Tobacco Use and Dependence. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/prevention/guidelines/index.html [accessed 2020-02-15]
12. Oussedik E, Foy CG, Masicampo EJ, Kammrath LK, Anderson RE, Feldman SR. Accountability: a missing construct in models of adherence behavior and in clinical practice. Patient Prefer Adherence 2017;11:1285-1294 [FREE Full text] [doi: 10.2147/PPA.S135895] [Medline: 28794618]
13. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]

14. Ajzen I, Kruglanski AW. Reasoned action in the service of goal pursuit. Psychol Rev 2019 Oct;126(5):774-786. [doi: 10.1037/rev0000155] [Medline: 31343213]
15. Pai C, Edington DW. Association between behavioral intention and actual change for physical activity, smoking, and body weight among an employed population. J Occup Environ Med 2008 Sep;50(9):1077-1083. [doi: 10.1097/JOM.0b013e31817d0795] [Medline: 18784556]
16. Nayak P, Pechacek TF, Weaver SR, Eriksen MP. Electronic nicotine delivery system dual use and intention to quit smoking: Will the socioeconomic gap in smoking get greater? Addict Behav 2016 Oct;61:112-116 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.020] [Medline: 27268063]
17. Collins BN, Lepore SJ. BMC Public Health 2017 Mar 14;17(1):249 [FREE Full text] [doi: 10.1186/s12889-017-4145-7] [Medline: 28288601]
18. Homa DM, Neff LJ, King BA, Caraballo RS, Bunnell RE, Babb SD, Centers for Disease Control and Prevention (CDC). Vital signs: disparities in nonsmokers' exposure to secondhand smoke--United States, 1999-2012. MMWR Morb Mortal Wkly Rep 2015 Feb 06;64(4):103-108 [FREE Full text] [Medline: 25654612]
19. Cokkinides VE, Halpern MT, Barbeau EM, Ward E, Thun MJ. Racial and ethnic disparities in smoking-cessation interventions: analysis of the 2005 National Health Interview Survey. Am J Prev Med 2008 May;34(5):404-412. [doi: 10.1016/j.amepre.2008.02.003] [Medline: 18407007]
20. Counseling about smoking cessation. American Academy of Pediatrics. URL: https://www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/Richmond-Center/Pages/Counseling-About-Smoking-Cessation.aspx [accessed 2020-02-15]
21. Massett H, Atkinson NL, Kraiger A, Grady M, Killam B, Barry B. NCI Quitpal: an evidence-based app to help people quit smoking. In: American Public Health Association (APHA) 141st Annual Meeting. 2013 Presented at: American Public Health Association (APHA) 141st Annual Meeting; November 6, 2013; Boston, MA.
22. Regmi D, Tobutt C, Shaban S. Quality and use of free smoking cessation apps for smartphones. Int J Technol Assess Health Care 2018 Jan;34(5):476-480. [doi: 10.1017/S0266462318000521] [Medline: 30226123]
23. Ryan C, Bergin M, Wells JS. Theoretical Perspectives of Adherence to Web-Based Interventions: a Scoping Review. Int J Behav Med 2017 Jul 20:17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
24. Imtiaz MH, Ramos-Garcia RI, Wattal S, Tiffany S, Sazonov E. Wearable Sensors for Monitoring of Cigarette Smoking in Free-Living: A Systematic Review. Sensors (Basel) 2019 Oct 28;19(21):4678 [FREE Full text] [doi: 10.3390/s19214678] [Medline: 31661856]
25. Rajani NB, Weth D, Mastellos N, Filippidis FT. Use of gamification strategies and tactics in mobile applications for smoking cessation: a review of the UK mobile app market. BMJ Open 2019 Jun 17;9(6):e027883 [FREE Full text] [doi: 10.1136/bmjopen-2018-027883] [Medline: 31213452]

Abbreviations

AAR: Ask, Advise, Refer
BLiSS: Babies Living Safe and Smokefree
MBI: multimodal behavioral intervention
QuitPal-m: modified QuitPal app
TSE: tobacco smoke exposure
WIC: Special Supplemental Nutrition Program for Women, Infants, and Children

Edited by G Eysenbach; submitted 24.02.21; peer-reviewed by B Bock, H Mehdizadeh; comments to author 17.03.21; revised version received 24.04.21; accepted 31.05.21; published 02.07.21.

Please cite as:
Lepore SJ, Collins BN, Killam HW, Barry B
Supportive Accountability and Mobile App Use in a Tobacco Control Intervention Targeting Low-Income Minority Mothers Who Smoke: Observational Study
JMIR Mhealth Uhealth 2021;9(7):e28175
URL: https://mhealth.jmir.org/2021/7/e28175
doi:10.2196/28175
PMID:34255698

Original Paper

User Perspectives on a Resilience-Building App (JoyPop): Qualitative Study

Aislin R Mushquash[1], PhD; Erin S Pearson[2], PhD; Kayla Waddington[2], HBK; Angela MacIsaac[1], MA; Shakira Mohammed[1], MA; Elizabeth Grassia[1], HBA; Savanah Smith[3], HBA; Christine Wekerle[3], PhD

[1]Department of Psychology, Lakehead University, Thunder Bay, ON, Canada

[2]School of Kinesiology, Lakehead University, Thunder Bay, ON, Canada

[3]Department of Pediatrics, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Aislin R Mushquash, PhD
Department of Psychology
Lakehead University
955 Oliver Road
Thunder Bay, ON,
Canada
Phone: 1 807 343 8771
Email: aislin.mushquash@lakeheadu.ca

Abstract

Background: Resilience is the capability, resources, and processes that are available to a person or system to adapt successfully in the face of stress or adversity. Given that resilience can be enhanced, using advances in technology to deliver and evaluate the impact of resilience interventions is warranted. Evidence supports the effectiveness of the resilience-building JoyPop app in improving resilience-related outcomes after use; however, experiential data from users is also needed to provide a more comprehensive account of its utility.

Objective: The aim of this study was to explore users' experiences with the JoyPop app and their perspectives on its utility.

Methods: This qualitative description study involved a combination of group and one-on-one semistructured interviews with a subset of first-year undergraduate students who participated in a larger evaluation of the JoyPop app. Participants used the app for a 4-week period and were subsequently asked about their frequency of app use, most and least used features (and associated reasons), most and least helpful features (and associated reasons), barriers to use, facilitators of use and continuation, and recommendations for improvement. Data were coded and categorized through inductive content analysis.

Results: The sample of 30 participants included 24 females and 6 males, with a mean age of 18.77 years (SD 2.30). App use ranged from 1 to 5 times daily (mean 2.11, SD 0.74), with the majority indicating that they used the app at least twice daily. The Rate My Mood, Journal, and SquareMoves features were reported to be used most often, while the Rate My Mood, Journal, and Breathing Exercises features were identified as the most helpful. A number of themes and subthemes pertaining to facilitators of app use (prompts, creating routine, self-monitoring opportunities, expressive opportunities), barriers to app use (editing, lack of variety, student lifestyle), outcomes of app use (increased awareness, checking in with oneself, helpful distraction, emotional control), and recommendations for app improvement (adding more features, enhancing existing features, enhancing tracking abilities, providing personalization) were identified.

Conclusions: This study provides insight into the aspects of the JoyPop app that motivated and benefitted users, as well as measures that can be taken to improve user experiences and promote longer-term uptake. Users were willing to engage with the app and incorporate it into their routine, and they valued the ability to self-monitor, express emotion, and engage in distraction.

(JMIR Mhealth Uhealth 2021;9(7):e28677) doi:10.2196/28677

KEYWORDS

resilience; smartphone; app; innovation; qualitative; perspective; mHealth; emotion; mental health

Introduction

Background

Resilience is the multifaceted capability of an organism or system to adapt to challenging circumstances [1,2]. As adaptive creatures, humans must continually cope with stressors as they arise and find ways to flourish despite setbacks. A large part of this ability to cope and adapt involves the use of foundational skills that underlie resilience, including self-regulation processes such as emotion regulation [3-7].

Because resilience is developed and shaped over time, the argument follows that resilience can be promoted [8]. It is beneficial to assist individuals during moderately stressful experiences by providing them with self-regulatory skills (eg, emotion regulation) when they need them so that these skills might be added to their coping repertoire over time and become habitual resilient responses in times of stress [8]. The objective of this study was to explore user experiences with a resilience-building intervention (the JoyPop app [9,10]) and their perspectives on its utility.

Building Resilience

Mobile health (mHealth) app use has greatly expanded in recent years, with the benefit of reaching a wider population than traditional health interventions [11] and providing greater ease of access [12,13]. Recent evidence has demonstrated that mHealth interventions, including apps, are feasible and effective methods of increasing symptom reporting and improving medication and treatment adherence and health knowledge in those with chronic health conditions [14-16]. Evidence to date has also supported an mHealth approach for many different mental health–related targets, such as mood disorders, stress, and substance use [17,18], with emerging support for mHealth approaches to resilience [9]. Focusing on building resilience through mHealth approaches is in line with the recent recommendations that mental health-related apps be geared toward general well-being and fostering self-regulatory skills (eg, emotion regulation) as opposed to focused on addressing specific mental health disorders [19].

Evaluating mHealth Apps

Although there is promise in promoting resilience with apps, evaluating an app's utility requires a multimethod approach that includes a quantitative examination of changes seen in target outcomes following app use, along with qualitative data exploring users' perspectives on the app and its utility. Although establishing the effectiveness of an app via quantitative data is important and often lacking [20], doing so does not ensure that an app will be used or accepted by the target audience [21]. Interestingly, the alignment of an app with evidence-based interventions is often unrelated to its popularity [22], and some of the more popular apps contain little evidence-based content [23]. Obtaining user perspectives can provide insight into patterns of use and continuation over time [18,24-26].

User acceptance and uptake of an app have often been measured through proxy metrics, such as utilization data and download rates; however, qualitative data can better reveal the features and functions that draw users in and keep them engaged [27,28]. This is made especially clear by the evident eagerness of app users to communicate their desires and preferences. For example, app reviews posted to popular web-based stores (eg, Google Play Store, Apple App Store) often contain requests for certain features and design changes [25,27]. Obtaining user input throughout the design and evolution process helps to ensure that the needs of the target audience are met and would likely increase the possibility of uptake [25,27]. Gleaning perspectives from users is also in line with a proposed Canadian assessment framework for e–mental health apps, which indicates that apps should be user-centered and user-desirable such that they reflect the needs and expectations of potential users [29].

Evaluation of the JoyPop App

The resilience-building JoyPop app (see Figure 1) targets self-regulation through features that promote self-reflection and self-awareness [10]. As documented by MacIsaac et al [9], the JoyPop app was developed to focus on daily self-regulation through evidence-based techniques. Rate My Mood is a self-monitoring feature meant to increase insight and regulation of emotions [30,31]. The Breathing Exercises are intended to decrease arousal in support of increased self-regulation [32]. The Journal feature encourages self-reflection and includes a positive focus that is supported by research [33,34]. SquareMoves is a Tetris-like game that is included for its ability to induce a "flow"-like state in which the user is completely involved in the activity, thereby providing a positive distraction and decreased negative self-focus [35]. The Art feature allows users to express their creativity through doodling, which can be used as a nonverbal emotional outlet [36]. Social connection to one's support network is provided with the Circle of Trust feature, which allows users to input contact information and quickly reach out to their supports when help or connection is needed. Direct access to established helplines is also provided through the app; this function has been identified as important to mHealth app users [19,25,27,37].

Figure 1. Features of the JoyPop app. LGBTQ: lesbian, gay, bisexual, transgender, queer.

The JoyPop app was designed with youth, who have demonstrated a willingness to use mHealth apps [28,38], as a target audience. Research conducted by a federally funded Canadian team examining pathways to resilience for youth who had experienced adversity informed early decisions on feature inclusion [39-43]. In addition, consultation with youth, service providers, and clinician-scientists informed initial discussions with app developers at Clearbridge Mobile (Wekerle and Smith 2021, unpublished data). Once an initial version of the app was developed, youth involved with child welfare and victim services as well as providers who worked with youth provided input that was used to revise the features and design. Additional information on JoyPop is available on the web [10].

Research to date suggests that the JoyPop app contributes to positive changes in outcomes related to resilience among youth. In a sample of undergraduate students making the transition to university who were asked to use the app for 4 weeks, improvements in emotion regulation and depression symptoms were evidenced with each additional day a participant used the app [9]. Further, improvements in emotion regulation were especially evident for individuals who had experienced a greater

degree of adversity during childhood. While the transition to university is associated with new and exciting opportunities, it is also marked by stress and an increased prevalence of mental health difficulties [44,45], making it an important time to support resilience and adaptive coping skills.

Gathering qualitative evidence on users' experiences with the app and perspectives of its utility is important for a comprehensive evaluation. For instance, because the JoyPop app contains numerous features that might differentially contribute to changes in outcomes, a qualitative assessment could elucidate which features and functions participants felt were most used and most helpful. Moreover, with the goal of promoting autonomous use of the app outside of a research context, a qualitative assessment would help gather feedback on factors facilitating continued long-term use.

Objective

The objective of this study was to describe users' experiences with the JoyPop app and their perspectives on its utility. A priori hypotheses were not specified due to the nature of the qualitative research design.

Methods

Design

This study was part of the larger program of research examining the outcomes associated with using the JoyPop app [9]. Qualitative description, a naturalistic inquiry approach, was used as an overarching framework, as this method is most appropriate when a straightforward description of a phenomenon is sought [46,47]. Ethical approval was received from Lakehead University's Research Ethics Board prior to commencing the study.

Participants

Undergraduate students who were enrolled in their first year at a mid-size Canadian university (ie, Lakehead University), were fluent in English, and had an iPhone were eligible for the larger study. As documented elsewhere [9], participants attended 3 group laboratory sessions: pre-app, mid-app (after 2 weeks of app use), and post-app (after 4 weeks of app use). During these sessions, participants received information about the JoyPop app, were provided with support in accessing the app, and completed self-report measures assessing quantitative outcomes. Participants were asked to use the JoyPop app at least twice per day over the 4-week study period; no additional requirements were made with respect to feature use or time spent using the app. Each morning and evening (ie, twice per day), participants were sent reminder emails to use the app.

Prior to their post-app session, participants were presented with the option of sharing their experiences using the app and perspectives on its utility by participating in this qualitative description study. Those who expressed interest were invited to participate and received CAD $10 (US $8.13) upon completion.

Data Collection

Data were collected from February to March 2019 through a combination of group and one-on-one semistructured interviews immediately following participants' completion of their self-report measures during the post-app laboratory session. Because the nature of each laboratory session was group based, this same format was selected for the interviews in an effort to maximize involvement while minimizing scheduling challenges and participant burden. This structure also enabled questions to be asked systematically in a direct manner by the interviewer to each participant on an individual level (in contrast with a focus group, where participants interact more with one another) [48-50]. One-on-one interviews were conducted when only 1 participant among those attending the post-app session expressed interest in participating or when a participant wished to participate but had a scheduling conflict with the group-based session. All interviews were conducted in a private board room on the Lakehead University campus.

The interview guide (see Multimedia Appendix 1), created for the purposes of the study, asked participants about their frequency of app use, most and least used features (and associated reasons), most and least helpful features (and associated reasons), barriers to use, facilitators of initial and continued use, and recommendations for improvement. The same interview guide was used during all group and one-on-one interviews. Interviewers (AM and SM) were enrolled in a master's degree program in clinical psychology at Lakehead University with training and supervision related to interviewing and qualitative data collection. Interviewers were supervised by ARM, who is a Registered Clinical Psychologist with extensive training in interviewing and experience conducting qualitative research. Interviewers expanded beyond the interview guide (ie, probes and prompts) throughout each interview to promote elaboration on given responses. To ensure that input was received from all participants, interviewers sought to encourage responses through direct questioning, provision of elongated pauses, and reflective listening [46]. Each interview was audio recorded and transcribed verbatim.

Data Analysis

All transcripts were deidentified and analyzed by ESP and KW independently. The data analysts first read and reread the transcripts to become familiar with the data. Inductive content analysis, whereby themes are not predetermined but emerge organically, was then used to code and categorize the text [51]. This involved the application of open coding (ie, writing detailed notes, headings, keywords, and commonly used phrases while reading). A list of reoccurring categories was then created and grouped into higher-order themes [46,51]. Following independent analysis, the data analysts met on multiple occasions and used an iterative, constant comparative method to discuss their findings and agree upon the organizing thematic framework. When differing interpretations occurred, the researchers continued discussions until consensus was reached [52]. Following this manual analysis phase, the transcripts were imported into NVivo Software 12 (QSR International) to verify the consistency and hierarchy of the resultant themes. Finally, the thematic framework was reviewed and discussed by all

members of the researcher team to ensure consensus and that interpretations aligned with participants' meanings and experiences [46,53]. Frequencies were applied to provide an account of overall app and specific feature use.

To enhance the trustworthiness of the data and limit researcher biases [46,54,55], several strategies were integrated into the data collection and analysis phases. For example, to promote credibility, member checking was done between questions to ensure the interviewers understood participant responses correctly [53]. The researchers also kept detailed notes throughout the duration of the study in the form of an audit trail to enable a transparent account of the data analysis process [46]. In addition, to enhance the dependability of the data and endorse neutrality, the analysis was conducted by two independent researchers (ESP, KW) who were trained in qualitative methodologies but removed from the conceptualization and data collection phases of the study.

Results

Participant Demographics

A sample of 30 first-year undergraduate students was enrolled in the study. Participants were predominantly female (24/30, 80%) and 18.77 years old on average (SD 2.30). Their ages ranged from 16-29 years, with the majority of the sample being aged 19 years or younger (28/30, 93%). Most participants self-identified as White (21/30, 70%), Asian (6/30, 20%), or Black (2/30, 7%). Close to half (14/30, 47%) of the participants were pursuing degrees within the area of health and behavioral sciences (eg, kinesiology, psychology, nursing), while 5/30 (17%) were undeclared/undecided. The remaining participants were pursuing degrees in science and environmental studies (7/30, 23%; eg, biology, geography), business administration (2/30, 7%; eg, business), or social sciences and humanities (2/30, 7%; eg, social work, political science). We conducted 8 group-based (3 with n=2; 4 with n=3; 1 with n=5) and 7 one-on-one interviews. On average, the group interviews were 16.19 minutes long, and the one-on-one interviews were 8.79 minutes long.

Overall Use

At the outset of each interview, participants were asked how often they used the JoyPop app over the preceding 4 weeks. Use ranged from 1 to 5 times daily (mean 2.11, SD 0.74), with the majority indicating that they used the app at least twice daily. Some participants specifically reported that they used it first thing in the morning and again in the evening before going to sleep.

Feature Use and Perceived Helpfulness

A summary of the use of specific features and their perceived helpfulness is presented in Table 1. Most participants indicated that the Rate My Mood, Journal, or SquareMoves features were used most often. Most participants identified either the Rate My Mood, Journal, or Breathing Exercises as most helpful. The Art feature was rated as the least used and reported most often as being the least helpful.

Table 1. Feature use and helpfulness (N=30).

Feature	Value, n (%)[a]			
	Most used	Least used	Most helpful	Least helpful
Rate My Mood	16 (53)	0 (0)	8 (27)	3 (10)
Journal	15 (50)	8 (27)	8 (27)	5 (17)
SquareMoves	16 (53)	5 (17)	3 (10)	3 (10)
Breathing Exercises	9 (30)	5 (17)	10 (33)	1 (3)
Art	2 (7)	13 (43)	0 (0)	8 (27)
Helplines	0 (0)	1 (3)	0 (0)	2 (7)
Circle of Trust	0 (0)	12 (40)	0 (0)	4 (13)

[a]Some participants identified >1 feature when asked about the most/least used and most/least helpful features.

User Perspectives

In addition to summarizing the reported use and helpfulness, analyses of the transcripts revealed a number of themes and subthemes pertaining to facilitators of app use, barriers to app use, outcomes of app use, and recommendations for app improvement. Each theme and its associated subthemes are described below along with illustrative quotes.

Facilitators of App Use

Participants described several positive attributes of the JoyPop app that facilitated their enjoyment and/or continued use. The most commonly discussed attributes included prompts; creating routine; self-monitoring opportunities; and expressive opportunities.

Prompts

Many participants noted their appreciation for the two types of prompts associated with the study design or app itself. Specifically, receiving email reminders to use the app and the prompts built into the Journal feature (ie, meant to provide suggestions for topics to write about) were identified as helpful.

> *I like to write and like, get my expressions out. But sometimes I like, struggle to start. So, having the prompts was nice...* [Participant 4]

> *At first, I felt like I didn't really need the app…when I got the reminder and I kept on using it, I felt like it was actually really helpful.* [Participant 27]

Creating Routine

Participants discussed how over time, they developed a routine with using the JoyPop app. Many shared how they used it in the morning to start their day and in the evening to "decompress" and process the day's events. Several participants also noted that the more they used the app, the easier and more enjoyable it became.

> *At first, it was just like "Oh, I have to use the app." … And after a while, I just found myself just kind of using it on my own, not really thinking about having to do it, but wanting to do it.* [Participant 8]

> *Yeah it definitely got better with time. And like, incorporating the breathing exercises into like, your everyday routine. I found it really helpful.* [Participant 10]

Self-Monitoring Opportunities

Many participants commented about appreciating the ability to look back at their calendar and journal entries over time, which enabled them to review and reflect on previous feelings and circumstances. This act of self-monitoring was deemed to be helpful in gaining perspective on one's feelings and current state and also facilitated continued use of the JoyPop app. Some participants also reported using various app features in tandem, such as journaling following their mood rating, which likely contributed to ongoing self-monitoring.

> *Look[ing] back, even if I was like, having a bad day I'd be like, "Oh, I was really happy on this day." And if like, I'm feeling bad one day I could look back and see like, "Oh, this week I was really, really happy" kind of thing.* [Participant 5]

> *I liked being able to rate my mood and then go into the journal setting and be able to like, write down things that were happening that day, so I could go and look back.* [Participant 13]

Expressive Opportunities

Some participants shared how using the JoyPop app enabled them to express themselves and that this was a positive feature that encouraged future use. Some noted prior difficulty communicating their feelings verbally and stated that through the app, they were able to express their emotions more readily via the Journal or Rate My Mood features.

> *It helps give students like, the opportunity to have a way to express themselves... the journal [feature] for example. But also, like keeping track of how they're feeling and keeping themselves in check by having like, that visual… showing them where they're at with their emotions.* [Participant 4]

> *[I found that journaling] was really helpful at the end of the day, just to kind of unload everything…[it] unpacked my day so I found that was really helpful.* [Participant 18]

Barriers to App Use

In addition to factors that facilitated enjoyment and ongoing use of the JoyPop app, participants noted some factors that they felt interfered with their use. The primary barriers were related to editing, lack of variety, and student lifestyle.

Editing

Some participants discussed the inability to make changes to their data in some of the features. For example, a few expressed that they did not feel comfortable writing in the Journal if they were not able to later edit or retract their entry. Similar sentiments were expressed by some participants regarding not being able to save their drawings in the Art feature. Participants expressed that the option to edit or save their data would have motivated them to use certain features more often.

> *I didn't like how you couldn't edit [the journal entries] because, [I'm] someone who can get anxious about making mistakes…* [Participant 1]

> *It didn't save my mood and stuff, you know?… I also didn't like that I couldn't save my drawings. Like, or attach a drawing to a mood or something like that.* [Participant 17]

Lack of Variety

Some participants commented on the limited prompt variety in the Journal feature, indicating that once they had responded to all the available prompts, they felt less inclined to continue to use the feature. Some participants similarly noted that because the Rate My Mood feature included a limited number of emotions, they felt that their true feelings may not have been captured.

> *I think the quotes being the same every time when you fill out the [journal] got kind of [frustrating].* [Participant 1]

> *I liked the rate my mood the least… it was just like, a happy or sad thing. Would've been nice be able to have something that it's like, there's different moods to rate. So like maybe I was feeling very happy but there would have been one to track, like how nervous I was feeling or how distressed like I feel. Like, more options would have been nice.* [Participant 3]

Student Lifestyle

Some participants commented that, at times, the busy lifestyle and fatigue associated with being a student interfered with using the app. These participants indicated that "life" affected their use of the JoyPop app, and they mentioned feeling that there was not enough time to use it due to competing priorities and responsibilities.

> *[If] I was like, really busy studying, like during midterms [I would use it less].* [Participant 3]

> *If I was busy at the moment…if I was working on schoolwork or something…it wouldn't happen.* [Participant 28]

Outcomes Related to App Use

During the interviews, participants expressed several positive outcomes related to their use of the JoyPop app. These were categorized into four main subthemes: increased awareness; checking in with oneself; helpful distraction; and emotional control. While most participants noted positive outcomes, a few mentioned experiencing increased stress while using the app.

Increased Awareness

Many participants expressed how the JoyPop app enhanced their self-awareness in terms of being better able to recognize and differentiate their emotions. In addition, a few participants shared that after using the app, they began thinking more about the "why" behind their emotions, and how their feelings were related to mental and physical health.

> *It kinda gave me some time to reflect back on…certain things…After a point, I started noticing a pattern of why this one emotion I kept linking with a certain event, and then it kinda made me internalize like, "Am I depending too much on this event to make me feel a certain way?"* [Participant 6]

> *I was really impressed with the fact that it brought my attention more to how I was feeling. So, instead of just acknowledging that I was stressed and then managing with that, it really brought attention to why I was feeling that and just acknowledging that, which I thought was really cool.* [Participant 18]

Checking in With Oneself

The notion of taking or making time to check in with oneself was discussed frequently. In particular, participants mentioned benefits associated with pausing and reflecting on how they were feeling and what they might need in the moment (eg, to focus on their breathing). Related to their increased awareness, participants also expressed that because they were more aware of how they were feeling, they were able to "reset" and get things "back in check."

> *It's just nice to start your day off or end your day off with…taking a moment for yourself and like…getting your breathing in check and your emotions back in check.* [Participant 4]

> *[It] kind of like, forced me to take time throughout the day to reflect upon how I was feeling.* [Participant 8]

Helpful Distraction

Many participants expressed that some features of the JoyPop app helped them to "take their mind off of" negative feelings or stressors they were experiencing. The distraction provided by the Journal and SquareMoves were deemed particularly useful and enabled the participants to focus on "something else."

> *I liked the journal one the most I think, 'cause it was like, different. It was kind of like, distracting to take your mind off of whatever you might have been feeling at the time.* [Participant 5]

> *There were times when I'd be like on my phone and like, bored and stressed out, and just looking for a distraction and then I was like, "Ok, this was something I could do right now."* [Participant 11]

> *I think the [SquareMoves feature] was probably the most helpful in terms of like, turning off my thoughts and focusing on something else.* [Participant 22]

Emotional Control

Related to their increased awareness, many participants shared how the JoyPop app helped them to have more active control over their emotions. Using specific features (eg, Journal, Breathing Exercises) was discussed often as a means of managing emotions directly, especially in relation to school-specific situations (eg, midterm examinations, studying).

> *Take a couple more minutes and like, go through the events of my day… I kinda found that…as a good way to kinda, balance my emotions.* [Participant 6]

> *It really helps to calm down: the breathing exercises. It really helps like when I'm…angry… it helped a lot.* [Participant 16]

> *It definitely helped me figure [my feelings] out… actually helped me to relate my mood… Often I would feel better afterwards because I'd be like, "Ok. I know how to deal with this now"… rather than just sitting there and stewing not knowing what to do.* [Participant 28]

Increased Stress

A few participants believed that the app or specific features of the app contributed to increased stress. For example, one participant explained that they felt there were too many buttons popping up which distracted them while using the app.

> *Too many buttons to hit or same quotes popping up or it was just…Even though it wasn't the app causing the stress, I found that in this situation it couldn't get me out of it… [It] was too many buttons and distracting colors to help.* [Participant 1]

> *[Using the breathing feature] …I feel stressed and stuff when I use this.* [Participant 18]

Recommendations for App Improvements

Suggestions for improving the JoyPop app involved four main subthemes: adding more features; enhancing existing features; enhancing tracking abilities; and providing personalization.

Adding More Features

Many participants suggested adding more features for future app users. Specifically, participants recommended adding more activities (eg, games and varied breathing exercises or strategies for relaxation). The notion of incorporating accountability was also discussed by a few participants, with suggestions to increase social connectedness through the app and adding a reward system tied to use patterns. Finally, a few participants felt that reminders sent through their phone (eg, via pop-up notifications) would lead to increased app use.

> *Some apps use like, this notification. Like, they notify you after not using [it] for an hour. So, maybe adding [more notifications on the app].* [Participant 26]

> *I think having more activities on it…or being able to connect with people through the app rather than having to call them…getting their opinion on coping strategies, or… maybe having them to… check in on you every once in a while… like have accountability.* [Participant 28]

Enhancing Existing Features

Some participants noted that more mood options would be beneficial in the Rate My Mood feature. Because only four emotions were included, the participants felt that their true feelings may not have been captured. Another common suggestion by participants was to include more prompts in the Journal feature. One participant suggested adding prompts to the Art feature as well.

> *[Adding] sound for the breathing so I can like, close my eyes or not like, stare down at my phone.* [Participant 2]

> *I like the idea of like the prompts for drawing or like a coloring book idea.* [Participant 18]

> *Um, definitely the rate the mood thing: adding more moods so then people can kinda have more options.* [P23]

Enhancing Tracking Abilities

Having the ability to track emotions over time was also discussed by several participants. For instance, participants suggested that having a chart or graph within the app would be helpful to look back at and compare their emotions over time.

> *Some means of actually being able to properly like… track all of my moods and how I'm feeling… it would be a good way of like, expressing myself and then being able to reflect on it if I wanted to.* [Participant 3]

Providing Personalization

A few participants suggested offering users the ability to personalize the app based on their personality or preferences. Others thought personalization based on locations might be beneficial.

> *Having different versions of it for your personality types. Like certain things that are geared more towards your personality than others… there are a lot more creative and artistic people that would gear from the creative parts, but then there are the people who are less creative that will not use it at all or can't draw. So…different versions of it too for each personality type would definitely… help it.* [Participant 21]

> *Like, you could put your location in [the app] and then it will bring up the local places around you [helplines, resources].* [Participant 23]

Discussion

Principal Findings

Consistent with recent recommendations to evaluate the utility and usability of mHealth apps [18,23-25], this study sought to obtain user perspectives on the JoyPop app. Overall, we found that different features of the app varied in their uptake and perceived benefit to users. The app seemed to remind participants to take a moment for themselves to self-monitor and express their emotions, distract themselves from the stress of daily life, and gain emotional awareness and control strategies in the process. The establishment of routine, potentially via prompts, also appeared to facilitate app use. At the same time, certain feature limitations and a busy lifestyle seemed to dissuade use, and for a few participants, the app was perceived to add to their stress. Participants were eager to share many recommendations for feature additions and changes. Complementary to previous research on the effectiveness of the JoyPop app [9], the current study speaks to its usability and functionality and aspects that can be improved in this regard, as both effectiveness and usability are vital to establishing the merit of an mHealth app.

Certain aspects of participants' experience with the JoyPop app emerged as key facilitators of continued use. First, one major subtheme was the creation of routine, whereby the drive to use the app seemed to progress from wanting to adhere to study requirements to eventually desiring to use the JoyPop app and feeling it had become a habitual part of one's day. This idea of habit formation is intriguing, as it is consistent with prior work on the factors associated with technology uptake [56,57]. Habit formation constitutes a process by which app use comes to be associated with certain environmental or contextual cues, and such cues may trigger the behavior in the future with some degree of automaticity [56]. Accordingly, one feature of the study that may have encouraged habit formation was the inclusion of consistently timed email prompts (each morning, each evening) reminding participants to use the app. This may have encouraged an association between time of day and app use. In line with this, appreciation of these prompts emerged as another subtheme, consistent with prior research [19,21,24,27]. Outside of the mHealth app literature, it is encouraging that app use became routine for these study participants considering the merits of practicing self-regulation skills regularly. When a person uses strategies like those provided in the JoyPop app when they are not necessarily distressed, the strategies are well-practiced and ready to be used when distress does occur, potentially obviating the need for later reactive strategies [58]. It should be noted, however, that a few participants commented that if they were too busy with school life, they would use the app less, suggesting that the power of routine may not be strong enough to promote app use during especially stressful times for students.

Other facilitators of app use endorsed by participants were the self-monitoring and expressive opportunities provided by the JoyPop app. Consistent with this, the features that most clearly support these functions were among the most used and most helpful features according to participants, such as the Rate My

Mood and Journal features. Prior studies have consistently indicated that self-monitoring and tracking of one's mood is a highly desired feature of mental health–related apps [18,21,24,27], and our findings are no exception. The importance of self-monitoring to participants was further demonstrated by their desire for more tracking abilities within the app, such as being able to look back at one's mood ratings or see charts and graphs of mood changes over time. This is consistent with prior research indicating that enhanced tracking ability is one of the most common app features requested by users [13,18]. Self-monitoring and the ability to identify patterns in emotional responding is integral to many forms of psychotherapy [59], and it seems to be an aspect that participants also value and desire to engage in further.

Related to the self-monitoring and expressive opportunities that facilitated app use were the perceived outcomes described by participants of checking in with oneself, increased awareness, and emotional control. The JoyPop app appeared to serve as a signal to participants to pause their day to check in on how they were doing and assess how they might improve their mood if needed. By engaging in this self-monitoring and expression of emotion, the participants seemed to gain awareness and understanding of their emotions to better regulate them. Specifically, they indicated that the Journal and Rate My Mood features helped them learn to label and communicate their feelings, while the Breathing Exercises also provided a more direct form of emotion regulation. These findings are in line with general theories of self-regulation, which emphasize a connection between self-reflection, understanding, and regulation [31,60]. Further, the participants' perceptions of improved emotion regulation were mirrored by our quantitative results, which spoke to a reduction in emotion dysregulation with each additional day of app use over 4 weeks [9]. As such, there is encouraging convergence of quantitative and qualitative findings.

Another outcome of app use was that participants found it acted as a helpful distraction. Both SquareMoves and the Journal feature were mentioned within this subtheme, such that participants reported they felt these activities helped them take their mind off negative emotions and thoughts and focus on something else. Interestingly, in our tally of feature use and helpfulness, SquareMoves was indicated as one of the most used features; however, it was rarely mentioned as the most helpful feature. This finding suggests that it was the entertainment or distraction value of the feature that fostered continued use. Models of technology acceptance highlight entertainment factors as an important motive to using mHealth apps [21,57], such that "gamification," or including game-like design elements, can be useful to increase user engagement with apps [61]. As such, the helpfulness of SquareMoves may have been masked by its entertainment value; however, distraction can still function as a valid form of emotion regulation, particularly in situations when emotion is overwhelming [62]. In essence, some students may have used the app for fun or as a distraction, which can still indirectly support emotion regulation.

In terms of constructive criticism, participants described barriers to app use and recommendations for app improvements. They desired enhanced functionality of certain features, such as the Journal (more prompts, the ability to edit entries) and the Rate My Mood feature (greater variety of mood states). They also desired more features overall and were forthcoming with suggestions about features that should be added. This is in line with prior research on mHealth apps, which suggests that app users desire a sufficient number and diverse array of features that are also flexible enough to accommodate various needs [18,21,25,27]. Indeed, reviews of mood-monitoring apps on popular app download sites often cite a lack of features as a primary reason for user dissatisfaction with apps [27]. Further, participants in our study recommended that features should be more personalized. Similarly, previous studies have found that users want customization and some degree of control over app features [21,25,27,37,63]. Finally, captured under the subtheme of increased stress, 2 participants commented that the appearance of the app contained too many buttons, distracting colors, and repetitive appearance of quotes, which rendered the app less helpful. Appearance and design tend to be major themes of user feedback on mHealth apps [18,21,24,25,27,29,37,64], and users may not use an app that is not user friendly even if they can see its benefit [24]. Therefore, these comments represent important feedback. However, the relatively few comments of this nature suggest that the general look-and-feel of the JoyPop app was acceptable to most participants.

Certain features did not appear instrumental to the uptake and perceived helpfulness of the JoyPop app. These included the Circle of Trust and Art features, which were either rarely mentioned or mentioned as the least used and least helpful features. With respect to the Circle of Trust feature, this finding is somewhat surprising, as having some aspect of social support and feeling part of a community have been cited as important to app users in prior research [18,25,37,63]. The social support in these studies, however, pertained to the ability to share app data with other users or in an online forum, including with close others or anonymously with other app users [21,24,25,37]. This is different from the social support included in the JoyPop app, which allows users to input and contact family, peers, and other support persons in times of need. Peng et al [21] noted that some participants felt that seeing others using the app through data sharing would motivate them to use the app. Similarly, our results indicated some participant desire to enhance the social aspect of the app in a similar fashion, such as being able to communicate with others directly in the app so that users might hold each other accountable. With respect to the Art feature, this activity may have been relatively unused due to the inability to save drawings, which was one drawback expressed by participants. However, ongoing evaluation among more diverse groups would be warranted, as it is possible that our student sample, which consisted mainly of students pursuing health or science degrees, may have been less interested in or drawn to the Art feature.

Strengths, Limitations, and Future Directions

This study had many strengths. Obtaining user input on experiences with the JoyPop app and their perspectives on its utility aligns with recommendations of the value of user involvement in the design of digital health interventions [64-67]. With respect to data collection and analysis, the authors who

conducted the analyses were independent from the study conceptualization and data collection to avoid interpretation bias. Additionally, the questions chosen for the interview guide served several functions that are important to comprehensive app evaluations, including discovering what is important to users for long-term uptake of the app, which remains an ongoing issue for many app endeavors [26].

A limitation to the current study is that participants were predominantly female, which reflects the composition of the larger sample of participants who used the app from which this subsample was drawn [9]. Certainly, there are potential gender differences in youths' preferences for and use of mental health-based apps [37]. Male students are also relatively underrepresented in studies of mental health interventions for university students in general [68]. Given that our sample consisted of university students in their late teens, additional research examining younger youths' perspectives on the app will be important before wider implementation occurs.

A further limitation is the variation in the number of participants who attended each interview session, which ranged from 1 to 5. A larger group may have resulted in some participants sharing more than others, which may have created differences in how much certain perspectives were amplified. In addition, participants may have been less willing to share negative feedback about the JoyPop app in the presence of the interviewers; however, certain questions were meant to specifically elicit negative or constructive feedback (eg, least used/least helpful features, barriers to use, recommendations for improvement), which should have encouraged participants to provide honest feedback about the app.

Continuing to seek and incorporate feedback from users into evolving versions of the JoyPop app is an important future direction. Understanding user perspectives on the utility, features, and design of the app will aid in ensuring its continued use and engagement [64-67]. Consistent with participant recommendations to add more features and enhance existing features, a new SleepEase feature offering relaxation exercises and tips for preparing the body for sleep has been added to the latest version [69]. In addition, more journal prompts have been added, and users are now informed during the orientation to the app about methods they can use to save their Art drawings (eg, by screenshotting and saving them directly to their devices). Consideration of participants' suggestions for increased personalization (eg, location-based and culture-related customization) of the JoyPop app will be important in future updates; for instance, the latest iteration of the JoyPop app includes a French-language version.

Although the JoyPop app was designed specifically to draw attention toward positive mood states, consistent with consumer desire for positivity in mHealth apps [20,40], it is important to consider participant requests for opportunities to rate and track a range of emotions over time. Should additional mood ratings and tracking abilities be added to subsequent versions of the JoyPop app, ongoing evaluation of the consequences of this change would be warranted to ensure that no unintended negative consequences emerge (eg, worse mood after reviewing past mood ratings). Following recommendations in two recent reviews, additional research exploring the economic impact of app-based interventions is an important next step [70,71]. For instance, it will be important to examine the economic impact, including cost-effectiveness and cost utility, of integrating an intervention such as the JoyPop app into school or mental health settings for youth. Finally, due to ongoing disruptions related to the COVID-19 pandemic, including prolonged changes to education, mental health service delivery, and social practices, optimizing digital approaches to service delivery (eg, mHealth apps) is a growing global priority [72,73]. As such, exploring whether the JoyPop app can aid in buffering against the negative effects of stress brought on by the ongoing global pandemic, or how users' perspectives on the app may vary within the context of the pandemic, is warranted.

Conclusion

In sum, this qualitative study of users' experiences with the JoyPop app provides insight into aspects of the app that motivated and benefitted users, as well as actions that can be taken to improve user experiences and promote longer-term uptake. Students in our study demonstrated a willingness to engage with the app and incorporate it into their routine, while valuing the ability to self-monitor, express emotion, and engage in distraction. They preferred and used some features over others and were forthcoming with various suggestions for improvement. Their feedback underscores the value of considering user input in the continual development and evolution of apps, and it will be instrumental in ensuring the JoyPop app meets the needs of its users.

Conflicts of Interest

The majority of the authors have no conflicts of interest to declare. CW is the creator of the JoyPop app. To mitigate any risk related to conflict of interest, CW was not involved in collecting or analyzing the data. CW's main role was to support the team in evaluating the JoyPop app, and to liaise between the research team and app developers. CW also reviewed the manuscript prior to submission.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 34 KB - mhealth_v9i7e28677_app1.pdf]

References

1. Masten AS, Barnes AJ. Resilience in children: developmental perspectives. Children (Basel) 2018 Jul 17;5(7) [FREE Full text] [doi: 10.3390/children5070098] [Medline: 30018217]
2. Wright M, Masten A, Narayan A. Resilience processes in development: four waves of research on positive adaptation in the context of adversity. In: Handbook of Resilience in Children. Boston, MA: Springer; 2013:15-37.
3. Buckner JC, Mezzacappa E, Beardslee WR. Characteristics of resilient youths living in poverty: the role of self-regulatory processes. Dev Psychopathol 2003;15(1):139-162. [doi: 10.1017/s0954579403000087] [Medline: 12848439]
4. Buckner JC, Mezzacappa E, Beardslee WR. Self-regulation and its relations to adaptive functioning in low income youths. Am J Orthopsychiatry 2009 Jan;79(1):19-30 [FREE Full text] [doi: 10.1037/a0014796] [Medline: 19290722]
5. Espeleta HC, Brett EI, Ridings LE, Leavens ELS, Mullins LL. Childhood adversity and adult health-risk behaviors: Examining the roles of emotion dysregulation and urgency. Child Abuse Negl 2018 Aug;82:92-101 [FREE Full text] [doi: 10.1016/j.chiabu.2018.05.027] [Medline: 29879586]
6. Fritz J, de Graaff AM, Caisley H, van Harmelen A, Wilkinson PO. A systematic review of amenable resilience factors that moderate and/or mediate the relationship between childhood adversity and mental health in young people. Front Psychiatry 2018;9:230 [FREE Full text] [doi: 10.3389/fpsyt.2018.00230] [Medline: 29971021]
7. Poole JC, Dobson KS, Pusch D. Anxiety among adults with a history of childhood adversity: Psychological resilience moderates the indirect effect of emotion dysregulation. J Affect Disord 2017 Aug 01;217:144-152. [doi: 10.1016/j.jad.2017.03.047] [Medline: 28410477]
8. Ungar M. Resilience after maltreatment: the importance of social services as facilitators of positive adaptation. Child Abuse Negl 2013;37(2-3):110-115 [FREE Full text] [doi: 10.1016/j.chiabu.2012.08.004] [Medline: 23260114]
9. MacIsaac A, Mushquash AR, Mohammed S, Grassia E, Smith S, Wekerle C. Adverse childhood experiences and building resilience with the JoyPop app: evaluation study. JMIR mHealth uHealth 2021 Jan 04;9(1):e25087 [FREE Full text] [doi: 10.2196/25087] [Medline: 33393908]
10. Wekerle C. Finding joy in an app. Futurum Careers. 2019. URL: https://doi.org/10.33424.FUTURUM02 [accessed 2021-05-01]
11. Prentice J, Dobson K. A review of the risks and benefits associated with mobile phone applications for psychological interventions. Can Psychol 2014 Nov;55(4):282-290 [FREE Full text] [doi: 10.1037/a0038113]
12. Turvey CL, Roberts LJ. Recent developments in the use of online resources and mobile technologies to support mental health care. Int Rev Psychiatry 2015;27(6):547-557. [doi: 10.3109/09540261.2015.1087975] [Medline: 26523397]
13. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
14. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940. [doi: 10.2196/10940] [Medline: 30026178]
15. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
16. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
17. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
18. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
19. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
20. Lui J, Marcus D, Barry C. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pr 2017 Jun;48(3):199-210 [FREE Full text] [doi: 10.1037/pro0000122]
21. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
22. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
23. Kertz S, MacLaren Kelly J, Stevens K, Schrock M, Danitz S. A review of free iPhone applications designed to target anxiety and worry. J Technol Behav Sci 2017 Jan 11;2(2):61-70 [FREE Full text] [doi: 10.1007/s41347-016-0006-y]

24. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
25. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
26. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
27. Widnall E, Grant CE, Wang T, Cross L, Velupillai S, Roberts A, et al. User perspectives of mood-monitoring apps available to young people: qualitative content analysis. JMIR mHealth uHealth 2020 Oct 10;8(10):e18140 [FREE Full text] [doi: 10.2196/18140] [Medline: 33037875]
28. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]
29. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR mHealth uHealth 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
30. Kauer SD, Reid SC, Crooke AHD, Khor A, Hearps SJC, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
31. Barrett L, Gross J, Christensen T, Benvenuto M. Knowing what you're feeling and knowing what to do about it: mapping the relation between emotion differentiation and emotion regulation. Cogn Emot 2001 Nov;15(6):713-724 [FREE Full text] [doi: 10.1080/02699930143000239]
32. Arch JJ, Craske MG. Mechanisms of mindfulness: emotion regulation following a focused breathing induction. Behav Res Ther 2006 Dec;44(12):1849-1858. [doi: 10.1016/j.brat.2005.12.007] [Medline: 16460668]
33. Pascual-Leone A, Yeryomenko N, Morrison O, Arnold R, Kramer U. Does feeling bad, lead to feeling good? Arousal patterns during expressive writing. Rev Gen Psychol 2016 Sep 01;20(3):336-347 [FREE Full text] [doi: 10.1037/gpr0000083]
34. King L. The health benefits of writing about life goals. Pers Soc Psychol Bull 2016 Jul 02;27(7):798-807 [FREE Full text] [doi: 10.1177/0146167201277003]
35. Rankin K, Walsh LC, Sweeny K. A better distraction: exploring the benefits of flow during uncertain waiting periods. Emotion 2019 Aug;19(5):818-828. [doi: 10.1037/emo0000479] [Medline: 30265082]
36. Dalebroux A, Goldstein T, Winner E. Short-term mood repair through art-making: positive emotion is more effective than venting. Motiv Emot 2008 Nov 1;32(4):288-295 [FREE Full text] [doi: 10.1007/s11031-008-9105-1]
37. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]
38. Johnson K, Kalkbrenner M. The utilization of technological innovations to support college student mental health: mobile health communication. J Technol Hum Serv 2017 Sep 11;35(4):314-339 [FREE Full text] [doi: 10.1080/15228835.2017.1368428]
39. Bell T, Romano E, Flynn RJ. Profiles and predictors of behavioral resilience among children in child welfare. Child Abuse Negl 2015 Oct;48:92-103. [doi: 10.1016/j.chiabu.2015.04.018] [Medline: 26002600]
40. Park T, Thompson K, Wekerle C, Al-Hamdani M, Smith S, Hudson A, et al. Posttraumatic stress symptoms and coping motives mediate the association between childhood maltreatment and alcohol problems. J Trauma Stress 2019 Dec;32(6):918-926. [doi: 10.1002/jts.22467] [Medline: 31742776]
41. Sanders J, Munford R, Thimasarn-Anwar T, Liebenberg L, Ungar M. The role of positive youth development practices in building resilience and enhancing wellbeing for at-risk youth. Child Abuse Negl 2015 Apr;42:40-53. [doi: 10.1016/j.chiabu.2015.02.006] [Medline: 25770347]
42. Strickland N, Wekerle C, Kehayes I, Thompson K, Dobson K, Stewart S. Self-compassion as a compensatory resilience factor for the negative emotional outcomes of alcohol-involved sexual assault among undergraduates. Int J Child Adolesc Resil 2019:52-69 [FREE Full text] [doi: 10.7202/1069076ar]
43. Tanaka M, Wekerle C, Schmuck ML, Paglia-Boak A, Research Team MAP. The linkages among childhood maltreatment, adolescent mental health, and self-compassion in child welfare adolescents. Child Abuse Negl 2011 Oct;35(10):887-898. [doi: 10.1016/j.chiabu.2011.07.003] [Medline: 22018519]
44. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
45. Bewick B, Koutsopoulou G, Miles J, Slaa E, Barkham M. Changes in undergraduate students' psychological well-being as they progress through university. Stud High Educ 2010 Aug 18;35(6):633-645 [FREE Full text] [doi: 10.1080/03075070903216643]
46. Kowalski K, McHugh T, Ferguson L, Sabiston C. Research Methods in Kinesiology. North York, ON: Oxford University Press; 2018.

47. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
48. Warner G, Kervin E, Pesut B, Urquhart R, Duggleby W, Hill T. How do inner and outer settings affect implementation of a community-based innovation for older adults with a serious illness: a qualitative study. BMC Health Serv Res 2021 Jan 07;21(1):42-13 [FREE Full text] [doi: 10.1186/s12913-020-06031-6] [Medline: 33413394]
49. Doing pedagogical research in engineering. Engineering Centre for Excellence in Teaching and Learning. 2011. URL: https://core.ac.uk/download/pdf/288385171.pdf [accessed 2021-06-30]
50. Morgan D. Focus Groups as Qualitative Research, Volume 16. Newbury Park, CA: SAGE Publications; 1996.
51. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
52. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research: norms and guidelines for CSCW and HCI practice. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-23 [FREE Full text] [doi: 10.1145/3359174]
53. Irwin JD, He M, Bouck LMS, Tucker P, Pollett GL. Preschoolers' physical activity behaviours: parents' perspectives. Can J Public Health 2005;96(4):299-303 [FREE Full text] [Medline: 16625802]
54. Creswell J. A Concise Introduction to Mixed Methods Research. Newbury Park, CA: SAGE publications; 2014.
55. Guba E. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Commun Technol 1981;29(2):75.
56. Venkatesh V, Thong, Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157 [FREE Full text] [doi: 10.2307/41410412]
57. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
58. Durand-Bush N, McNeill K, Harding M, Dobransky J. Investigating stress, psychological well-being, mental health functioning, and self-regulation capacity among university undergraduate students: Is this population optimally functioning? Can J Couns Psychother 2015:253-274 [FREE Full text]
59. Cohen J, Edmunds J, Brodman D, Benjamin C, Kendall P. Using self-monitoring: Implementation of collaborative empiricism in cognitive-behavioral therapy. Cogn Behav Pract 2013 Nov;20(4):419-428 [FREE Full text] [doi: 10.1016/j.cbpra.2012.06.002]
60. Lyons K, Zelazo P. Monitoring, metacognition, and executive function: elucidating the role of self-reflection in the development of self-regulation. In: Advances in Child Development and Behavior. Cambridge, MA: Academic Press; 2011.
61. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. In: MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; September 28-30, 2011; Tampere, Finland URL: https://doi.org/10.1145/2181037.2181040
62. Sheppes G, Scheibe S, Suri G, Gross JJ. Emotion-regulation choice. Psychol Sci 2011 Nov;22(11):1391-1396. [doi: 10.1177/0956797611418350] [Medline: 21960251]
63. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR mHealth uHealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
64. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
65. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR mHealth uHealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
66. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
67. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
68. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
69. Clearbridge mobile adds new features to mental health app to help adolescents and young adults increase psychological resilience. Cision. 2021 Feb 09. URL: https://www.newswire.ca/news-releases/clearbridge-mobile-adds-new-features-to-mental-health-app-to-help-adolescents-and-young-adults-increase-psychological-resilience-820071240.html [accessed 2021-05-01]
70. Badawy SM, Kuhns LM. Economic evaluation of text-messaging andsmartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR mHealth uHealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]

71. 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]
72. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
73. Serlachius A, Badawy SM, Thabrew H. Psychosocial challenges and opportunities for youth with chronic health conditions during the COVID-19 pandemic. JMIR Pediatr Parent 2020 Oct 12;3(2):e23057 [FREE Full text] [doi: 10.2196/23057] [Medline: 33001834]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 10.03.21; peer-reviewed by A Sarradon-Eck, S Badawy, MD, MS; comments to author 31.03.21; revised version received 31.05.21; accepted 31.05.21; published 08.07.21.

Please cite as:
Mushquash AR, Pearson ES, Waddington K, MacIsaac A, Mohammed S, Grassia E, Smith S, Wekerle C
User Perspectives on a Resilience-Building App (JoyPop): Qualitative Study
JMIR Mhealth Uhealth 2021;9(7):e28677
URL: https://mhealth.jmir.org/2021/7/e28677
doi:10.2196/28677
PMID:34255696

Original Paper

Exploring Associations Between Children's Obesogenic Behaviors and the Local Environment Using Big Data: Development and Evaluation of the Obesity Prevention Dashboard

Dimitris Filos[1], MSc, PhD; Irini Lekka[1], BSc, MSc; Vasileios Kilintzis[1], BSc, MSc, PhD; Leandros Stefanopoulos[1], BSc, MSc; Youla Karavidopoulou[1], BSc, MSc; Christos Maramis[1], MSc, PhD; Christos Diou[2,3], MSc, PhD; Ioannis Sarafis[3], PHO; Vasileios Papapanagiotou[3], PhD; Leonidas Alagialoglou[3]; Ioannis Ioakeimidis[4], PhD; Maria Hassapidou[5], PhD; Evangelia Charmandari[6], MD, PhD; Rachel Heimeier[7], PhD; Grace O'Malley[8], PhD; Shane O'Donnell[9], PhD; Gerardine Doyle[10], BSc, PhD; Anastasios Delopoulos[3], MSc, PhD; Nicos Maglaveras[1], MSc, PhD

[1]Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, Aristotle University, Thessaloniki, Greece

[2]Department of Informatics and Telematics, Harokopio University of Athens, Athens, Greece

[3]Multimedia Understanding Group, Aristotle University, Thessaloniki, Greece

[4]Department of Biosciences and Nutrition, Karolinska University, Stockholm, Sweden

[5]International Hellenic University, Thessaloniki, Greece

[6]Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[7]Internationella Engelska Skolan, Stokholm, Sweden

[8]School of Physiotherapy, Division of Population Health Sciences, RCSI University of Medicine and Health Sciences, Dublin, Ireland

[9]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland

[10]College of Business, University College Dublin, Dublin, Ireland

Corresponding Author:
Nicos Maglaveras, MSc, PhD
Lab of Computing, Medical Informatics and Biomedical Imaging Technologies
Aristotle University
University Campus
Box 323
Thessaloniki, 54124
Greece
Phone: 30 2310999281
Fax: 30 2310999263
Email: nicmag@auth.gr

Abstract

Background: Obesity is a major public health problem globally and in Europe. The prevalence of childhood obesity is also soaring. Several parameters of the living environment are contributing to this increase, such as the density of fast food retailers, and thus, preventive health policies against childhood obesity must focus on the environment to which children are exposed. Currently, there are no systems in place to objectively measure the effect of living environment parameters on obesogenic behaviors and obesity. The H2020 project "BigO: Big Data Against Childhood Obesity" aims to tackle childhood obesity by creating new sources of evidence based on big data.

Objective: This paper introduces the Obesity Prevention dashboard (OPdashboard), implemented in the context of BigO, which offers an interactive data platform for the exploration of objective obesity-related behaviors and local environments based on the data recorded using the BigO mHealth (mobile health) app.

Methods: The OPdashboard, which can be accessed on the web, allows for (1) the real-time monitoring of children's obesogenic behaviors in a city area, (2) the extraction of associations between these behaviors and the local environment, and (3) the evaluation of interventions over time. More than 3700 children from 33 schools and 2 clinics in 5 European cities have been monitored using a custom-made mobile app created to extract behavioral patterns by capturing accelerometer and geolocation data. Online databases were assessed in order to obtain a description of the environment. The dashboard's functionality was evaluated during a focus group discussion with public health experts.

Results: The preliminary association outcomes in 2 European cities, namely Thessaloniki, Greece, and Stockholm, Sweden, indicated a correlation between children's eating and physical activity behaviors and the availability of food-related places or sports facilities close to schools. In addition, the OPdashboard was used to assess changes to children's physical activity levels as a result of the health policies implemented to decelerate the COVID-19 outbreak. The preliminary outcomes of the analysis revealed that in urban areas the decrease in physical activity was statistically significant, while a slight increase was observed in the suburbs. These findings indicate the importance of the availability of open spaces for behavioral change in children. Discussions with public health experts outlined the dashboard's potential to aid in a better understanding of the interplay between children's obesogenic behaviors and the environment, and improvements were suggested.

Conclusions: Our analyses serve as an initial investigation using the OPdashboard. Additional factors must be incorporated in order to optimize its use and obtain a clearer understanding of the results. The unique big data that are available through the OPdashboard can lead to the implementation of models that are able to predict population behavior. The OPdashboard can be considered as a tool that will increase our understanding of the underlying factors in childhood obesity and inform the design of regional interventions both for prevention and treatment.

(JMIR Mhealth Uhealth 2021;9(7):e26290) doi:10.2196/26290

KEYWORDS

public health authorities; childhood obesity; children's behavior; environment; COVID-19; big data; mHealth; uHealth; intervention

Introduction

Obesity remains a major health problem worldwide, increasing the risk for the development of noncommunicable diseases, such as diabetes, coronary heart disease, and cancer [1,2]. Its global prevalence has increased dramatically in the last 40 years, resulting in a great economic burden for health care systems [3]. A concerning increase in overweight and obesity among children has also been noted. In 2013, over 42 million children were considered overweight or obese, and approximately 70 million children will be overweight or obese by 2025 [4,5]. In Europe, the prevalence of overweight and obesity in children varies across countries [6], peaking at just over 40% [4,7]. Children with obesity are more likely to be obese until adulthood [8], thereby increasing the risk for chronic disease development [9,10].

Existing evidence indicates that interventions targeting various elements of children's behavioral patterns, such as *what* and *how they eat* and *how they move* [11], in addition to the living environmental factors to which they are exposed (elsewhere referred to as "ecological" parameters), can yield positive outcomes [8]. To date, the primary sources of evidence for the evaluation of childhood obesity in a given area are self-reporting questionnaires, patient electronic health records, as well as a limited number of focused studies that use objective measurements reflecting children's behavioral patterns [8].

On a social level, certain environmental factors and socioeconomic parameters have been shown to be associated with obesity in children, such as (1) the ethnicity of the population [12] (pointing to a generalized genetic effect on a population level but mostly toward specific socioeconomic conditions and cultural effects [13], (2) the immediate socioeconomic environment (eg, the "neighborhood" [14], including the immediate proximity to fast food restaurants [15] and the availability of exercise facilities [16]), and (3) exposure to food-related advertising [17].

However, the local environmental context, including the local food environment (eg, the proximity to fast foods), and the local urban characteristics are particularly difficult to quantify, usually requiring extensive, expensive, and long-term studies, making it, until recently, extremely difficult to use those results in a context-based fashion in society.

Fortunately, the widespread adoption and maturation of new technologies, such as the use of mobile phones or smartwatches [18], especially by youth, allow for the collection of a significant volume of information under real-life conditions, which can lead to the extraction of user behavioral patterns [19,20]. Equally, the extensive use of the internet and social media has led to the creation of open databases, such as Google Maps, related to the availability of certain type of places in the local environment. These new technologies, elsewhere referred to as "mobile sensing" [21], open new horizons as a source of objective information. They allow for the more detailed evaluation of environmental effects, not only of the immediate place of residence of the individuals, but also of the areas which they frequent (a data collection challenge previously identified under the term "residential fallacy effect") [22].

The Big Data Against Childhood Obesity program [23] is a European-funded research project aiming to tackle the problem of childhood obesity using big data collected by children using their smartphones. BigO collects and analyzes big data on children's behavior and their environment to enable public health authorities to plan and execute effective programs addressing obesity. The BigO system has been developed as a tool for public health authorities to evaluate localized obesogenic behaviors at the population level and to facilitate decision making as well as intervention planning through a powerful analytical framework and a purpose-built dashboard.

This paper presents the Obesity Prevention Dashboard (OPdashboard), which has been developed in the context of the BigO study and can be used to visually explore aggregated children's localized behaviors and environment characteristics and to automatically extract associations among them. The potential users of the OPdashboard are health researchers in the domains of epidemiology and childhood obesity, as well as any relevant public health agency or authority. A custom-made

mobile app was used to collect big data from participating children and adolescents in 5 different European countries (see the *Data* section); the extracted information has been visualized in the OPdashboard. In brief, the functionality includes: (1) real-time monitoring and visualization of children's behavior, (2) extraction of associations between behavior and environment in the areas close to schools, (3) evaluation of interventions as well as (4) the design of interventions using behavioral prediction models. Each of the functionalities are described in more detail in the *OPdashboard Functionality* section. In the *Results* section, the effect of the living environment close to schools on children's behavior is described.

Lastly, we also explored the impact of COVID-19. The emergence of the COVID-19 pandemic resulted in the adoption of measures, at a national level, to decelerate disease spread in the population. Those interventions varied across countries, ranging from social distancing recommendations to partial or total lockdown. Data analysis based on wearable activity trackers has shown that home confinement policies affected physical activity levels in the general population [24]. An evaluation of these measures, in terms of behavioral changes associated with characteristics of the local environment, may reveal additional information to health authorities.

In addition, the effect of COVID-19–related measures implemented in Greece on the children's behavior is presented in the *Results* section. The evaluation of the OPdashboard's functionality and usefulness was performed by experts in the field of childhood obesity and public health policy, as described in the *Evaluation* section. The OPdashboard will be made freely available to the scientific community via a web interface after the conclusion of the BigO program. The expected impact of the OPdashboard, as well as its limitations, are discussed in the final section.

Methods

Data

Within BigO, the cornerstone of the data collection process is *citizen science*, which is a relatively new scientific approach to gather big data, where the general population actively contributes to scientific research [25]. In that vein, school students, aged 9 to 18 years, were recruited and following an informed parental consent process, agreed to participate in data collection as part of citizen science projects, using a custom mHealth (mobile health) app available for both Android and iOS devices [26]. Several questions related to children's daily mood and sleeping behavior are asked through the myBigO app. In addition, the app allows the users to submit photos related to eating behavior. Awards, in terms of virtual badges, are provided to frequent users in order to maximize adherence to data collection. Furthermore, objectively collected GPS and accelerometer data, via the myBigO app, were analyzed in order to extract aggregated population-level obesogenic behavioral indicators identified within a specific geographical region [20]. The geohash geocoding system [27] was adopted as a spatial structure describing a broader geographical area, the size of which varies according to the length of the codes considered. In the OPdashboard, geohashes of 6 and 7 digits were used to reflect an area of 1.22 km × 610 m and 153 m × 153 m on the equator, respectively, with a reduction in width the further away from the equator. The aggregation to geohash level was done as a privacy preservation measure in order to avoid potential subject identification from GPS data.

Per Diou et al [20], personal sensory data, such as GPS and accelerometry, were used to extract individual indicators. The aggregated behaviors were calculated either based on the unique individuals visiting a geohash, or based on all values across the visits, even if they are visits from the same individual, while various geohash sizes are supported through the OPdashboard. Furthermore, aggregations are also performed based on the individuals living in a geohash. The rationale behind this approach is the "residential fallacy" effect [22], which underlines the importance of the nonresidential places visited in order to estimate residential effects.

In order to have a more reliable estimation of a behavior in a geohash area, the aggregation was performed only in case more than 5 values were recorded there (ie, from unique visits or unique visitors, respectively). All data were transmitted with end-to-end encryption and stored on secure servers. The stored data did not contain any directly identifiable information (eg, names or emails).

In keeping with co-design principles, a Delphi panel was carried out with public health experts across Europe with a remit in obesity prevention in order to prioritize the elements of interest (behavior and environmental and socioeconomic parameters) that should be measured and visualized as part of the OPdashboard [28]. Proposals were then assessed based on the feasibility of the measurements during the BigO project, and finally 38 behavior indicators, 40 environmental parameters, and 6 socioeconomic parameters were defined.

Behavior indicators were distinguished into two main categories. The first category was related to eating behavior and addressed food consumption and frequency of visitations to food-related locations, whereas the second category was related to physical activity behavior and addressed the frequency of visits to sports-related places and measurements of physical activity, such as number of steps per day. More information related to the methodology adopted for the extraction of the behavioral indicators based on the raw accelerometry and geolocation data can be found in Papapanagiotou et al [29]. Table 1 provides a subset of the behavioral indicators that have been calculated in the context of BigO. Through the OPdashboard, only aggregated behaviors are presented and no individual data are available.

Similarly, 40 environment characteristics of interest were identified and included in the BigO system (Table 2). These broadly describe the availability of diverse type of places in a geographical region of a geohash and were collected through open and online data sources [20].

Finally, the socioeconomic characteristics of the local area were also highlighted by public health obesity experts as being crucial for the evaluation of childhood obesity [13]. In this respect, taking also into account the availability of the data in a regional level from national databases and Eurostat, a number of socioeconomic factors were also considered. However, this

information is not available at the geohash level and was thus measured in the broader geographical area of a municipality.

Table 1. A subset of the behaviors used in BigO.

Behavior	How it was computed
Average steps per hour across visits to the region	For each visit to the region, take the total number of steps during the visit, as well as the duration of the visit and compute the average steps/hour. This behavior is the average of these values across visits (even if they are visits from the same individual). For the computation of this indicator, only visits that last 1 minute or more are considered.
Average steps per hour across visitors of the region	For each unique visitor of the region, take the total number of steps during his or her visits, as well as the total time he or she spent in the region, and compute the average steps/hour. This indicator is the average of these values across unique visitors. For the computation of this indicator, only visits that last 1 minute or more are considered.
Percentage of visits to the region that include at least one visit to a food-related location	Compute the percentage of visits to the region that include at least one visit to a food-related place. For the computation of this indicator, only visits that last 1 minute or more are considered.
Percentage of visits that include at least one visit to an athletics or sports facility	For each visit to the region, take the number of visits to an athletics or sports facility. This indicator is the average across visits. For the computation of this indicator, only visits that last 10 minutes or more are considered.
Average daily number of steps for residents of the region	For each resident of the region, compute his or her steps using the recorded data across all the areas that the resident visited. Based on this, compute his or her daily average number of steps for days with more than 60 minutes of recorded data (days with fewer data points are discarded). This indicator is the average of this value across the residents of the region but corresponds to the behaviors that might have happened anywhere on the map.

Table 2. A subset of the environmental characteristics used in BigO.

Environmental characteristic	How it was computed
Number of athletics or sports facilities in the region	Using publicly available data sources, compute the number of athletics or sports facilities in the region.
Number of fast food outlets in the region	Using publicly available data sources, compute the number of fast food outlets in the region.
Average number of restaurants within a 100-meter radius from locations within the region	Create a 30-meter point grid inside the region. For each point, compute the number of restaurants within a 100-meter radius. This value is the average across all points inside the region.

OPdashboard

The OPdashboard is a web application for public health scientists and policy makers, which supports decision making and planning of localized interventions to reduce the prevalence of childhood obesity. The OPdashboard is freely available and the beta version can be accessed online [30].

Requirements

Methodologically, the development of the OPdashboard has taken jointly into account (1) expert-driven predevelopment requirements and (2) mid-development expert evaluation with regards to relevance within the childhood obesity and public health domains.

Based on public health experts' knowledge, a set of mock-ups were designed. In order to ensure that the specific functionality of the OPdashboard meets the practical needs of public health authorities, focus groups and interviews with experts in nutrition, obesity, behavior, and public health were arranged to demonstrate and elicit feedback on the proposed functionality of the system. Based on the presented methodology and through systematic evaluation of the collected feedback, a list of requirements were defined reflecting both the functionality requirements and requirements relevant to the developed interface design.

According to the feedback from the focus groups, the following functionalities were recommended for integration into the OPdashboard:

- Explore: description of aggregated local data on children's behaviors, socioeconomics, and environment;
- Explain: visual comparisons between behaviors and the environment in a given area, as well as the extraction of associations;
- Compare: comparison of behaviors across time, which can be used to monitor and evaluate an intervention;
- Predict: simulation of behavioral change upon changes in an environment characteristic.

The first version of the OPdashboard focused on the first three of these functionalities. In the future, it is anticipated that predictive models of high accuracy will be developed based on the ongoing big data collection, supporting the design of future interventions.

Technical Implementation

All the data required by the OPdashboard are stored in a Mongo database (MongoDB), which is a cross-platform

document-oriented database management system. A number of RESTful web services [31] have been developed using the Jersey framework [32]. These services access the MongoDB with the help of the MongoDB Java Driver [33], using a Mongo account defined specifically for the service that is granted with the proper set of permissions (only read/select access on specific Mongo collections). The endpoints offered by the service require authorization via a valid, BigO-generated, JSON (JavaScript Object Notation) web token (JWT) [34].

The interactive web application was implemented using the R Studio Shiny server [35]. Shiny allows simple HTML pages to interactively execute R scripts and takes advantage of R's visualizing capabilities to graphically present the analysis results on the webpage. In order to consolidate the user experience, Shiny applications can be further extended by using JavaScript and jQuery. The R-script execution and the use of the analysis and visualization mechanisms inside the HTML pages are controlled via the user interface–rendering engine of the Shiny web server.

Extraction of Associations

Regarding the extraction of associations between the behaviors and the environment, a number of studies propose that the proximity of fast food outlets to schools may have an effect on adolescent eating patterns and may contribute to obesity [15,36]. In this respect, the analysis was based on the extraction of associations between the population behavior and the environment close to schools. For the definition of this area, only the geohashes where the distance between the center of the geohash and the location of the school is less than 1000 m were considered. For this analysis, a 7-digit geohash was used.

For these areas, each behavior detected within a geohash area was correlated with the environmental characteristics of the respective geohash in order to investigate the existence of a correlation between them. The Pearson correlation coefficient was computed which is defined as:

where X and Y denote the 2 variables, cov is the covariance, and σ the standard deviation of a variable. The P values for the estimation of the statistical significance of the correlation coefficients are also computed. The linear correlation analysis is made available on the dashboard in order to provide a preliminary understanding of the data.

It is worthy to mention that cultural and ethnic backgrounds also have the potential to affect users' behavior and their association with the environment [37], especially considering the case of BigO where data are distributed across different countries. In order to minimize the effect of possible confounders that can bias the results, the analysis was performed separately for each city.

Comparison in Time: The COVID-19 Paradigm

The compare functionality of the OPdashboard allows for the comparison of obesogenic behaviors in different time frames. These time frames represent the period before and after interventions that focus on the modification of obesogenic behaviors or exposure to obesogenic environments to support a healthier lifestyle. Due to the lack of typical obesity-related interventions in the before period, mainly because of the COVID-19 pandemic, the paradigm presented here was based on a scenario that monitored children's behavioral change as a result of the nation-wide measures undertaken to limit the spread of COVID-19 in Greece [38].

Greece adopted strict measures to restrain the pandemic, leading to the closure of schools on March 11, 2020, while a lockdown was implemented on March 23. Since the goal of the analysis was to investigate the effect of the adopted policies on children's behavior, the analysis was carried out on the Greek population, and in the city of Thessaloniki in particular, where the majority of the data were collected. The target behavior was steps per hour since visits to specific places were prohibited during the lockdown and only outdoor activity was allowed for all citizens.

The analysis was made for all 8 municipalities of Thessaloniki, while the behavior for each municipality was calculated as the aggregation of all the geohashes within the boundaries of each municipality where the behavior from at least one user was identified. For this analysis, 7-digit geohashes were considered.

In this respect, the aggregated steps per hour in each municipality of Thessaloniki, before and after school closure, were compared. The period before was defined as ranging from January 8, 2020, when schools in Greece opened after the Christmas holidays, to March 10, 2020. The after period was defined as March 11-31, 2020. Recruitment was low after this date, and data for the post–March 31 period was not considered in the analysis. The two-sample t test was used to investigate the existence of statistically significant differences in behavior across time.

Predictive Models

The OPdashboard allows for the downloading of the aggregated behaviors at the geohash level of detail. This functionality offers the possibility to further analyze the data and to implement models that are able to predict the behavior of the population based on the environment. Initial works carried out by the BigO team have used such data in order to predict children's physical activity, achieving a prediction accuracy of 81% [39]. However, the implementation of such models as well as the investigation of their accuracy is beyond the scope of this work. Nonetheless, such models have been integrated in the OPdashboard, offering a visual interpretation of behavior in a city when no behavioral data are available in specific regions.

In addition, the OPdashboard offers the possibility to design specific interventions in a region and predict their effect on certain obesogenic behaviors in children. The design of the interventions can be made graphically by modifying the characteristics of the environment (eg, adding or removing public parks in a region), while the effect of the intervention can be assessed using the map view.

OPdashboard Evaluation

The OPdashboard was demonstrated to a wide audience during the European Childhood Obesity Group Conference while a focus group comprising experts in public health, statistical

epidemiology, and clinical endocrinology was organized in order to obtain feedback on the dashboard's usability.

After the live demonstration of the OPdashboard, the following questions were posed to the experts:

- In your opinion, what is the value of the BigO system for local population behavioral assessment?
- In your opinion, what is the value of the BigO system in discovering associations between localized behaviors and environments? What additional steps would be necessary to discover causal relationships?
- In your opinion, which users/public authorities would most benefit from BigO? (eg, epidemiologists, population health consultants, local governments, educational policy makers)
- In your opinion, what are the main barriers for the adoption of BigO by public health authorities?

This discussion was recorded using a microphone, and a report was produced. The outcome of this focus group helped define the improvements needed and are described in the *Evaluation* section.

Results

Data Overview

The OPdashboard constitutes a live system, where more and more schools are added over time. By April 2020, BigO was deployed in 33 schools and 2 pediatric clinics in 5 different European cities (Athens, Thessaloniki, and Larissa from Greece; Stockholm from Sweden; and Dublin from Ireland) [40]. Ethical approval was received from each country separately. Each child was asked to use the myBigO app for at least 2 weeks. A steady flow of recruitment was followed; however, an expected decrease occurred during the summer because of the summer holidays and school closure. Most of the children used the app during the 2-week period, while a small minority continued to use it for more than 2 weeks

More than 3700 children, aged 9 to 18 years, participated in data collection, between the start of March 2018 until the end of March 2020, providing approximately 107 years of accelerometry data and 73 years of GPS data [40]. This was the first time such an amount of big data has been collected contemporaneously, on an individual basis within a region, allowing for the extraction of childhood obesogenic behavior accurately. The majority of the schools included in the analysis are located in the city of Thessaloniki, Greece, and Stockholm, Sweden, and thus the analysis in the following sections will be based on the schools from these cities only, as seen in Table 3.

Table 3. Details on the collected data from Thessaloniki, Greece, and Stockholm, Sweden.

Characteristic	Location	
	Thessaloniki	Stockholm
Children, n	839	671
Age (years), mean (SD)	13.5 (2.2)	15.3 (1.8)
Sex, n		
Male	454	375
Female	380	265
Not available	5	31
Z-score >1 (%)	26.1	18.3
Schools, n	24	5
Accelerometer (hours)	179,969.48	78,903.17
GPS (hours)	112,624.13	225,337.67

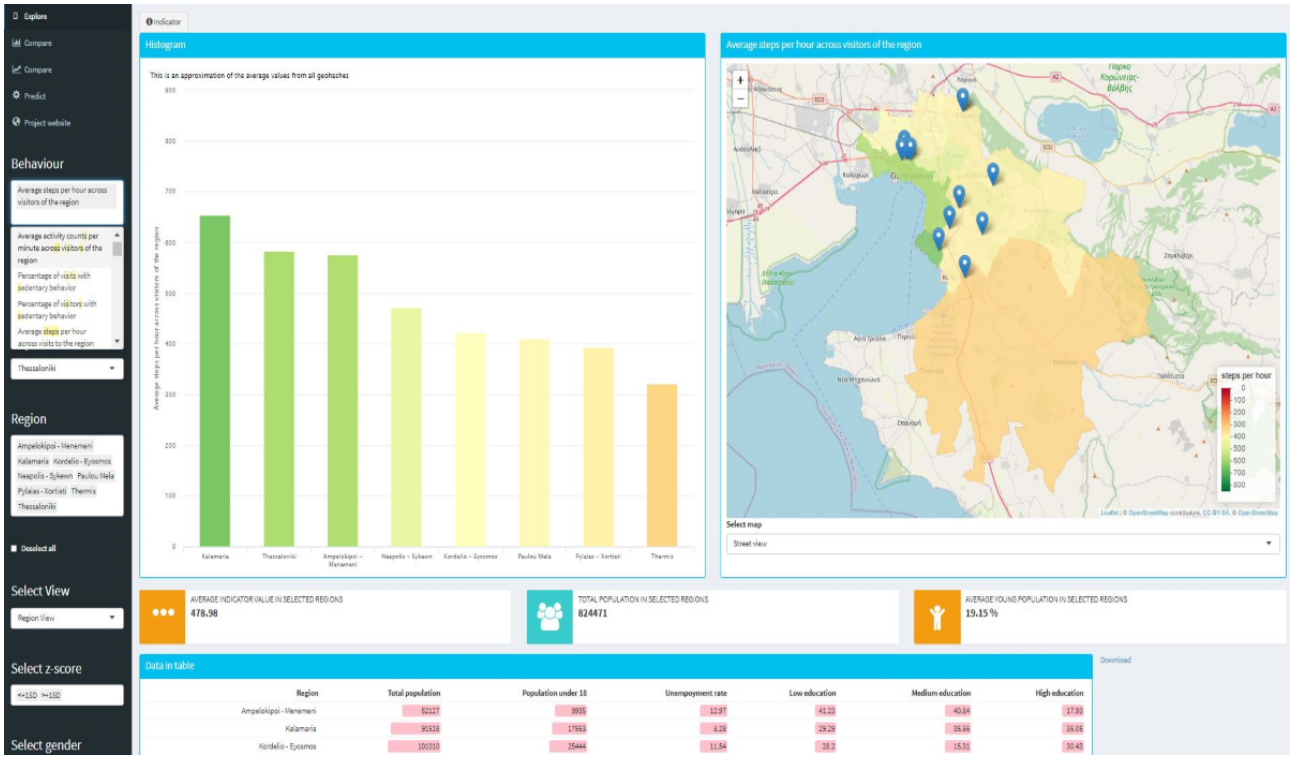
OPdashboard Functionality

In this section, the current version of the OPdashboard is presented. It focused on the data exploration ("explore") and the extraction of associations ("explain") between children's behavior and the environment close to schools. Regarding the "compare" functionality, the analysis focused on the comparison of the population's steps per hour before and after the implementation of national health policies due to COVID-19. The "predict" functionality was not presented here for the reasons made clear earlier.

Data Exploration Overview

The main screen of the web interface is shown in Figure 1. It is divided into two main areas, one focusing on the selection of the variables to be analyzed (control panel), whereas in the second area the results are visualized (main panel). The available selections include behavior and environmental characteristics as well as the city of interest, in terms of the broader metropolitan area, which can be further divided into the level of available municipalities.

Figure 1. The OPdashboard main page with the control panel (left) and the main panel (right).



Furthermore, the user can select the representation level, including the region view, where the population-aggregated value of the variable is depicted for the selected municipalities of the city, and the geohash view, where the aggregation is made at the geohash level. Filters related to the BMI z-score and gender can also be applied in order to focus on specific groups of children. Finally, a list of regional socioeconomic information related to (1) the total population, (2) their educational level, (3) the percentage of young people, and (4) the unemployment rate is also provided in a tabular format.

At the top of the analysis panel, the button serves as a tool to provide additional information regarding the selected behavior or environment characteristic, highlighting its importance as well as describing how it was measured.

The dashboards are provided within the analysis panel, allowing the user to have a clear overview of the indicators and their distribution across the regions. For data exploration, both bar charts and map views are provided to the user. A tabular representation of the socioeconomic data for the selected regions is provided as this information can be considered as a confounder contributing to a specific behavior or environmental characteristic.

The bar charts provide in a sorted format the aggregated value in each municipality of the city. This comprehensive visualization allows for the direct comparison of different variables across several municipalities of the city, as well as between each municipality and the average value across the selected regions.

The map view allows user to quantify the value of each variable within a geographical area. This information can be depicted as an aggregation of the behavior across the municipality, or it can be represented in more detail using the geohash visualization (Figure 2). Blue pins on the map highlight the location of the schools that have been included in the BigO pilots.

The coloring format used in the aforementioned visualizations ranges from red to green or vice versa, according to the type of indicator. In other words, the coloring changes from red to green as the number of steps per hour increases. On the other hand, when the small value represents a healthy state (eg, low number of visits to fast food restaurants, low number of food outlets in the region), then the coloring changes from green to red to indicate an increase in the number of food outlets in the region. The maximum value for each variable was extracted either by a review of the literature, or derived as the 95th percentile based on the collected data. In addition, the OPdashboard allows for a visual comparison between behaviors and/or the environment through bar charts (Figure 3). The user can extract the data, in a JSON format, in order to perform further analysis.

The visualization of a behavior at the geohash level of detail was made only for those geohashes where a behavior was observed.

Figure 2. Geohash distribution of food-related locations within a 100-meter radius from each geohash (left) and the steps per hour (right) for the city of Thessaloniki. Using mouseover, the user can access more information related to the number of contributors and the actual value in each geohash. Note: the geohash is visible only in cases where the aggregation was based on more than 5 contributions. The visual exploration reveals the high density of food-related places in the downtown of Thessaloniki and high activity among the children near the coastline of the city. The blue pins denote the location of schools. Source: OpenStreetMap [41].

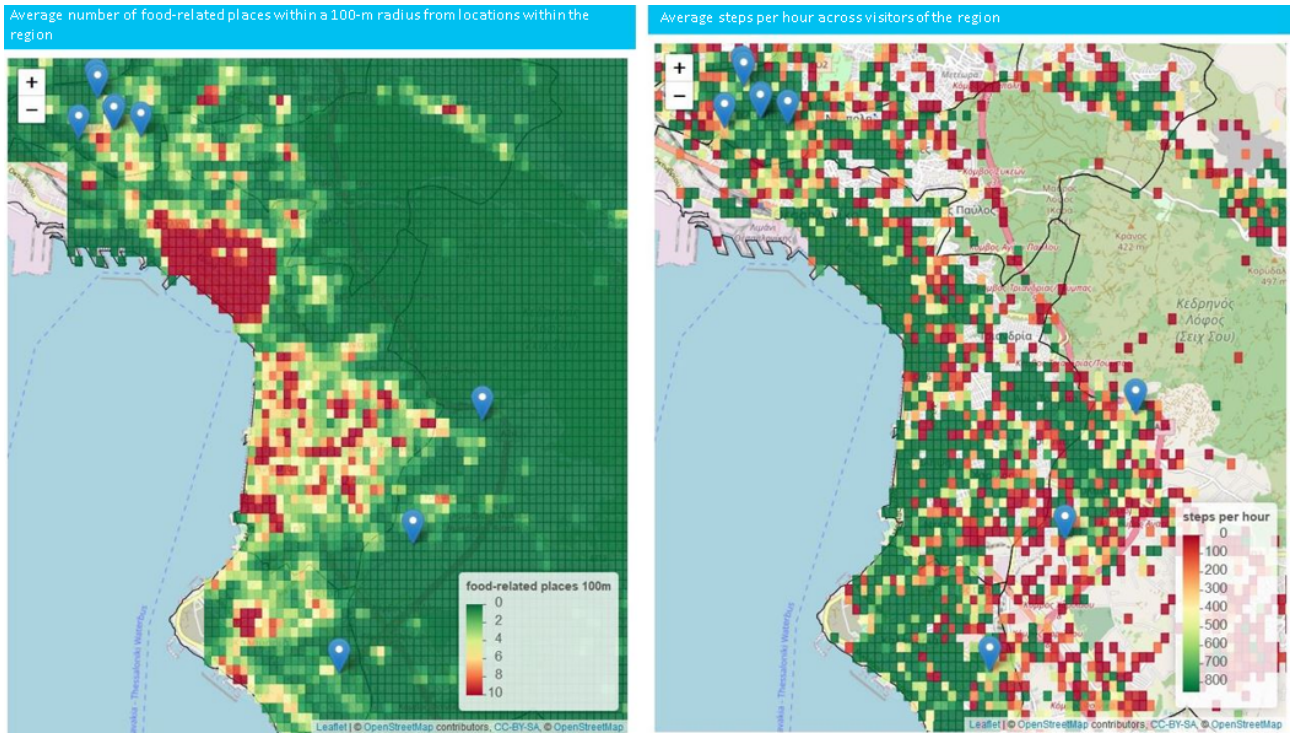
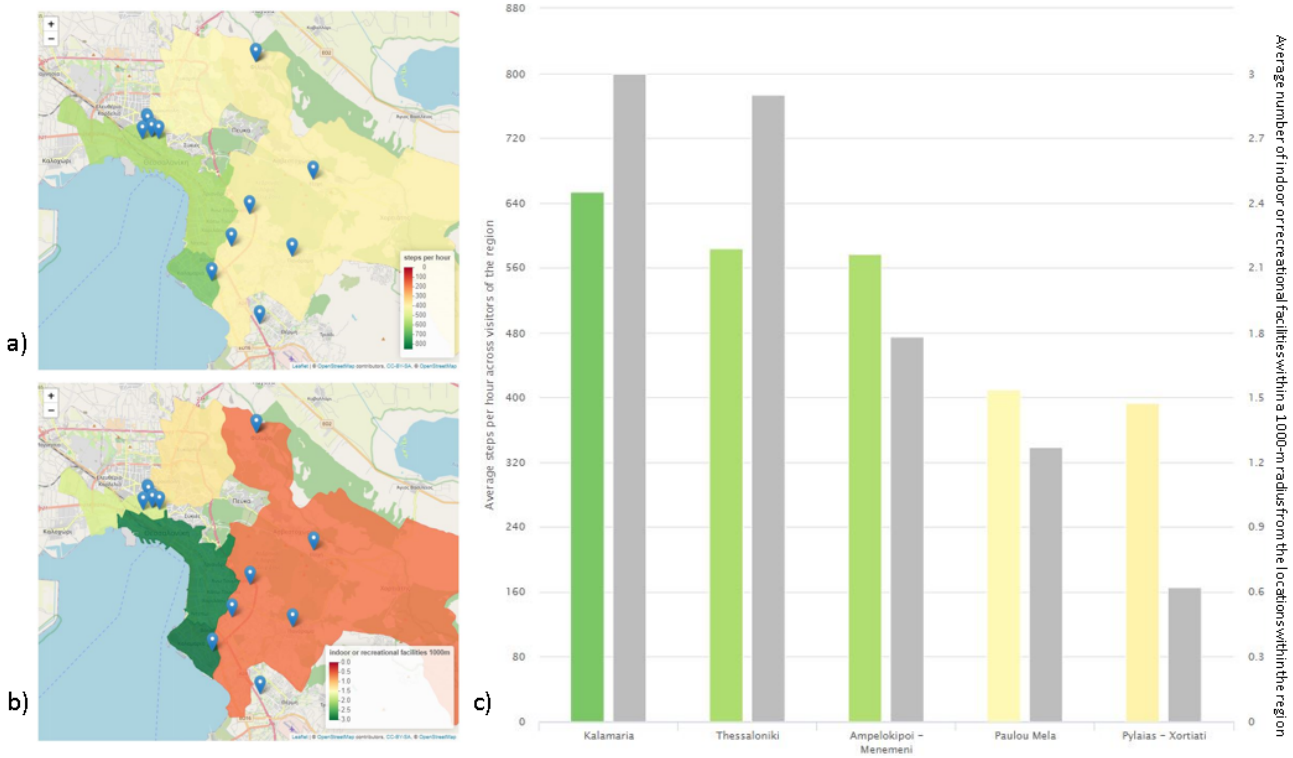


Figure 3. Regional representation of the steps per hour in the regions where schools are located (A) and the number of indoor recreational or sports facilities in those areas (B). The visual comparison between the two variables reveals an association between them (C). In (C), the color of the second indicator is set to gray in order to avoid any confusion to the user.



Associations Between the Environment and Behaviors Close to Schools

The analysis performed here was based on the data collected in Thessaloniki and Stockholm, where the majority of schools are located. In order to increase the reliability of the results, only the geohashes visited by more than 5 unique users or visited more than 5 times were analyzed. Figure 4 provides a visual representation of the behaviors identified close to schools and the respective environment in the city of Thessaloniki. Through the OPdashboard, the user can select a subset of schools in order to focus the analysis on a specific population.

In total, 12 behavioral indicators were analyzed. Those behaviors were categorized into two groups—the first one reflects the activity (eg, steps/minute or % of sedentary behavior) while the second reflects the visits to a specific place (eg, % of visits to fast food places). For the first group of behaviors, the associations were computed based on environmental characteristics either reflecting the availability of food stores or sports-related facilities. For the second group, the analysis was focused on the extraction of associations between the behavior and the density of the respective places (eg, % of fast food or takeaway visits with the density of fast food or takeaway shops in the geohash).

The OPdashboard offers automated analysis for Stockholm and Thessaloniki. The results of the Pearson correlation analysis are provided in Table 4. The results for Thessaloniki reveal the existence of statistically significant correlations between physical activity levels and the availability of food stores and sports facilities close to schools—that the higher the density of these shops or facilities in an area, the greater the chances a child will visit them (Table 4). In other words, in Thessaloniki, the use of resources in the areas close to schools was high. In order to explain this difference, additional confounders must be taken into account, such as the total population in these areas, the age of the children, and psychological factors [16].

Figure 4. An area of Thessaloniki where 4 schools are located (blue pins). The blue circles (1000-meter radius) define the regions under analysis close to schools. In the right image, the steps per hour for each geohash are depicted, whereas the density of food-related places for the same region is depicted on the left. Visual observation of the images revealed that there should exist a correlation between the density of food-related places close to schools and the steps per hour performed by the children in those areas. Source: OpenStreetMap [41].

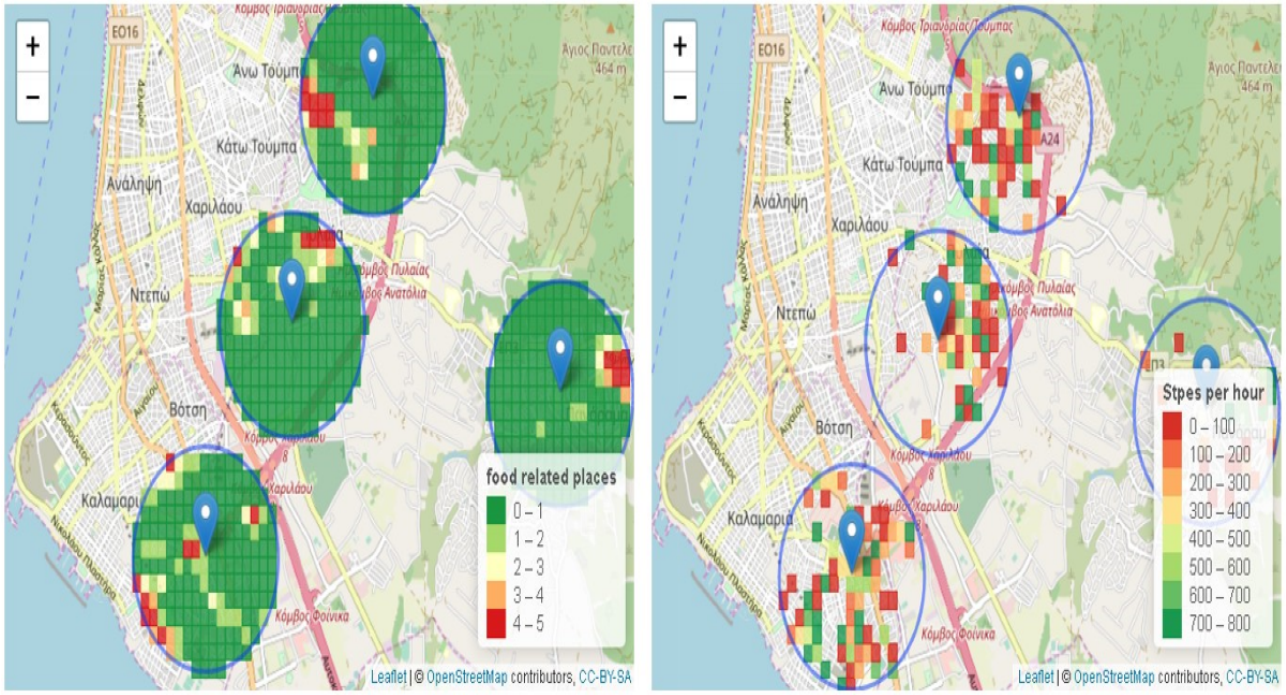


Table 4. Correlation analysis for each behavior and the most relevant environmental characteristic within a 1-km radius from a school for students in Thessaloniki (24 schools). "Food" and "Sports" are computed as the summation of all relevant environmental characteristics reflecting the availability of food places or sports-related facilities. "Visitors" indicate the aggregation of behaviors from unique children visiting a geohash, while "visits" indicate aggregation from all the visits to a geohash even when performed by the same children. For the areas close to schools, 281 geohashes with more than 5 unique visitors and 523 geohashes with more than 5 visits were analyzed.

Behavior and environment	Thessaloniki	
	Correlation (CI)	*P* value
Steps/hour (visitors)		
Food	0.123 (0.006-0.237)	.04
Sports	0.128 (0.011-0.241)	.03
Steps/hour (visits)		
Food	0.129 (0.043-0.212)	.003
Sports	0.141 (0.056-0.224)	.001
% visits to		
Food places	0.221 (0.138-0.301)	<.001
Fast food or takeaway places	0.302 (0.222-0.378)	<.001
Supermarkets or grocery stores	0.209 (0.211-0.368)	<.001
Athletics or sports facilities	0.325 (0.246-0.400)	<.001
Indoor recreation facilities	0.305 (0.225-0.381)	<.001
Public parks	0.248 (0.166-0.327)	<.001

Comparison of Children's Activity in Time

For the period before school closures (January 8 to March 10, 2020), 1802 geohashes were analyzed where at least one visit by a child was performed and a behavior was detected (Figure 5). On the other hand, in the period between March 11-31, 2020, 427 geohashes were analyzed. This expected decrease in the number of geohashes is attributed to the restrictions on population mobility as the result of the COVID-19 pandemic. A decrease in the average steps per hour was detected for the visitors of 6 municipalities within the Thessaloniki metropolitan area, while on the other hand, a slight increase was found for 2 municipalities. In particular, for the municipalities of Ampelokipoi-Menemeni and Kordelio-Evosmos, a statistically significant reduction in steps per hour was observed. Upon closer analysis, one can observe that the municipalities where an increase in physical activity was observed were suburban areas with ample open spaces. This observation implies that school closures provided an opportunity for children to exercise outside. On the contrary, the highest decrease was observed in the westside regions, possibly due to the lack of open spaces.

Finally, statistically significant differences were also observed in the downtown area of the metropolitan part of Thessaloniki, covering the area close to the coast. Under normal conditions, children visit the coast to spend their free time; however, after the closing of schools and probably due to the fear of COVID-19, children ceased visiting this area, as shown in Table 5.

Figure 5. Effect of school closure due to COVID-19 on children's physical activity levels. The steps per hour for the period between January 8, 2020 (opening of schools after the Christmas holidays) and March 10, 2020 (school closure) are depicted on the left, while on the right, the steps per hour for the period between March 11 and 31 (end of data collection) is depicted. The black lines define the boundaries of each municipality in the city of Thessaloniki. Source: OpenStreetMap [41].

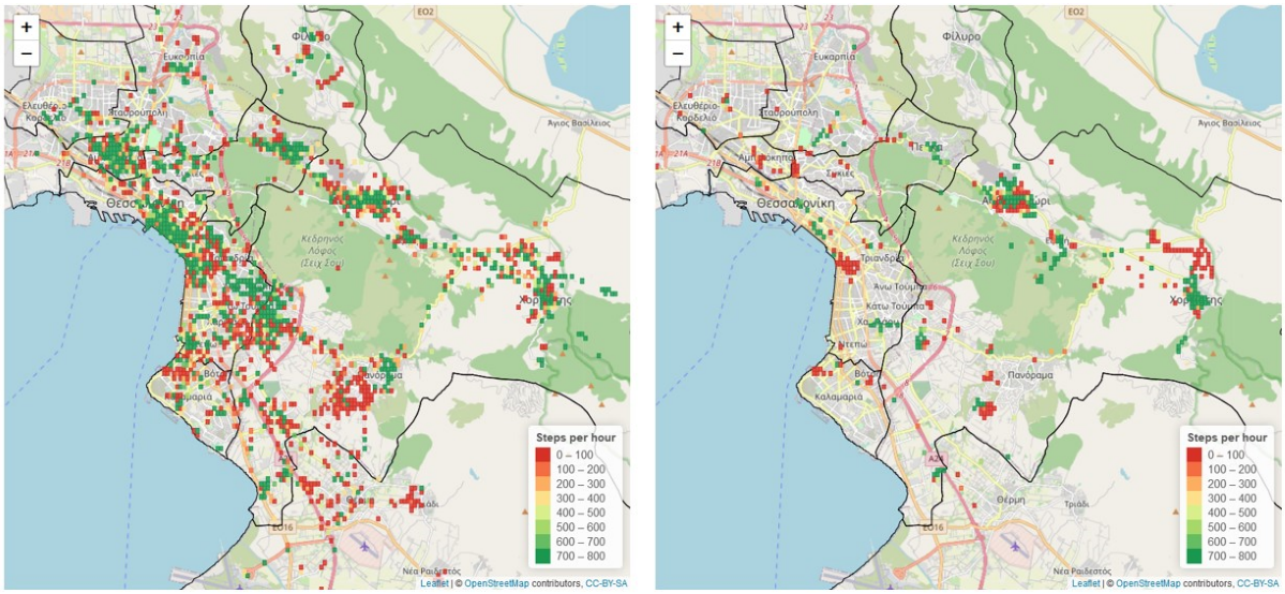


Table 5. Effect of school closures on the aggregated number of steps per hour in Thessaloniki as a measure to restrain the COVID-19 pandemic.

Municipality	Before		After		*P* value
	Geohashes, n	Average steps per hour	Geohashes, n	Average steps per hour	
Thermis	132	163.09	7	206.07	.74
Pylaia-Chortiatis	793	379.32	275	401.90	.37
Kalamaria	81	304.20	15	160.85	.09
Thessaloniki	421	468.90	60	358.62	.03
Ampelokipoi-Menemeni	87	561.37	14	46.45	<.001
Neapoli-Sykies	138	484.42	24	388.25	.22
Kordelio-Evosmos	45	516.75	11	226.35	.02
Paulou Mela	105	434.89	21	415.46	.81
Total	1802	404.54	427	367.86	.06

Evaluation

The evaluation of the first version of the OPdashboard regarding its usability and prospect of use, mainly focusing on data exploration, was performed in a focus group that took place in Katowice, Poland, on November 14, 2019, as a parallel session of the ongoing meeting of the European Children Obesity Group. During the focus group, a live demonstration of the OPdashboard was performed, and no technical issues arose.

The focus group participants mentioned the existence of many applications, software, and geographical information systems (GIS) that take data from national authorities or Eurostat and present them in map views and dashboards; however, it was noted that there was nothing available with this type of behavioral data from children at the local level and in such detail. The integration of these data with nationally collected statistics was considered by the participants to be a major strength of the OPdashboard.

In addition, during the focus group, it was mentioned that the OPdashboard potentially offers a very powerful interface with a dashboard and maps, providing insightful information on local data. This visual tool could be very useful to local health authorities to explore localized childhood behaviors and big effects. The OPdashboard therefore offers a contribution to science and public health (and clinical management) as well as the collection and use of big data in finding solutions to real-life problems.

Furthermore, regarding the analysis of the available data, it was highlighted that the analysis must focus on local use and avoid comparisons across countries. The main reason mentioned was a lack of standardization in the classifications of environmental parameters when using data from various countries to look at relationships. A large number of confounders must also be considered in order to extract causal relations between obesogenic behaviors and the environment.

One additional aspect that may affect the analysis is the estimation of the lack of standardization for the characterization of the environment. In the BigO system, environmental parameters such as points of interest use the data from GIS and online data sources, such as Google Maps, Foursquare, and OpenStreetMap. However, there are big differences in this terminology between countries. For example, fast food stores offer different types of foods in Greece (healthy foods can also be available in Greek fast food stores) and Sweden. If international comparisons are made, the analysis that is made through the OPdashboard must take into account the heterogeneity of classifications in environmental variables.

The OPdashboard can be used for the evaluation of interventions, through analysis made longitudinally with time, or for the design of health policies through the application of predictive models. In this respect, during the focus group, it was mentioned that the users who would benefit most from the OPdashboard will be local authorities, central governments, educational policy makers, and researchers (eg, epidemiologists) since these user groups play an important role in the development of effective interventions or actions to address local childhood obesity–related issues. Finally, engaged citizens may also find the system valuable due to high interest in exploring data from their region and country.

It was identified that confidentiality and privacy of data may be the main barrier against the adoption of the OPdashboard, and BigO in general, by health authorities.

In conclusion, it was underlined that BigO, and the OPdashboard in particular, can be very helpful to public health authorities for planning interventions and observing their effect in the local population.

Discussion

Principal Findings

The increased prevalence of obesity and overweight in children in European countries has been identified as a major public health issue that must be addressed. Child and adolescent obesity can occur into adulthood and is associated with increased morbidity and mortality. Modern technologies and their widespread adoption by children and adolescents can facilitate a better understanding of children's obesogenic behavior, particularly those that increase the likelihood of developing obesity early in life. In addition, the characteristics of the environment in which children move, eat, and live can promote obesogenic behaviors, and these must be considered in order to design effective localized interventions at the population level.

In this paper, we presented the OPdashboard, which is an interactive tool that may be used for the extraction of associations between children's obesogenic behaviors and the local environment using big data. The OPdashboard is freely available and the beta version can be accessed online [30]. The main functionalities of the OPdashboard include data exploration, extraction of associations, evaluation of interventions, and prediction of a behavior after an intervention. The OPdashboard was implemented as a web interface for potential users like (1) local health authorities and policy makers who develop and deliver actions to change obesogenic factors and reduce obesity, (2) researchers focusing on understanding obesity and the factors that influence it (eg, data scientists, epidemiologists), and (3) public and local educational and social services. In addition, citizens who are interested in exploring the data of their region and/or country can also benefit from the provided tools. The OPdashboard was implemented in the context of the BigO program [23], which aims to tackle the problem of childhood obesity.

Over 3700 children from more than 33 schools and 2 clinics in 5 European cities used a custom-made mHealth app, through which accelerometry and geolocation data were recorded. The big data collected led to the extraction of population behaviors in specific geographical areas while online databases were used for the collection of data related to the characteristics of the environment.

The web interface of the OPdashboard was implemented in R while RESTful web services were developed in order to access data stored in the MongoDB database. The data exploration functionality was evaluated by a focus group, where experts in the field of public health, statistical epidemiology, and clinical endocrinology participated, highlighting the usefulness of such a tool for the understanding of obesity in association with the environment.

Our findings indicated that the greater the density of food- or sports-related places in the areas proximal to schools, the greater the probability that children will visit them. These results highlight the effect of the local environmental factors on children's behavior, as previously documented in several studies [15,16].

The outbreak of the COVID-19 pandemic triggered an assessment of children's behavioral change as a result of the health-related policies implemented. In this respect, the effect of school closure in Greece on children's physical activity was studied. It was found that school closure led to an expected decrease in children's physical activity in the city of Thessaloniki. However, the decrease was statistically significant in urban areas, while in the suburbs, the decrease was not so apparent. On the contrary, in some municipalities, an increase in children's physical activity levels was observed. This finding might be the result of a different mix of underlying socioeconomic, cultural, and environmental factors related to obesogenic behaviors among children that was not addressed in the design of this study.

A recent systematic review on the use of telehealth for the treatment of childhood obesity was shown to be promising, particularly for reaching rural and less accessible patients, and carefully designed mHealth interventions have the potential for improving this reach, given the increasing popularity of mobile devices [42]. However, the prevention of obesity, or nutritional disorders in general, using mHealth apps is still under development, with some studies evaluating the feasibility of using such apps for the prevention of nutritional disorders using food intake metrics, activity level metrics, and questionnaires [43]. The impact of preventive interventions through advanced mHealth systems is expected to increase due to a lack of large-scale databases, a gap that our OPdashboard system tries

to close. Furthermore, given that pediatricians and other health professionals involved in the management of childhood obesity appear inclined to incorporate mHealth systems into their practice [44], we expect the OPdashboard to be adopted by the medical community as is the case with the educational community.

The unique big data that are available through the OPdashboard can lead to the implementation of models that are able to predict population behavior, allowing for the design of localized interventions. The OPdashboard can be considered as a tool that will increase our understanding of the underlying factors in childhood obesity and the design of regional interventions both for prevention and treatment.

Limitations

Limitations include problems regarding data sparseness such as missing data; lack of continuous monitoring at the individual level; and variations in the accuracy of the estimates, which may lead to bias or inaccurate correlation estimates. Overcoming these limitations is in progress.

There is also an underestimation of daily physical activity as the dashboard relies on smartphone-based accelerometry measurements, and smartphones are not carried continuously by children. Sports activities are usually missed. In some countries such as Greece, students may not be allowed to use their smartphones in school, further limiting the collection of accelerometry data during the day.

Furthermore, to ensure completeness of the analysis, additional confounders affecting children's behaviors must be studied, which were not considered here. These cofounders include, among others, psychological, socioeconomical, or cultural factors. Additionally, the age, gender, and BMI z-score of the children was not considered in the analysis. An analysis of the type of meals based on photos taken by the children using the myBigO app could provide more insights on children's obesogenic behaviors; however, this was out of the scope of this study.

Finally, another limitation of the study was the high administrative load due to the exclusive use of paper-based parental consent for children's participation, which resulted in low participant numbers in some schools. Following the decision of the ethics review boards in Greece and Sweden, electronic consent via the mHealth app, which could facilitate user registration, was deemed inappropriate due to the young age of the participants and the sensitive nature of the collected data.

Acknowledgments

The authors would like to thank the participating researchers of the International Hellenic University, the teachers and headmasters of the participating schools (Internationella Engelska Gymnasiet, Ellinogermaniki Agogi Scholi Panagea Savva SA, Ekpaideftiria N Mpakogianni, Ekpaideftiria Vassiliadi, and 23 public schools in Thessaloniki) as well as the researchers at the Biomedical Research Foundation, Academy of Athens, for their hard work in recruiting users and assisting with data collection. In addition, we acknowledge the pediatric obesity and public health experts from Temple Street Children's Hospital in Dublin and Wageningen University for their valuable feedback during the design stages of the BigO project. Finally, the authors would like to thank the BigO consortium for their support and for addressing technical issues.

The work leading to these results was part of the European Union (EU) H2020 project, *BigO: Big Data Against Childhood Obesity* (grant 727688). This project was part of the European Community's Health, Demographic Change and Well-being Programme of EU H2020.

Conflicts of Interest

None declared.

References

1. Damasceno A. Chapter 5 Noncommunicable Disease. In: Stewart S, Sliwa K, Mocumbi A, Damasceno A, Ntsekhe M, editors. The Heart of Africa: Clinical Profile of an Evolving Burden of Heart Disease in Africa. Hoboken, NJ: John Wiley & Sons, Ltd; 2016.
2. Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, et al. Body-mass index and mortality among 1.46 million white adults. N Engl J Med 2010 Dec 02;363(23):2211-2219 [FREE Full text] [doi: 10.1056/NEJMoa1000367] [Medline: 21121834]
3. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. The Lancet 2011 Aug 27;378(9793):815-825. [doi: 10.1016/S0140-6736(11)60814-3] [Medline: 21872750]
4. Report of the commission on ending childhood obesity. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/204176/9789241510066_eng.pdf?sequence=1 [accessed 2021-06-17]
5. Cuschieri S, Grech S. COVID-19: a one-way ticket to a global childhood obesity crisis? J Diabetes Metab Disord 2020 Nov 06:1-4 [FREE Full text] [doi: 10.1007/s40200-020-00682-2] [Medline: 33173756]

6. Lissner L, Wijnhoven TMA, Mehlig K, Sjöberg A, Kunesova M, Yngve A, et al. Socioeconomic inequalities in childhood overweight: heterogeneity across five countries in the WHO European Childhood Obesity Surveillance Initiative (COSI-2008). Int J Obes (Lond) 2016 May;40(5):796-802 [FREE Full text] [doi: 10.1038/ijo.2016.12] [Medline: 27136760]
7. Wijnhoven TMA, van Raaij JMA, Spinelli A, Rito AI, Hovengen R, Kunesova M, et al. WHO European Childhood Obesity Surveillance Initiative 2008: weight, height and body mass index in 6-9-year-old children. Pediatr Obes 2013 Apr 21;8(2):79-97. [doi: 10.1111/j.2047-6310.2012.00090.x] [Medline: 23001989]
8. Lobstein T, Jackson-Leach R, Moodie ML, Hall KD, Gortmaker SL, Swinburn BA, et al. Child and adolescent obesity: part of a bigger picture. The Lancet 2015 Jun 20;385(9986):2510-2520 [FREE Full text] [doi: 10.1016/S0140-6736(14)61746-3] [Medline: 25703114]
9. Juonala M, Magnussen CG, Berenson GS, Venn A, Burns TL, Sabin MA, et al. Childhood adiposity, adult adiposity, and cardiovascular risk factors. N Engl J Med 2011 Nov 17;365(20):1876-1885. [doi: 10.1056/NEJMoa1010112] [Medline: 22087679]
10. Twig G, Afek A, Shamiss A, Derazne E, Landau Rabbi M, Tzur D, et al. Adolescence BMI and trends in adulthood mortality: a study of 2.16 million adolescents. J Clin Endocrinol Metab 2014 Jun;99(6):2095-2103 [FREE Full text] [doi: 10.1210/jc.2014-1213] [Medline: 24601695]
11. Wang W, Yu Y, Wilson R, Bleich S, Cheskin L, Weston C, et al. Childhood Obesity Prevention Programs: Comparative Effectiveness Review and Meta-Analysis. Report No.: 13-EHC081-EF. Rockville, MD: Agency for Healthcare Research and Quality (US); 2013.
12. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. JAMA 2012 Feb 01;307(5):483-490 [FREE Full text] [doi: 10.1001/jama.2012.40] [Medline: 22253364]
13. Petrauskienė A, Žaltauskė V, Albavičiūtė E. Family socioeconomic status and nutrition habits of 7-8 year old children: cross-sectional Lithuanian COSI study. Ital J Pediatr 2015 Apr 23;41:34 [FREE Full text] [doi: 10.1186/s13052-015-0139-1] [Medline: 25903745]
14. Ludwig J, Sanbonmatsu L, Gennetian L, Adam E, Duncan GJ, Katz LF, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. N Engl J Med 2011 Oct 20;365(16):1509-1519 [FREE Full text] [doi: 10.1056/NEJMsa1103216] [Medline: 22010917]
15. Davis B, Carpenter C. Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity. Am J Public Health 2009 Mar;99(3):505-510. [doi: 10.2105/ajph.2008.137638]
16. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics 2006 Feb;117(2):417-424. [doi: 10.1542/peds.2005-0058] [Medline: 16452361]
17. Folkvord F, Anschütz DJ, Boyland E, Kelly B, Buijzen M. Food advertising and eating behaviour in children. Appetite 2016 Dec;107:681. [doi: 10.1016/j.appet.2016.08.038]
18. Katz RL, Felix M, Gubernick M. Technology and adolescents: Perspectives on the things to come. Educ Inf Technol 2013 May 4;19(4):863-886. [doi: 10.1007/s10639-013-9258-8]
19. Zandian M, Ioakimidis I, Bergström J, Brodin U, Bergh C, Leon M, et al. Children eat their school lunch too quickly: an exploratory study of the effect on food intake. BMC Public Health 2012 May 14;12(1):351 [FREE Full text] [doi: 10.1186/1471-2458-12-351] [Medline: 22583917]
20. Diou C, Sarafis I, Papapanagiotou V, Ioakimidis I, Delopoulos A. A methodology for obtaining objective measurements of population obesogenic behaviors in relation to the environment. SJI 2019 Dec 10;35(4):677-690. [doi: 10.3233/sji-190537]
21. Chaix B. Mobile Sensing in Environmental Health and Neighborhood Research. Annu Rev Public Health 2018 Apr 01;39:367-384. [doi: 10.1146/annurev-publhealth-040617-013731] [Medline: 29608869]
22. Chaix B, Duncan D, Vallée J, Vernez-Moudon A, Benmarhnia T, Kestens Y. The "Residential" Effect Fallacy in Neighborhood and Health Studies: Formal Definition, Empirical Identification, and Correction. Epidemiology 2017 Nov;28(6):789-797. [doi: 10.1097/EDE.0000000000000726] [Medline: 28767516]
23. BigO. URL: https://bigoprogram.eu/ [accessed 2021-06-07]
24. Pépin JL, Bruno RM, Yang R, Vercamer V, Jouhaud P, Escourrou P, et al. Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis. J Med Internet Res 2020 Jun 19;22(6):e19787 [FREE Full text] [doi: 10.2196/19787] [Medline: 32501803]
25. Socientize Consortium. Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research. European Commission. URL: https://digital-strategy.ec.europa.eu/en/library/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research [accessed 2021-06-04]
26. Maramis C, Ioakimidis I, Kilintzis V, Stefanopoulos L, Lekka E, Papapanagiotou V, et al. Developing a Novel Citizen-Scientist Smartphone App for Collecting BehavioralAffective Data from Children Populations. In: O'Hare G, O'Grady M, O'Donoghue J, Henn P, editors. Wireless Mobile Communication and Healthcare. MobiHealth 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 320. Cham: Springer International Publishing; 2020.
27. Geohash. URL: http://geohash.org [accessed 2021-06-07]

28. O'Donnell S, Doyle G, O'Malley G, Browne S, O'Connor J, Mars M, et al. Establishing consensus on key public health indicators for the monitoring and evaluating childhood obesity interventions: a Delphi panel study. BMC Public Health 2020 Nov 17;20(1):1733 [FREE Full text] [doi: 10.1186/s12889-020-09814-y] [Medline: 33203390]
29. Papapanagiotou V, Sarafis I, Diou C, Ioakimidis I, Charmandari E, Delopoulos A. Collecting big behavioral data for measuring behavior against obesity. 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24; Montreal, Canada p. 5296-5299 URL: http://arxiv.org/abs/2005.04928 [doi: 10.1109/embc44109.2020.9175313]
30. OPdashboard. URL: http://bigo.med.auth.gr:3838 [accessed 2021-06-07]
31. REST API Tutorial. RESTful. URL: https://restfulapi.net/ [accessed 2021-06-07]
32. Eclipse Jersey. Eclipse Foundation. URL: https://jersey.github.io/documentation/latest/index.html [accessed 2021-06-07]
33. MongoDB Java Drivers. MongoDB. URL: https://mongodb.github.io/mongo-java-driver/ [accessed 2021-06-07]
34. JWT Web Tokens. JWT. URL: https://jwt.io/ [accessed 2021-06-07]
35. Shiny from R Studio. RStudio, PBC. URL: http://shiny.rstudio.com/ [accessed 2021-06-07]
36. Day PL, Pearce J. Obesity-promoting food environments and the spatial clustering of food outlets around schools. Am J Prev Med 2011 Feb;40(2):113-121. [doi: 10.1016/j.amepre.2010.10.018] [Medline: 21238858]
37. Kumanyika SK. Environmental influences on childhood obesity: ethnic and cultural influences in context. Physiol Behav 2008 Apr 22;94(1):61-70. [doi: 10.1016/j.physbeh.2007.11.019] [Medline: 18158165]
38. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
39. Sarafis I, Diou C, Papapanagiotou V, Alagialoglou L, Delopoulos A. Inferring the Spatial Distribution of Physical Activity in Children Population from Characteristics of the Environment. 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24; Montreal, Canada URL: http://arxiv.org/abs/2005.03957 [doi: 10.1109/embc44109.2020.9176246]
40. Diou C, Sarafis I, Papapanagiotou V, Alagialoglou L, Lekka I, Filos D, et al. BigO: A public health decision support system for measuring obesogenic behaviors of children in relation to their local environment. 2020 Jul 20 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24; Montreal, Canada URL: http://arxiv.org/abs/2005.02928 [doi: 10.1109/embc44109.2020.9175361]
41. OpenStreetMap. URL: https://www.openstreetmap.org/#map=2/71.3/-96.8 [accessed 2021-06-07]
42. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile Health for Pediatric Weight Management: Systematic Scoping Review. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
43. Maramis C, Moulos I, Ioakimidis I, Papapanagiotou V, Langlet B, Lekka I, et al. A smartphone application for semi-controlled collection of objective eating behavior data from multiple subjects. Comput Methods Programs Biomed 2020 Oct;194:105485. [doi: 10.1016/j.cmpb.2020.105485] [Medline: 32464588]
44. SanGiovanni CB, Morella K, Roberts JR. Primary Care Providers Welcome Smartphone Apps That Assist in Pediatric Weight Management. Clin Pediatr (Phila) 2019 Jun 27;58(6):665-670. [doi: 10.1177/0009922819832642] [Medline: 30813759]

Abbreviations

GIS: geographical information system
JSON: JavaScript Object Notation
mHealth: mobile health
OPdashboard: Obesity Prevention dashboard

Edited by L Buis; submitted 05.12.20; peer-reviewed by V Vercamer, D Berrigan, J Alvarez Pitti; comments to author 12.01.21; revised version received 02.02.21; accepted 19.03.21; published 09.07.21.

Please cite as:
Filos D, Lekka I, Kilintzis V, Stefanopoulos L, Karavidopoulou Y, Maramis C, Diou C, Sarafis I, Papapanagiotou V, Alagialoglou L, Ioakeimidis I, Hassapidou M, Charmandari E, Heimeier R, O'Malley G, O'Donnell S, Doyle G, Delopoulos A, Maglaveras N
Exploring Associations Between Children's Obesogenic Behaviors and the Local Environment Using Big Data: Development and Evaluation of the Obesity Prevention Dashboard
JMIR Mhealth Uhealth 2021;9(7):e26290
URL: https://mhealth.jmir.org/2021/7/e26290
doi:10.2196/26290
PMID:34048353

Original Paper

Effect of Adherence to Smartphone App Use on the Long-term Effectiveness of Weight Loss in Developing and OECD Countries: Retrospective Cohort Study

Myeunghee Han[1], PhD; Sang Youl Rhee[2], PhD, MD

[1]Daegu Health High School, Daegu, Republic of Korea

[2]Department of Endocrinology and Metabolism, Kyung Hee University School of Medicine, Seoul, Republic of Korea

Corresponding Author:
Sang Youl Rhee, PhD, MD
Department of Endocrinology and Metabolism
Kyung Hee University School of Medicine
23 Kyungheedae-ro, Seoul, 02447, Korea
Seoul
Republic of Korea
Phone: 82 029610274
Email: bard95@hanmail.net

Abstract

Background: Globally, 71% of deaths occur due to noncommunicable diseases (NCDs). Poor diet quality and physical activity have a significant impact on NCDs. At present, behavior change interventions using smartphone apps have rapidly increased worldwide to prevent NCDs. However, most previous studies on the use and effectiveness of apps have been conducted in Organization for Economic Co-operation and Development (OECD) countries. As such, relevant research in low-income countries is scarce.

Objective: This retrospective cohort study aims to investigate the characteristics of adherence to the use of the Noom app. We also aim to compare the effects of using the app on body weight changes over time according to adherence to the use of the app between users in low-income and OECD countries. In addition, the differences in weight loss are compared among users who use the free and paid versions of the app.

Methods: A secondary data analysis was conducted using repeated measures. The data were collected from users in low-income countries (n=312) and OECD countries (n=8041) who used the app for 12 months. The app provided programs for the self-monitoring of physical activity, dietary intake, and body weight. Descriptive statistics, independent two-tailed *t* tests, chi-square tests, and linear mixed models were used for the analysis.

Results: During the first 3 months of using the Noom app, users from OECD countries entered data into the app more frequently; however, users in low-income countries entered data more frequently from 3 months to 12 months. Users in OECD countries consumed significantly more calories than those in low-income countries for 12 months. The body weight of all users significantly decreased over time (−1.8 kg; $P<.001$); however, no statistically significant differences in the change in body weight for 12 months were observed between users from low-income and OECD countries (β=−.2; $P=.19$). The users who frequently monitored their lunch (β=−.1; $P<.001$), dinner (β=−.1; $P<.001$), body weight (β=−.1; $P<.001$), evening snack (β=−.1; $P<.001$), and exercise (β=−.03; $P<.001$) exhibited significant weight loss over time. We found no significant differences in the body weight changes between users who used the free and paid versions of the app (β=−.2; $P=.19$).

Conclusions: This study found that using the app has a significant effect on weight loss regardless of users' country of residence. The results of this study suggest that the frequency of monitoring health-related behaviors by entering data into the app plays a pivotal role in losing weight. In conclusion, regardless of where users live and what versions of the app they use, it is important to monitor health-related behaviors by frequently entering data into the app to efficiently lose weight.

(JMIR Mhealth Uhealth 2021;9(7):e13496) doi:10.2196/13496

KEYWORDS

low-income countries; Organization for Economic Co-operation and Development; body weight; mobile app self-management; diet; exercise; mobile phone

Introduction

Background

Annually, approximately 41 million people (or 71% of all deaths globally) die of noncommunicable diseases (NCDs), such as cardiovascular diseases, cancers, and diabetes [1]. Accordingly, poor diet quality and physical inactivity, which are the key factors for the prevalence of NCDs and mortality worldwide, are considered priority areas for global action [2]. Improving physical activity (PA) and dietary intake have been emphasized to prevent NCDs [3].

Given the global scale of NCDs, effective preventative interventions that can reach a wide range of populations at low costs are urgently needed [4]. It has been recognized that technology can support health improvements worldwide [1]. Specifically, because of the lowering prices of smartphones and the easy access they provide, mobile technology is expected to play a particularly important role in improving health-related behaviors in low-, middle-, and high-income countries [5]. Currently, smartphones are arguably the most prosperous and expeditiously adopted modern technology in the world. In low-income countries, access to smartphones increased from 4% to 94% from 2000 to 2015 [6].

Objectives

The growth in mobile technologies has stimulated the growth of smartphone health and fitness apps [7]. The apps that target health promotion have become a central part of people's lives and have demonstrated large increases in use [8]. According to previous studies, smartphone health apps have a positive impact on the improvement of health-related behaviors and outcomes, such as dietary intake, PA, and weight loss [9].

However, despite the growing use of apps that target health-related behavior change, the long-term effects of such apps on targeted health behaviors and outcomes, such as diet, PA, and weight loss, particularly among users in low-income countries, remains unclear [10,11]. Most previous studies were conducted in wealthy countries with highly developed technologies. To fill this gap in the literature, this study was conducted to compare the characteristics of adherence with using a smartphone app (Noom) and the effectiveness of the apps between low-income and Organization for Economic Co-operation and Development (OECD) countries.

The hypotheses tested in this study were as follows:

- H1: There would be differences in adherence to the use of self-monitoring data entry of diet, PA, and body weight between users from low-income countries and OECD countries.
- H2: There would be differences in body weight changes over time between users from low-income and OECD countries.
- H3: There would be differences in body weight changes between users of the paid and free versions of the app.

Methods

Study Design

This study was a retrospective cohort study that aimed to compare the use and effectiveness of a smartphone health app on changes in body weight over time between low-income and OECD countries.

Setting

Low-Income Countries

The list of low-income countries was derived from a report by the International Statistical Institute. The gross national income, derived from the World Bank County classification, was used as a measure of a country's income. Countries with gross national income per capita slightly over US $12,476 were considered as low-income countries [12]. In this study, we focused on data from 31 low-income countries (Multimedia Appendix 1).

OECD Countries

The OECD is an intergovernmental economic organization with 36 member countries [13]. Most OECD countries are high-income economies with a high human development index and are regarded as high-income countries [14]. The data of users from 32 OECD countries were included in this study (Multimedia Appendix 2).

Participants

The data were provided by the Noom Coach (Noom Inc) app company. Individuals who used the Noom Coach app for 12 months between October 2012 and April 2014 and provided relevant data (demographic characteristics, exercise, dietary intake, and weight) were included. The participants used the exercise data entry function, dietary data entry function, or weight data entry function. As the age of 42 years is the default value in the age tab of the app, we excluded all users who indicated 42 years as their age, assuming that all these users did not correctly indicate their age when they started using the app. Accordingly, from 48,095 cases in the original data set, 4026 cases were removed. From the remaining 44,069 users, we selected 12,173 users who used the app for 12 months. On deletion of users who did not enter all necessary data (n=3638), the final data set contained a total of 8353 users.

Intervention: Noom Coach

Noom Coach is a smartphone app for weight loss that tracks dietary intake, PA, and body weight. Created in 2012, this app is now available in 5 languages (English, Korean, Japanese, German, and Spanish) from the Google Play store; the iOS version is also available. With more than 10 million downloads worldwide, this app has been consistently ranked as the most effective weight loss app [15,16].

When app users log in for the first time, they are asked to enter the expected body weight and record their present body weight and height. During the period of using the Noom Coach app, users are requested to record their daily dietary intake and the number of footsteps as their PA. On the basis of the data entered by users, the app reports the trends in body weight changes, calories, and nutritional summaries. To support the achievement of the desired body weight, the app provides tailored feedback, including types of exercise. The app is available in 2 versions: free and paid. The free version includes functions such as food logging, weight tracking, and helpful tips. The paid version offers more services, including supportive advice from a specialist, new recipes of dietary intake, workout guides, tracking progress, and one-to-one coaching.

Statistical Methods

All statistical analyses were performed using SPSS (version 24.0; IBM Corporation). Appropriate descriptive statistic analyses were conducted on baseline variables, such as age, sex, body weight, BMI, and the frequency of data entry on each section of diet and exercise. To analyze the differences between users of the app in OECD countries and low-income countries, independent two-tailed sample *t* tests were conducted. Moreover, the progression of body weight over time was described graphically. A linear mixed model (LMM) was used to evaluate the differences in body weight changes over time between the low-income and OECD countries. Statistical significance was determined at $P<.05$ (2-sided). Before conducting the analysis, all assumptions were checked and met. The dispersion of the outcome variable of body weight was checked before conducting the LMM. All assumptions were met to conduct an LMM.

LMM with random intercepts was used to evaluate the effects of time and the effects of gender, age, group (OECD countries vs low-income countries), frequency of exercise data entry, frequency of breakfast data entry, frequency of morning snack data entry, frequency of lunch data entry, frequency of afternoon snack data entry, frequency of dinner data entry, and frequency of evening snack data entry on the weight changes over time. From the unconditional model, the value of intraclass correlations was 0.9, which confirmed the use of the LMM for further analysis. From the results of the conditional model, the random effects model was selected for the LMM. Individual univariate LMMs were conducted with each independent variable to select all significant variables. All significant variables were entered into the LMM model, and then backward eliminations were conducted until the minimum values of Akaike information criterion and Bayesian information criterion were reached, indicating the best model that predicted the body weight changes over time.

Variables

In this study, adherence to app use was defined as the frequency of exercise, diet (breakfast, breakfast snacks, lunch, afternoon snacks, dinner, and dinner snacks), or the frequency of weight data entry [11]. The outcome variable was body weight (kg) changes from baseline to 12 months. If there were no body weight indications at baseline or at 3, 6, 9, and 12 months, the values were replaced with the average weight calculated by averaging the body weight before and after 7 days at each time point. Age and sex were used as the self-reported baseline values. BMI (kg/m^2) was calculated using weight in kilograms divided by height in meters squared. BMI was divided into the following 4 categories: underweight (≤18.5 kg), normal (18.5-24.9 kg), overweight (25-29.9 kg), and obese (≥30 kg) [17]. The frequencies of data entry for exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack were the sum of the number of days with values every 3 months. The average calories for exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack were calculated by averaging the calories consumed before and after 7 days at each time point.

Results

Characteristics of Participants

The baseline characteristics of the participants (N=8343) are summarized in Table 1. Of the 8343 users, 8041 (96.38%) were from OECD countries and 312 (3.88%) were from low-income countries. The mean ages of users from low-income countries and OECD countries were 32.43 years (SD 9.5; range 18-66) and 36.1 years (SD 11.7; range 13-76), respectively. Most users in each group were female (6024/8343, 72.2%). The mean values of body weight of users from low-income and OECD countries were 74.6 kg and 82.2 kg at baseline, respectively. Most users in low-income countries (208/312, 66.7%) and OECD countries (5468/8041, 68%) were obese.

Table 1. Demographic characteristics at baseline (N=8353).

Characteristics	Organization for Economic Co-operation and Development countries (n=8041)	Low-income countries (n=312)
Age (years), mean (SD; range)	36.1 (11.7; 13.0-76.0)	32.4 (9.5; 18.0-66.0)
Gender, n (%)		
Male	2017 (25.08)	112 (35.9)
Female	6024 (74.92)	200 (64.1)
Weight (kg), mean (SD; range)	82.2 (21.2; 39.0-188.7)	74.6 (17.1; 49.0-158.8)
BMI (kg/m^2), n (%)		
Underweight	63 (0.78)	2 (0.64)
Normal weight	1400 (17.41)	60 (19.23)
Overweight	1110 (13.80)	42 (13.46)
Obese	5468 (68)	208 (66.67)
Frequency of exercise entry, mean (SD; range)	123.6 (97.4; 0.0-360.0)	126.2 (96.6; 0.0-360.0)
Frequency of breakfast data entry, mean (SD; range)	198.5 (99.6; 0.0-360.0)	199.3 (103.8; 0.0-360.0)
Frequency of morning snack data entry, mean (SD; range)	89.0 (85.8; 0.0-360.0)	114.8 (99.7; 0.0-360.0)
Frequency of lunch data entry, mean (SD; range)	183.1 (99.0; 0.0-360.0)	179.7 (104.5; 0.0-360.0)
Frequency of afternoon snack data entry, mean (SD; range)	102.0 (84.9; 0.0-360.0)	120.4 (97.1; 0.0-360.0)
Frequency of dinner data entry, mean (SD; range)	152.5 (100.8; 0.0-360.0)	149.4 (106.3; 0.0-360.0)
Frequency of evening snack data entry, mean (SD; range)	57.5 (70.1; 0.0-360.0)	62.7 (76.0; 0.0-360.0)
Exercise, mean (SD; range)	305.8 (192.2; 0.0-1300.0)	280.4 (190.1; 0.0-1300.0)
Breakfast, mean (SD; range)	288.1 (110.3; 0.0-700.0)	293.8 (112.0; 0.0-700.0)
Morning snack, mean (SD; range)	179.7 (95.4; 0.0-1244.4)	187.4 (107.6; 0.0-1200.0)
Lunch, mean (SD; range)	413.9 (151.0; 0.0-1419.7)	226.6 (107.2; 0.0-1400.0)
Afternoon snack, mean (SD; range)	225.0 (122.0; 0.0-1500.0)	401.9 (171.2; 0.0-1108.2)
Dinner, mean (SD; range)	458.8 (173.0; 0.0-1436.7)	223.7 (125.0; 0.0-714.7)
Evening snack, mean (SD; range)	235.5 (123.7; 0.0-978.0)	157.9 (74.6; 0.0-157.9)

The average frequencies of data entry of users from OECD countries for exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack were 123.6, 198.5, 89.0, 183.1, 102.0, 152.5, and 57.5, respectively. The average frequencies of data entry of users from low-income countries for exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack were 126.2, 199.3, 114.8, 179.7, 120.4, 149.4, and 62.7, respectively. The mean calories of exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack of users from OECD countries were 305.8, 288.1, 179.7, 413.9, 225.0, 458.8, and 235.5, respectively. The mean calories of exercise, breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack of users from low-income countries were 280.4, 293.8, 187.4, 226.6, 401.9, 223.7, and 157.9, respectively. The values corresponding to each item represent the average value over a period when users use the app.

Comparison of Frequency of Self-Monitoring Data Entry Between Low-Income and OECD Countries at Each Time Point

Independent two-tailed sample *t* tests were conducted to compare the frequency of data entry and calories between users from low-income and OECD countries. There were significant differences in the frequency of data entering for breakfast (t_{8351}=−2.6; *P*=.009), morning snack (t_{8351}=3.1; *P*=.002), lunch (t_{8351}=−3.1; *P*=.002), and dinner (t_{8351}=−2.9; *P*=.003) data between baseline and 3 months. At this time point, users from OECD countries entered their data (except for those on morning snack) more frequently than those from low-income countries. Between 3 and 6 months, users from low-income countries

entered morning snack (t_{8351}=5.0; P<.001) and afternoon snack data (t_{8351}=3.6; P<.001) significantly more frequently. Users in low-income countries entered morning snack (t_{8351}=4.3; P<.001) and afternoon snack data (t_{8351}=2.8; P=.005) more times than those in OECD countries between 6 and 9 months. Between 9 and 12 months, there was a significant difference in the values for morning snack (t_{8351}=3.8; P<.001) and afternoon snack data (t_{8351}=2.3; P<.02) entering between the users from low-income and OECD countries (Table 2). These results support hypothesis 1, as there were differences in adherence to the use of self-monitoring data entry between users from low-income countries and OECD countries.

Table 2. Results of independent two-tailed *t* tests to compare the difference in frequency of data entry between Organization for Economic Co-operation and Development (n=8041) and low-income countries (n=312).

Characteristics	Time 1 (baseline to 3 months), mean (SD)		Time 2 (3-6 months), mean (SD)		Time 3 (6-9 months), mean (SD)		Time 4 (9-12 months), mean (SD)	
	OECD[a]	Low-income	OECD	Low-income	OECD	Low-income	OECD	Low-income
Frequency of exercise data entry	35.1 (23.0)	36.2 (24.5)	28.4 (23.6)	29.7 (24.3)	22.2 (22.5)	23.3 (23.0)	18.5 (21.1)	19.4 (21.1)
Frequency of breakfast data entry	51.5 (30.0)[b]	46.9 (32.1)[b]	50.9 (28.1)	51.7 (27.3)	40.0 (29.7)	40.4 (30.5)	32.9 (29.1)	33.9 (29.9)
Frequency of morning snack data entry	22.4 (22.2)[b]	26.4 (26.4)[b]	21.7 (22.3)[b]	28.2 (25.8)[b]	16.4 (20.5)[b]	21.5 (24.6)[b]	13.4 (19.2)[b]	17.8 (23.4)[b]
Frequency of lunch data entry	48.4 (29.4)[b]	43.1 (32.1)[b]	47.1 (47.5)	47.5 (28.4)	36.2 (28.9)	36.8 (30.3)	29.5 (28.2)	30.3 (29.3)
Frequency of afternoon snack data entry	27.7 (23.0)	29.2 (26.4)	25.5 (22.6)[b]	30.3 (24.9)[b]	18.8 (21.2)[b]	22.3 (24.3)[b]	15.2 (19.8)[b]	17.9 (22.6)[b]
Frequency of dinner data entry	42.8 (29.1)[b]	37.8 (30.7)[b]	38.7 (28.0)	38.8 (22.6)	29.1 (27.8)	29.3 (28.2)	23.6 (26.3)	23.8 (27.1)
Frequency of evening snack data entry	16.5 (18.1)	17.1 (20.5)	13.6 (17.5)	15.0 (19.3)	21.1 (16.8)	21.1 (18.0)	8.0 (14.9)	8.9 (16.1)

[a]OECD: Organization for Economic Co-operation and Development.

[b]Significant at the P<.05 level.

Body Weight Changes Between Users in Low-Income and OECD Countries From Baseline to 12 Months

Figure 1 shows the body weight (kg) changes of users from the low-income and OECD countries at each time point. The body weight of users from the low-income and OECD countries was 78.3 and 82.3 at baseline, respectively. From baseline to 6 months, users in both types of countries exhibited a dramatic reduction in their body weight, about 5 kg, which then slightly decreased until 12 months (Figure 1). There were no statistically significant differences in the degree of weight loss between users from the low-income and OECD countries at each time point (baseline to 3 months, 3-6 months, 6-9 months, and 9-12 months; Table 3).

Figure 1. Comparison of body weight changes over time between Organization for Economic Co-operation and Development (n=8041) and low-income countries (n=312). OECD: Organization for Economic Co-operation and Development.

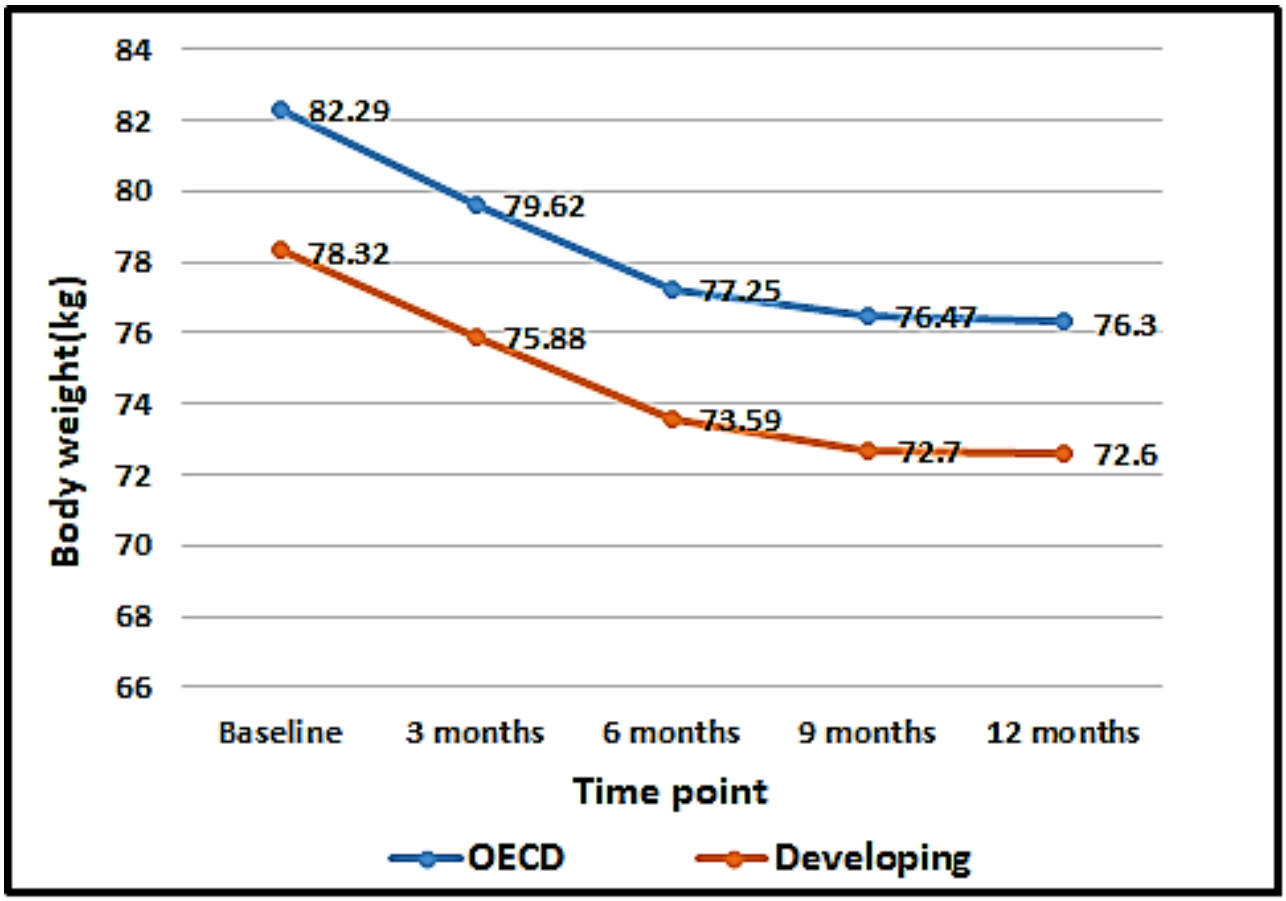


Table 3. Difference in the degree of weight loss between Organization for Economic Co-operation and Development (n=8041) and low-income (n=312) countries at each time point.

Time	Organization for Economic Co-operation and Development countries, mean (SD)	Low-income countries, mean (SD)	*t* test (*df*)	*P* value	95% CI
Baseline to 3 months	−2.7 (5.1)	−2.4 (7.7)	0.7 (8351)	.46	−0.4 to 0.8
3-6 months	−2.4 (3.8)	−2.3 (3.2)	0.4 (8351)	.67	−0.3 to 0.5
6-9 months	−0.8 (3.7)	−0.9 (2.1)	−0.5 (8351)	.63	−0.5 to 0.3
9-12 months	−0.2 (4.8)	−0.03 (2.1)	0.6 (8351)	.54	−0.4 to 0.7

Comparison of Body Weight Changes Over Time Between Users From Low-Income and OECD Countries and Between Users of the Paid and Free Versions of the App

The change in body weight over time was not significantly different between the users from low-income and OECD countries (β=−.2; *P*=.19). However, there were differences in body weight changes over time according to adherence to the app. For every increase of 1 unit in frequency of exercise (β=−.004; *P*<.001), lunch (β=−.01; *P*<.001), dinner (β=−.01; *P*<.001), evening snack (β=−.01; *P*<.001), or weight (β=−.01; *P*<.001), body weight statistically significantly decreased. The changes in body weight differed by gender, as demonstrated by the interaction between time and sex (β=.7; *P*<.001). For every increase of 1 unit in age, the body weight increased by 0.002 (*P*=.03; Table 4). There was no significant difference in body weight changes over time based on the version of the app (free version vs paid version; β=−.01; *P*=.91). On the basis of the results, the hypothesis that there would be differences in body weight change over time according to smartphone adherence between users from low-income and OECD countries and the hypothesis that there would be differences in body weight changes between users of the paid and free versions of the app had to be rejected.

Table 4. Compare the difference in body weight changes over time between low-income (n=312) and Organization for Economic Co-operation and Development (n=8041) countries.

Parameter	Estimate (SE; 95% CI)	*t* test (*df*)	*P* value
Intercept	89.8 (1.2; 87.4–92.3)	72.2 (8380.0)	<.001
Time	−1.7 (0.1; −1.9 to −1.4)	−12.8 (8524.7)	<.001
Age	0.1 (0.01; 0.04–0.1)	6.6 (8347.3)	<.001
Gender	−19.0 (0.5; −20.0 to −18.0)	−37.4 (8809.2)	<.001
OECD[a]	5.7 (1.2; 3.4–8.0)	4.8 (8346.2)	<.001
Frequency of lunch data entry	−0.04 (0.01; −0.1 to −0.03)	−7.7 (30380.0)	<.001
Frequency of dinner data entry	0.03 (0.01; 0.02–0.1)	5.2 (30493.1)	<.001
Frequency of weight data entry	−0.02 (0.003; −0.02 to −0.01)	−5.0 (31933.7)	<.001
Frequency of evening snack data entry	0.02 (0.01; 0.01–0.03)	3.2 (31694.5)	.001
Interaction between time and gender	0.7 (0.1; 0.6–0.8)	12.7 (8369.0)	<.001
Interaction between time and frequency of lunch data entry	0.01 (0.001; 0.01–0.01)	5.7 (32911.1)	<.001
Interaction between time and frequency of dinner data entry	−0.01 (0.002; −0.01 to −0.01)	−4.9 (33057.6)	<.001
Interaction between time and frequency of body weight data entry	−0.01 (0.001; −0.01 to −0.003)	−4.1 (33075.5)	<.001
Interaction between time and frequency of evening snack data entry	−0.01 (0.002; −0.01 to −0.004)	−4.4 (33007.3)	<.001
Interaction between time and frequency of exercise data entry	−0.003 (0.0004; −0.004 to −0.003)	−7.6 (33510.8)	<.001
Interaction between time and age	0.002 (0.001; −0.0001 to 0.003)	1.8 (8299.9)	.03
Interaction between time and OECD	−0.2 (0.12; −0.4 to 0.1)	−1.3 (8304.4)	.19

[a]OECD: Organization for Economic Co-operation and Development.

Comparison of Consumed Calories Between Users From the Low-Income and OECD Countries at Each Time Point

There were significant differences in calorie consumption for each category between the users from low-income and OECD countries. Between baseline and 3 months, users from OECD countries consumed significantly more calories for breakfast (t_{8351}=−2.5; *P*=.01), lunch (t_{8351}=−3.5; *P*<.001), afternoon snack (t_{8351}=−2.0; *P*=.04), dinner (t_{8351}=−6.7; *P*<.001), and evening snacks (t_{8351}=−2.2; *P*=.03). Compared with users from low-income countries, users from OECD countries also consumed significantly more calories for exercise (t_{8351}=−2.7; *P*=.008) and dinner (t_{8351}=−5.2; *P*<.001) between 3 and 6 months. There was also a statistically significant difference in calories of exercise (t_{8351}=−2.6; *P*=.01) and dinner (t_{8351}=−3.9; *P*<.001) for both groups between 6 and 9 months. Users from OECD countries consumed more calories for exercise and dinner. Between 9 months and 12 months, users from the OECD countries consumed significantly higher amounts of dinner calories than users from low-income countries (t_{8351}=−4.7; *P*<.001; Table 5).

Table 5. Results of independent two-tailed *t* tests of average consumed calories between Organization for Economic Co-operation and Development (n=8041) and low-income countries (n=312).

Characteristics	Time 2 (baseline to 3 months), mean (SD)		Time 3 (3-6 months), mean (SD)		Time 4 (6-9 months), mean (SD)		Time 5 (9-12 months), mean (SD)	
	OECD[a]	Low-Income	OECD	Low-Income	OECD	Low-Income	OECD	Low-Income
Calories of exercise	277.0 (189.1)	259.8 (192.6)	286.3 (217.9)[b]	252.9 (182.8)[b]	277.6 (232.3)[b]	243.4 (202.7)[b]	270.4 (241.1)	231.6 (214.9)
Calories of breakfast data	263.6 (130.8)[b]	245.0 (145.1)[b]	280.8 (125.7)	280.0 (126.7)	276.5 (134.2)	280.3 (141.1)	270.1 (142.3)	281.8 (151.9)
Calories of morning snack	152.1 (149.6)	149.6 (122.1)	169.1 (120.8)	169.4 (105.0)	158.5 (126.6)	171.6 (132.7)	149.6 (136.4)	162.4 (132.5)
Calories of lunch	366.8 (175.9)[b]	331.0 (192.1)[b]	402.0 (163.8)	388.0 (170.4)	392.4 (382.6)	382.6 (199.1)	381.5 (197.9)	374.6 (210.2)
Calories of afternoon snack	188.9 (113.7)[b]	175.2 (121.6)[b]	207.4 (118.4)	216.2 (108.8)	197.8 (132.6)	203.8 (133.6)	185.6 (140.0)	188.5 (134.8)
Calories of dinner	402.5 (208.4)[b]	321.4 (192.9)[b]	442.2 (197.2)[b]	383.1 (176.6)[b]	417.7 (223.0)[b]	367.5 (227.0)[b]	405.4 (244.1)[b]	339.1 (235.8)[b]
Calories of evening snack	192.4 (138.7)[b]	174.3 (146.2)[b]	203.0 (149.4)	187.9 (152.3)	178.2 (171.4)	173.8 (163.0)	162.3 (166.3)	149.5 (164.8)

[a]OECD: Organization for Economic Co-operation and Development.

[b]Significant at the *P*<.05 level.

Comparison of the Total Frequency of Data Entry and Overall Calorie Consumption Between Users From Low-Income and OECD Countries

Independent two-tailed sample *t* tests were conducted to compare the total frequency and overall calorie consumption between users from low-income and OECD countries. There was a statistically significant difference in the frequency of data entry of morning snack (t_{8351}=−5.4; *P*<.001) and afternoon snack (t_{8351}=3.9; *P*<.001) between the two groups of users (Table 6). There were also significant differences in the average calories for exercise (t_{8351}=−2.4; *P*=.02) and dinner (t_{8351}=−5.9; *P*<.001; Table 7).

Table 6. Results of independent two-tailed *t* tests to compare the total frequency of data entry between Organization for Economic Co-operation and Development (n=8041) and low-income countries (n=312).

Characteristics	Organization for Economic Co-operation and Development countries (n=8041), mean (SD)	Low-income countries (n=321), mean (SD)	*t* test (*df*)	*P* value	*t*-value (95% CI)
Total frequency of exercise data entry	123.6 (97.4)	126.2 (96.6)	0.1 (8351)	.65	−8.4 to 13.6
Total frequency of breakfast data entry	198.5 (99.4)	199.3 (103.8)	−0.2 (8351)	.88	−10.4 to 12.1
Total frequency of morning snack data entry	88.0 (85.1)	114.8 (99.7)	−5.4 (8351)	<.001	17.1 to 36.5
Total frequency of lunch data entry	183.2 (98.7)	179.7 (104.5)	−0.9 (8351)	.53	−14.8 to 7.6
Total frequency of afternoon snack data entry	101.3 (84.4)	120.4 (97.1)	3.9 (8351)	<.001	9.5 to 28.7
Total frequency of dinner data entry	152.6 (100.6)	149.36 (106.3)	1.1 (8351)	.58	−15.3 to 8.8
Total frequency of evening snack data entry	57.3 (69.9)	62.74 (76.0)	−0.3 (8351)	.18	−7.3 to 9.2

Table 7. Results of independent two-tailed *t* tests to compare the overall calories between Organization for Economic Co-operation and Development (n=8041) and low-income countries (n=312).

Characteristics	Organization for Economic Co-operation and Development countries, mean (SD)	Low-income countries, mean (SD)	*t* test (*df*)	*P* value	*t*-value (95% CI)
Exercise	306.8 (192.2)	280.4 (190.1)	−2.4 (8351)	.02	−48.2 to −4.7
Breakfast	287.9 (110.2)	293.8 (112.0)	1.0 (8351)	.35	−6.6 to 18.4
Morning snack	179.4 (94.8)	187.4 (107.6)	1.1 (8351)	.15	−2.8 to 18.8
Lunch	414.1 (150.5)	408.6 (163.4)	−0.6 (8351)	.53	−22.5 to 11.6
Afternoon snack	224.9 (122.5)	226.6 (107.2)	1.2 (8351)	.81	−12.1 to 15.5
Dinner	461.0 (172.7)	401.9 (171.2)	−5.9 (8351)	<.001	−78.7 to −39.6
Evening snack	236.0 (123.6)	223.7 (125.0)	−1.8 (8351)	.09	−26.2 to 1.8

Discussion

Principal Findings

The results of this study suggest that there are significant differences in using the app between users from low-income and OECD countries. At most time points, users from low-income countries entered more data into the app than users from OECD countries. However, all users exhibited a decrease in the frequency of app use over time. There has been insufficient research on adherence to mobile health (mHealth) apps, especially regarding users in low-income countries. Accordingly, it is difficult to understand the level of adherence to the use of mHealth apps and to compare differences in adherence rates among different groups of users. However, achieving long-term health-related goals, such as weight loss, requires constant and committed participation in using the mHealth app. In addition, the low retention rate of using mHealth apps has been recognized as a critical problem. Therefore, further research on adherence to using mHealth apps is needed.

According to a recent study, more than two-thirds of people who downloaded an mHealth app used it only once or stopped using it within a short time [18]. Similarly, Lee et al [19] reported the use of mHealth apps to gradually reduce over time. To benefit from an mHealth app, users should continue to use it for a sufficient amount of time so that it can be incorporated into their daily lives [20]. Furthermore, users should put much effort into using mHealth apps for a long time because changing habitual behavior takes a substantial amount of time [19]. Accordingly, researchers and app developers should investigate and adopt essential features that would encourage users to keep using mHealth apps to accomplish their health outcomes.

In our results, the frequency of data entry was significantly associated with weight loss over time. Specifically, users who frequently monitored their lunch, dinner, body weight, evening snack, or exercise exhibited significant weight loss over time. As previous studies established that eating a late and large amount of dinner is associated with weight gain [21,22], it can be assumed that those users who more frequently track their exercise and dinner calories try to increase their movement and reduce dinner calories or make an effort to eat dinner in small quantities and earlier than usual. Similar to our findings, a systematic review found a significant relationship between the frequency of self-monitoring for diet and PA and weight loss [23]. Furthermore, Conroy et al [24] found that higher mean rates of PA self-monitoring were associated with a greater reduction in weight. Moreover, Peterson et al [25] reported that the total number of food records can be a predictor of weight loss, regardless of the type of meals. On the basis of this evidence, it can be concluded that people who more consistently track their food, exercise, and body weight by using the app are more likely to lose weight [26].

In fact, there is insufficient research on the importance of having separate data entries for meals and snacks on weight loss. However, it can be assumed that people can understand their current eating habits and the nutrient value of food by tracking meals and snacks. Through this process, they might enhance their ability to balance total calories and macronutrients throughout the day and reduce their body weight in the long term [27]. Regular weighing is another essential and simple way to lose weight by boosting motivation [28]. A study found that people who never weighed themselves or only weighed themselves once per week did not exhibit weight reduction. However, those who weighed themselves almost daily exhibited significant weight loss within 1 year [29]. The monitoring and logging of exercise can encourage people to move more throughout the day; therefore, people can efficiently reduce their body weight [30]. To summarize, many studies have found that monitoring health-related behaviors or weight can increase people's perceptions of the effects of changing behaviors related to weight loss [29,31,32].

It was found that there were no significant differences in weight change between users living in low-income or OECD countries. Specifically, we found that self-monitoring of PA, diet, or body weight through a smartphone app can be a useful tool to lose weight regardless of the users' country of residence. Users from the low-income and OECD countries showed statistically significant weight loss at 9 months after starting using the app; however, users from OECD countries regained their weight and

returned to their original body weight in 12 months. Lifestyle, including not only diet and PA patterns, varies considerably across countries [33]. Specifically, although high-income countries generally have better diets based on healthy foods, they also have substantially poorer diets because of a higher intake of unhealthy foods compared with low-income countries [34]. In addition, high-income countries are often linked to low levels of PA levels [35]. These reasons could have influenced the weight gain of users from OECD countries in this study. However, further research on the reasons that made users from OECD countries gain weight is needed.

In this study, we assumed that the users of the paid version of the app would show more body weight reduction than users who used the free version of the app. Our prediction was based on the fact that the paid version offers a variety of features, including customized diet planning, one-on-one coaching, social support, weight recording logs, and food and exercise tracking, whereas the free version only allows users to log and track their data on diet, exercise, and weight [36]. However, the results showed that there were no statistically significant differences in body weight changes over time between the users of the 2 versions of the app. A previous systematic review found that self-monitoring of diet, exercise, and weight is the core feature of behavioral weight loss intervention programs [37]. However, there is a lack of information about the features of smartphone health apps that are the most valuable in terms of weight loss. To develop an efficient and effective app for weight loss, further research is required to identify such features.

The imminent global threat of NCDs calls for urgent solutions that would extend existing health systems into the community [38]. In recent years, smartphone health care apps have attracted significant attention as effective interventions to prevent NCDs [39]. However, despite the widespread use of smartphone apps as interventions to lose weight, previous research on body weight changes according to using smartphone health apps in low-income countries remains scarce [40]. This study compared the use of an app and its effectiveness on body weight changes over time in low-income and OECD countries. The results showed no significant differences in the use and effectiveness of the app between users from the two groups of countries. Although the app provides a specialized diet and PA monitoring programs in many languages and targets diverse populations around the world, the number of users in OECD countries is more than 4 times the number in low-income countries. Although it is difficult to conclude that our data set contained all users' information about the app, this information provides a need to consider the barriers and challenges of smartphone use in low-income countries. As smartphone apps can provide extraordinary health opportunities to users from low-income countries, which seriously lack health infrastructure and clinical resources, many efforts to enhance the use of smartphone health apps are needed so that these apps to reach their full potential in low-income countries [6].

This study has several limitations. First, as our study was a retrospective cohort study, the results may be limited by the observational nature of the data set. To accurately evaluate the effectiveness of the app, well-developed randomized controlled trials should be conducted. Second, our data set included only a representative selection of OECD and low-income countries, which might have affected the generalizability of our findings. Third, our analysis did not include some important demographic factors, socioeconomic factors, and lifestyle variables that could affect the use of the smartphone app and body weight changes. Fourth, the users downloaded and used the app with the intention of monitoring their diet and PA and to voluntarily lose weight. Therefore, this might have made the results significant in this analysis. Fifth, it is possible that users may engage in a suitable diet and PA without logging or self-reporting. Finally, the analyzed data were self-reported, which might have affected the accuracy of our conclusions. Despite the aforementioned limitations, the results of this study provide valuable insights for investigators, engineers, politicians, policymakers, government authorities, health care providers, and the general population worldwide in terms of highlighting the benefits of using health care smartphone apps in health promotion. Our results also highlight the strong potential of the studied app [41].

Conclusions

In conclusion, this study is the first to investigate the effectiveness of using a smartphone app on losing body weight between users from low-income and OECD countries. On the basis of the results, it can be concluded that the frequent input of self-monitoring data, the main function of the Noom app, can be an effective approach to weight loss. We also found no significant differences in body weight changes between users from the two groups of countries, suggesting that the smartphone app can be an effective and general way to lose weight regardless of the users' country of residence. In addition, we also found that using even the free version of the app increases the frequency of self-monitoring and thus positively affects weight loss.

Acknowledgments

The authors would like to thank Emeritus Professor Young Seol Kim at Kyung Hee University for his teaching and inspiration, which encouraged us to conduct this study. This research was supported by a grant from the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute, which is funded by the Ministry of Health and Welfare, Republic of Korea (grant HI16C2048).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Low-income countries.
[DOCX File , 14 KB - mhealth_v9i7e13496_app1.docx]

Multimedia Appendix 2
Organization for Economic Co-operation and Development countries.
[DOCX File , 15 KB - mhealth_v9i7e13496_app2.docx]

References

1. Noncommunicable Diseases. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2020-01-24]
2. Lim S, Vos T, Flaxman D, 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. Lachat C, Otchere S, Roberfroid D, Abdulai A, Seret FM, Milesevic J, et al. Diet and physical activity for the prevention of noncommunicable diseases in low- and middle-income countries: a systematic policy review. PLoS Med 2013 Jun 11;10(6):e1001465 [FREE Full text] [doi: 10.1371/journal.pmed.1001465] [Medline: 23776415]
4. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct 22;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
5. Mayes J, White A. How smartphone technology is changing health care in developing countries. Journal of Glob Heal 2017;7(2):36-38.
6. Kahn J, Yang J. Mobile health needs and opportunities in developing countries. Health Aff 2010;29(2):258.
7. Gruessner V. The Advantages of Mobile Health Apps Today and Tomorrow. mhealth Intelligence. 2015. URL: https://mhealthintelligence.com/news/the-advantages-of-mobile-health-apps-today-and-tomorrow [accessed 2020-02-18]
8. Kaplan W. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? Global Health 2006 May 23;2:9 [FREE Full text] [doi: 10.1186/1744-8603-2-9] [Medline: 16719925]
9. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
10. Milne-Ives M, Lam C, de Cock C, van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
11. Ustulin M, Keum C, Woo J, Woo J, Rhee SY. Effects of climatic variables on weight loss: a global analysis. Sci Rep 2017 Jan 20;7(1):40708-40707 [FREE Full text] [doi: 10.1038/srep40708] [Medline: 28106167]
12. Bhurosy T, Jeewon R. Overweight and obesity epidemic in developing countries: a problem with diet, physical activity, or socioeconomic status? Sci World J 2014;2014:964236 [FREE Full text] [doi: 10.1155/2014/964236] [Medline: 25379554]
13. List of OECD Member Countries-ratification of the Convention on the OECD. The Organisation for Economic Co-operation and Development. 2018. URL: https://www.oecd.org/about/members-and-partners/ [accessed 2020-03-01]
14. World Economic Outlook Database. International Monetary Fund. 2018. URL: https://www.imf.org/en/Publications/SPROLLs/world-economic-outlook-databases [accessed 2020-03-03]
15. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
16. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 7;6(1):34563-34568 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
17. BMI Formula. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/childrens_bmi_formula.html [accessed 2020-02-27]
18. Byrnes N. Mobile Health's Growing Pains. MIT Technol Review. 2014. URL: http://www.technologyreview.com/news/529031/mobile-healths-growing-pains/ [accessed 2020-04-03]
19. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018 Jul 26;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
20. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. Am Psychol 1992;47(9):1102-1114. [doi: 10.1037/0003-066x.47.9.1102]

21. Jakubowicz D, Barnea M, Wainstein J, Froy O. High caloric intake at breakfast vs dinner differentially influences weight loss of overweight and obese women. Obesity (Silver Spring) 2013 Dec 2;21(12):2504-2512 [FREE Full text] [doi: 10.1002/oby.20460] [Medline: 23512957]
22. Eating an Early Dinner Can Help You Burn Fat, Lower Your Blood Sugar. Healthline. 2020. URL: https://www.healthline.com/health-news/eating-an-early-dinner-can-help-you-burn-fat-lower-your-blood-sugar [accessed 2020-02-18]
23. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
24. Conroy M, Yang K, Burke L. Physical activity self-monitoring and weight loss: 6-months results of the SMART trial. Med Sci Sport Exerc 2011;43(8):1574. [doi: 10.1249/mss.0b013e31820b9395]
25. Peterson ND, Middleton KR, Nackers LM, Medina KE, Milsom VA, Perri MG. Dietary self-monitoring and long-term success with weight management. Obesity (Silver Spring) 2014 Sep 13;22(9):1962-1967 [FREE Full text] [doi: 10.1002/oby.20807] [Medline: 24931055]
26. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014 Sep;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
27. Lesko A. 8 Reasons Why You Should Try Tracking Your Food. Fit-Flavors. 2019. URL: https://fit-flavors.com/blogs/nutrition/8-reasons-why-you-should-try-tracking-your-food [accessed 2020-06-07]
28. Jenny. What to Track When You're Losing Weight. Progress Blog. 2017. URL: https://theprogressapp.com/blog/what-track-losing-weight/ [accessed 2020-06-30]
29. American HA. Daily Weighing May Be Key to Losing Weight. Sicence News. 2018. URL: https://www.sciencedaily.com/releases/2018/11/181105081735.htm [accessed 2020-05-12]
30. Hartmann-Boyce J, Boylan A, Jebb SA, Aveyard P. Experiences of self-monitoring in self-directed weight loss and weight loss maintenance: systematic review of qualitative studies. Qual Health Res 2019 Jan 9;29(1):124-134. [doi: 10.1177/1049732318784815] [Medline: 29984630]
31. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
32. Aromatario O, Van Hoye A, Vuillemin A, Foucaut A, Crozet C, Pommier J, et al. How do mobile health applications support behaviour changes? A scoping review of mobile health applications relating to physical activity and eating behaviours. Public Health 2019 Oct;175:8-18 [FREE Full text] [doi: 10.1016/j.puhe.2019.06.011] [Medline: 31374453]
33. Maher A, Sridhar D. Political priority in the global fight against non-communicable diseases. J Glob Health 2012 Dec;2(2):020403-020423 [FREE Full text] [doi: 10.7189/jogh.02.020403] [Medline: 23289078]
34. The Lancet Global Health: Unhealthy Eating Habits Outpacing Healthy Eating Patterns in Most World Regions. The Lancet. 2015. URL: https://www.sciencedaily.com/releases/2015/02/150218191719.htm [accessed 2020-03-07]
35. Physical Activity Fact Sheet. World Health Organization. 2016. URL: https://www.euro.who.int/en/health-topics/disease-prevention/physical-activity/data-and-statistics/physical-activity-fact-sheets [accessed 2020-07-03]
36. What is Noom Diet Plan? A Nutritionist Honest Review. For Care Education and Research. URL: https://fcer.org/noom/ [accessed 2020-05-17]
37. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
38. Kwan A. Mhealth Opportunities for Non-communicable Diseases Among the Eldery. Click Medix. 2017. URL: https://clickmedix.com/2013/10/31/mhealth-opportunities-for-non-communicable-diseases-among-the-eldery/ [accessed 2020-09-07]
39. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583]
40. Xiao Y, Zhao N, Wang H, Zhang J, He Q, Su D, et al. Association between socioeconomic status and obesity in a Chinese adult population. BMC Public Health 2013 Apr 17;13(1):355-323 [FREE Full text] [doi: 10.1186/1471-2458-13-355] [Medline: 23590682]
41. Latif S, Rana R, Qadir J, Ali A, Imran MA, Younis MS. Mobile health in the developing world: review of literature and lessons from a case study. IEEE Access 2017;5:11540-11556. [doi: 10.1109/access.2017.2710800]

Abbreviations

LMM: linear mixed model
mHealth: mobile health
NCD: noncommunicable disease
OECD: Organization for Economic Co-operation and Development
PA: physical activity

Edited by L Buis; submitted 25.01.19; peer-reviewed by C Loum, H Ranjani, M Andrews, C Mauch, J Ceasar, S Hughes; comments to author 27.04.19; revised version received 27.03.20; accepted 17.05.21; published 12.07.21.

Please cite as:
Han M, Rhee SY
Effect of Adherence to Smartphone App Use on the Long-term Effectiveness of Weight Loss in Developing and OECD Countries: Retrospective Cohort Study
JMIR Mhealth Uhealth 2021;9(7):e13496
URL: https://mhealth.jmir.org/2021/7/e13496
doi:10.2196/13496
PMID:34255708

Original Paper

A Mobile-Based Intervention to Increase Self-esteem in Students With Depressive Symptoms: Randomized Controlled Trial

Alina Bruhns[1], MSc; Thies Lüdtke[1], PhD; Steffen Moritz[1], Prof Dr; Lara Bücker[1], PhD

University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany

Corresponding Author:
Lara Bücker, PhD
University Medical Center Hamburg-Eppendorf (UKE)
Martinistraße 52
Hamburg, 20251
Germany
Phone: 49 40741055868
Email: l.buecker@uke.de

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2022/5/e39448

Abstract

Background: Depressive symptoms are one of the most common and ever-increasing mental health problems among students worldwide. Conventional treatment options, particularly psychotherapy, do not reach all students in need of help. Internet- and mobile-based interventions are promising alternatives for narrowing the treatment gap.

Objective: In the framework of a randomized controlled trial, we aim to investigate the effectiveness, acceptance, and side effects of a self-help smartphone app (*MCT & More*) based on cognitive behavioral therapy, mindfulness, acceptance and commitment therapy, and metacognitive training in a sample of students with self-reported depressive symptoms. Furthermore, we were interested in examining the influence of treatment expectations and attitudes toward internet- and mobile-based interventions on treatment adherence and effectiveness.

Methods: A total of 400 students were recruited via open access websites and randomized to either the intervention group (n=200), who received access to the self-help smartphone app *MCT & More* for a period of 4 weeks, or to a wait-list control group (n=200). The Patient Health Questionnaire-9 (depression) served as the primary outcome parameter, and the Rosenberg Self-esteem Scale (self-esteem) and the global item of the World Health Organization Quality of Life-abbreviated version (quality of life) served as the secondary outcome parameters. The Attitudes Towards Psychological Online Interventions was used to measure attitudes toward internet- and mobile-based interventions. Outcome expectations were assessed using the Patient Questionnaire on Therapy Expectation and Evaluation, and side effects were assessed using the Inventory for Assessing Negative Effects of Psychotherapy.

Results: Per-protocol (PP), complete-case, and intention-to-treat analyses showed a significantly higher reduction in depressive symptoms (PP: $F_{1,222}$=3.98; P=.047; d=0.26) and a significantly higher increase in self-esteem (PP: $F_{1,220}$=8.79; P=.003; d=0.40) in the intervention group than in the wait-list control group. Most participants regularly used the self-help smartphone app (91/120, 75.8%, at least once a week). The more positive the attitude toward internet- and mobile-based interventions (r=0.260; P=.004) and the more positive the outcome expectation (r=0.236; P=.009), the more frequently the self-help smartphone app was used.

Conclusions: The effectiveness of the self-help smartphone app *MCT & More* was demonstrated among students with depressive symptoms compared with a wait-list control group. The app could be offered regularly as a low-threshold intervention to enhance students' health.

Trial Registration: German Clinical Trials Register DRKS00020941; https://tinyurl.com/pr84w6er

(JMIR Mhealth Uhealth 2021;9(7):e26498) doi:10.2196/26498

KEYWORDS

mHealth; depression; depressive symptoms; students' mental health; self-help smartphone app; mobile phone; self-esteem

Introduction

Background

Universities worldwide are confronted with increasing rates of mental health problems among students [1]. In Germany, 15.6% of the university students state that they are currently affected by depressive symptoms (women: 16.9% and men: 14%) [2]. Adjustments to the new living environment (eg, moving away from home and new social environment), expectations of academic performance (eg, final grade in the master's program), and other stressors resulting from the university program (50% of male and 65% of female students cite their studies as the most common cause of stress) render students particularly vulnerable to developing mental disorders [3]. Mental disorders such as depressive symptoms affect the social and general functioning of students, thereby negatively influencing the course of their study [4] and leading to a deterioration in academic performance. High depression scores among students have been shown to be associated with a low grade point average [5], and depressive symptoms during examination periods predict low future grades [6]. In addition, students with health problems (mental and physical) take long study durations, change their course of study or university more often, and are less likely to have a secure livelihood [4].

Despite their negative impact on functioning, mental disorders among students often remain undertreated [7]. The treatment gap can be attributed to the lack of available psychotherapy, especially in rural areas [8]; self-stigmatization [9]; fear of being stigmatized by others [10]; the preference to solve the problem independently [8]; fear of having to talk about one's own problems to a psychotherapist; or the high treatment costs that may arise. Furthermore, depressive symptoms are often not recognized or misinterpreted by primary care physicians [11], leading to reduced help-seeking behavior [12]. Due to the treatment gap, universities are encouraged to initiate help offers that better reach the affected students [2].

The Potential of Internet-Based Interventions in the Treatment of Depression

In Germany, virtually all individuals (>99%) aged between 16 and 44 years use the internet [13]. German students belong to the generation of *digital natives* (confident in using computer technology), so it is assumed that students can easily use internet-based interventions [14]. The benefits of internet-based interventions are, inter alia, the high level of autonomy and privacy. They can be used from any location and are often available free of charge or at a low cost. They are not intended to replace traditional psychotherapy but to expand conventional care [15]. In the last decade, numerous internet-based interventions, especially for the treatment of anxiety disorders and depression, have been developed and tested for their efficacy [16-18]. Systematic reviews and meta-analyses have shown that they can improve mental health problems such as depression, anxiety, and stress among students [14,19,20].

Internet-based interventions can be categorized as either guided or self-guided. Although guided internet-based interventions are supported by a therapist or a trained person (eg, via frequent email correspondence or telephone support), self-guided ones do not provide additional human support. A recent meta-analysis indicated that guided internet-based interventions show higher effect sizes than self-guided ones (guided: g=0.65 and unguided: g=0.27) [21]. However, self-guided internet-based interventions have the advantage that they can be made available to a broad population requiring few resources (no psychotherapists required, can be used at any time without waiting time, and low costs for users [22]), making self-guided internet-based interventions easier to implement at universities. Studies have shown that guided and unguided internet-based interventions show similar effectiveness in direct comparison (ie, when the same intervention is evaluated as guided and unguided) [23,24].

Advantages of Mobile-Based Interventions

Smartphones are the most used technological devices among students on campus [25]. On average, students use their smartphones for approximately 5 hours a day and check their smartphones 28 times daily, which suggests that mobile-based interventions could be highly appealing to students. Mobile-based interventions have already proven to be an effective strategy for improving health-promoting behavior in the general population (eg, physical activity and weight control) [26]. Recent studies have also indicated the effectiveness of self-help smartphone apps in treating depressive symptoms in university students [27,28]. One of the benefits of mobile-based interventions is that smartphones can be accessed almost anytime and are independent of location [29,30]. Therefore, these exercises can be easily integrated into everyday life of students. Furthermore, mobile-based interventions provide the possibility to link users with other forms of support (eg, telephone numbers for acute crises) and to send reminders to the users. By sending reminders, the adherence of the users can be increased in self-guided treatment for anxiety and depression (eg, course completion with reminders: 58% and course completion without reminders: 35%) [31].

However, not all health care apps adequately protect the sensitive health data of the users (eg, commercialization of app users' data) [32], and many German-language depression apps show limitations in quality (eg, in functionality, information quality, esthetics, and user involvement) [33]. Therefore, there is a need for high-quality apps, which are being investigated with regard to their benefits and risks [33]. Previous meta-analyses have found small effect sizes (Hedges g=0.22-0.33) [34,35] in reducing depressive symptoms. Furthermore, a recent study suggested that the actual treatment outcome (reduction in depressive symptoms) for mobile-based interventions can be predicted by the expected treatment outcome (usefulness for the patient: B=0.364 and perception of how logical the treatment is: B=0.528) [36].

Objective of the Study

The overall goal of the study is to improve the health of students at German universities. The aim of this study is to examine the acceptance and efficacy of the self-help smartphone app *MCT & More* among German students with depressive symptoms in comparison with a wait-list control group. To our knowledge, little research has been conducted on the effectiveness and

acceptance of self-help smartphone apps for students with depressive symptoms (especially in German-speaking countries).

A previous version of the app was positively evaluated by the users. In a randomized controlled trial (intervention group and wait-list control group) comprising 90 participants with reported depressive symptoms, it was shown that the app was effective in reducing depressive symptoms when used regularly (ie, several times a week, *P*=.05) app was effective in reducing depressive symptoms when used regularly (ie, several times a week) [37]. It was expected that the use of the self-help smartphone app would lead to a stronger reduction in depressive symptoms and to a higher increase in self-esteem and quality of life in the intervention group than in the wait-list control group after the intervention period. Another novel aspect we aimed to investigate was whether the effect of the app can be predicted by the attitudes toward the internet- and mobile-based interventions and the expected outcome. Moreover, we examined the possible side effects of self-help smartphone apps, which, to our knowledge, have barely been studied to date. Furthermore, an exploratory moderation analysis was conducted to identify possible moderators that affect differential symptom improvement (per-protocol [PP] sample).

Methods

Design

Two web-based assessments were performed at baseline (t0) and 4 weeks later (t1). All participants provided web-based informed consent at the beginning of the baseline assessment. No personal information was requested at any time, except for an anonymous email address (instructions to create an anonymous email address were given) and a personal codeword (consisting of the first letters of the parents' names and some figures of their dates of birth). The collected data were anonymized and stored electronically on a password-protected computer. By providing the codeword or the anonymized email address, the data could be deleted at the request of the participants. At the end of the postassessment period, both groups were given access to a self-help manual, as an incentive, to improve emotional problems. As common in web-based trials, blinding of participants was not possible. The study was conducted in accordance with the Declaration of Helsinki. The local psychological ethics committee of the Center for Psychosocial Medicine of the University Medical Center Hamburg-Eppendorf assessed the study project as ethically unobjectionable (approval number: LPEK-0122).

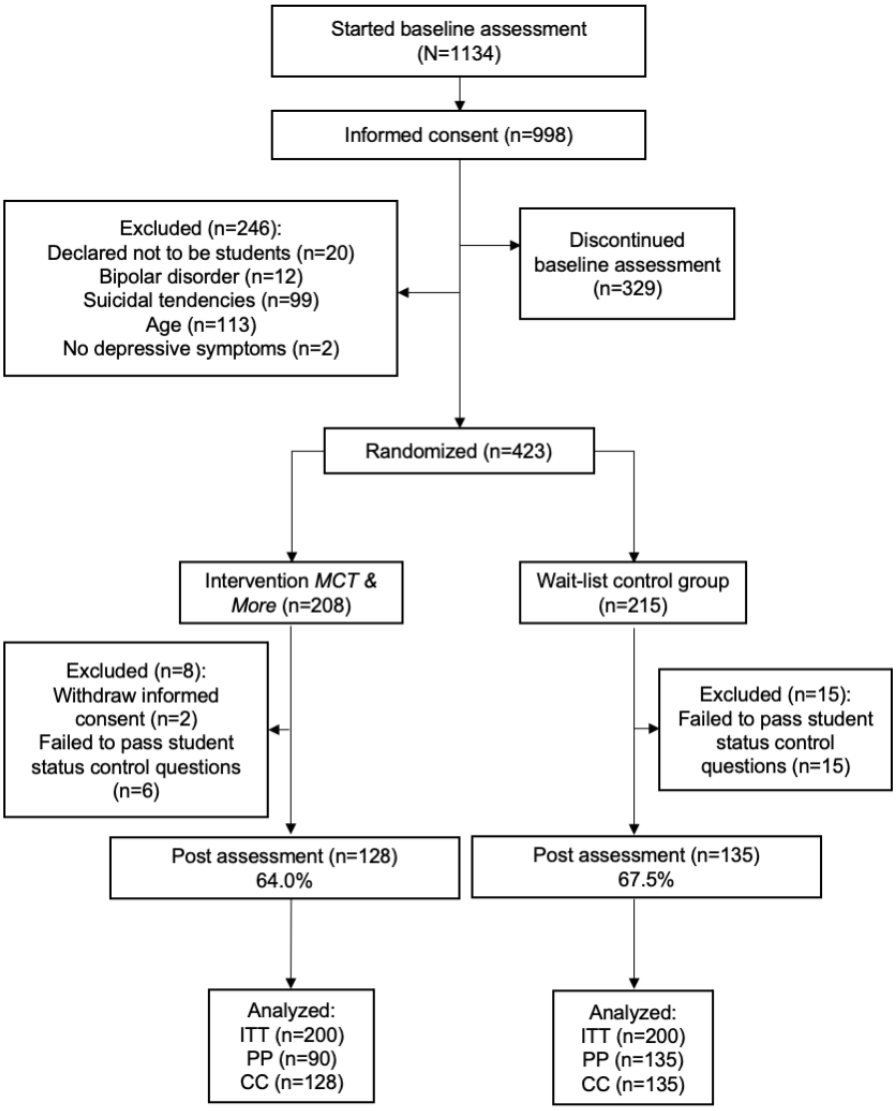
Participants

We recruited participants via web-based platforms and forums by posting an invitation to the study with a link to the web-based baseline assessment. At the beginning of the baseline assessment, participants received detailed information about the study's goals and procedures and were informed about the underlying data protection. An electronic informed consent form was obtained from each participant.

The following inclusion criteria had to be met: student at a German university (whether the participants were actually students was checked by asking questions about the study system in Germany, which are difficult to answer correctly for nonstudents, eg, "What scoring system is used to measure your academic performance?"), aged at least 18 years, willing to provide informed consent, having access to the internet and a smartphone, having depressive symptoms (measured by Patient Health Questionnaire [PHQ-9], total score>0), willingness to participate in 2 pseudonymous web-based assessments, willingness to use the self-help smartphone app for a period of 4 weeks on one's own responsibility, willingness to leave an anonymous email address, no acute suicidal tendencies (measured with item 9 of PHQ-9, cut-off>1), and no current or past bipolar or psychotic disorder. Other psychiatric diagnoses were not a criterion for exclusion. Parallel treatments (eg, psychotherapy or pharmacotherapy) could be continued during participation. If the inclusion criteria were not met, the participants were automatically excluded from the web-based assessment. Then, participants were informed about the reason for exclusion and received information about other help-seeking resources, such as telephone numbers for acute crisis.

Data collection took place in Germany from March 16, 2020 (first baseline assessment) to July 06, 2020 (last postassessment). During this period, Germany experienced the first wave of the COVID-19 pandemic. A total of 246 participants had to be excluded because the inclusion criteria were not met. The final sample consisted of 400 individuals (Figure 1).

Figure 1. Flowchart. CC: complete-case; ITT: intention-to-treat; MCT: metacognitive training; PP: per-protocol.



Procedure

At both measurement points (t0 and t1), data were collected using the survey software Qualtrics. Multiple registrations from one device were detected and prevented by the program. In the baseline assessment, sociodemographic and psychopathological data as well as the attitude toward internet- and mobile-based interventions and expected treatment outcomes were assessed. After the 4-week intervention period, all participants were invited via email to participate in the postassessment and were asked to provide their anonymous email address and personal code again to ensure a correct matching of predata and postdata. Afterward, the participants filled out the same psychopathological questionnaires used in the baseline assessment. In addition, the participants were asked about use frequency ("How often have you used the app during the last 4 weeks?"), side effects, and satisfaction with the self-help smartphone app (refer to the *Measures* section). The study was conducted at Hamburg-Eppendorf University Medical Center (Germany).

Randomization

Randomization was performed using Qualtrics survey software after the baseline assessment. The option *equal distribution* ensured that there was a balanced distribution between the 2 groups. The allocation rule was set to 1:1.

Sample Size

The calculation of the sample size for an analysis of covariance (ANCOVA) with 2 groups was performed using G*Power. The results indicated a sample size of 351 participants based on a small effect of f=0.15, with α=.05, and a power of 0.80. Considering a dropout rate of 15%, the final sample should include 413 participants. The calculation is based on the results of a meta-analysis investigating the effectiveness of smartphone app interventions for depression [34].

Measures

PHQ-9: Depression Module

The self-assessment questionnaire PHQ-9 [38] is the depression module of the Patient Health Questionnaire and is used to measure symptoms of major depression according to the

Diagnostic and Statistical Manual of Mental Disorders-IV. The symptoms are assessed using 9 items on a 4-point rating scale ranging from *not at all* (0) to *almost daily* (3). A total score between 0 and 27 can be calculated. Sum scores of 0-4 indicate none or minimal depressive symptoms, 5-9 indicate mild depressive symptoms, 10-14 indicate moderate depressive symptoms, and 15-27 indicate severe depressive symptoms. The internal consistency ranges from Cronbach α=.86 to .89 [38,39].

Rosenberg Self-esteem Scale

The Rosenberg Self-esteem Scale (RSE) is a self-assessment questionnaire presenting 10 statements on self-esteem, which are rated on a 4-point rating scale (1 to 4) from *strongly agree* to *strongly disagree*. The total score ranges from 10 to 40 points. High scores indicated high self-esteem. Its internal consistency ranges from Cronbach α=.77 to .88 [40].

World Health Organization Quality of Life-Abbreviated Version

In this study, the first item of the World Health Organization Quality of Life-abbreviated version (WHOQOL-BREF; "How would you assess your quality of life?") with the response options *very poor* (1) to *very good* (5) was chosen to measure quality of life. The WHOQOL-BREF [41] is a self-assessment tool with 26 items (4 domains) and represents a short version of the questionnaire World Health Organization Quality of Life-100, which is based on the World Health Organization's concept of quality of life. All domain scores were fairly to moderately correlated with the global quality of life [42,43]. In a sample of medical students, the internal consistency was Cronbach α=.896 [44].

Attitude Toward Psychological Online Interventions

Attitude Toward Psychological Online Interventions (APOI) [45] is a self-assessment tool consisting of 4 dimensions: (1) skepticism and risk perception, (2) trust in therapeutic efficacy, (3) perception of deficits in mechanization, and (4) perception of the advantages of anonymity. The questionnaire consists of 16 items and can be rated on a 5-point rating scale ranging from *do not agree at all* to *fully agree*. The total scale ranges from 16 to 80. A high total score indicates a positive attitude. All 4 dimensions were equally weighted. The APOI has an internal consistency of Cronbach α=.77 [45].

Patient Questionnaire on Therapy Expectation and Evaluation

The Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV) [46] is a self-assessment questionnaire that measures therapy expectations and consists of 10 items. The instrument covers 3 subscales: (1) hope of improvement, (2) fear of change, and (3) suitability. Ten statements are presented, which are rated on a 5-point rating scale ranging from *not correct at all* to *completely correct*. The higher the sum of the subscales, the stronger the hope of improvement, fear of change, and suitability. The total scale ranges from 11 to 55, and the internal consistency ranges from Cronbach α=.73 to .83. The questionnaire was adapted to internet- and mobile-based interventions (eg, "I consider the treatment principle of psychological internet- and mobile-based interventions to be reasonable").

Patient Satisfaction (Fragebogen zur Patientenzufriedenheit)

The instrument *Fragebogen zur Patientenzufriedenheit* (ZUF-8) [47] is the German version of the Client Satisfaction Questionnaire-8. The self-assessment questionnaire consists of 8 items that are used to assess patient satisfaction with a treatment, such as psychotherapy, in a 1D and global way. The items can be rated on a 4-point rating scale (eg, *excellent*, *good*, *less good*, and *bad*). A total score (8-32) can be calculated, whereby a high score indicates a high level of satisfaction. The internal consistency ranges from Cronbach α=.87 to .93 [47,48].

Inventory for Assessing Negative Effects of Psychotherapy

The Inventory for Assessing Negative Effects of Psychotherapy (INEP) [49] is a German self-assessment tool with 21 items and focuses on the side effects of psychotherapy regarding intrapersonal changes, partnership, stigma and financial worries, family, friends, dependency, and therapeutic relationship. The instrument consists of 2 scales: side effects (scale 1) and therapeutic misbehavior (scale 2). For the first 6 questions, a 7-point rating scale (−3 to +3, bipolar response format) can be used to indicate the extent to which the respective areas of life have developed positively or negatively from the start of the intervention or whether they have remained unchanged. A unipolar response format is used for questions 7-21 to determine whether a negative effect is experienced and with what intensity (0 to +3) it is perceived. A total score can be calculated for items 1-15, reflecting the number of experienced side effects. Furthermore, the intensity of the experienced side effects can be determined by calculating an average score (1-3, where 3 indicates a high intensity of the side effect). The INEP had an internal consistency of Cronbach α=.86 [49]. As no therapeutic relationship could be developed during the use of the self-help smartphone app, items 16-21 were excluded from the assessment. The wording was slightly adapted (self-help app instead of psychotherapy).

Intervention

During the 4-week intervention period, the intervention group had free access to the self-help smartphone app *MCT & More* (Textbox 1), which is primarily intended for individuals with depressive symptoms.

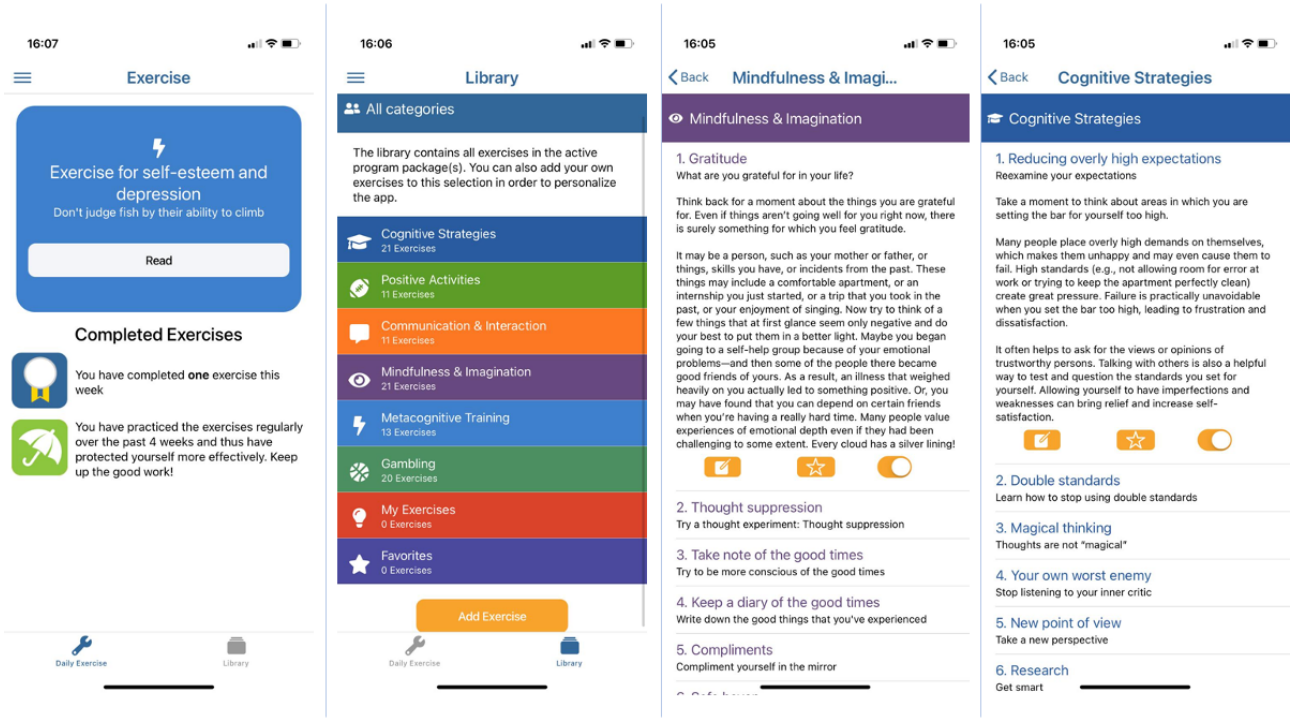
The basic package of the self-help smartphone app *mood* comprises 57 short exercises on the following topics: cognitive strategies, communication and interaction, positive activities, and mindfulness and imagination. The program package *gambling* was developed especially for individuals with gambling problems, and the program package *metacognitive training* was intended for individuals with psychotic experiences. These program packages are deactivated by default settings but can also be useful for people who are not affected by the addressed symptoms. They can be activated in the app by the users themselves. The exercises take only a few minutes and are designed to be easy to use in the everyday life of students (Figure 2).

Textbox 1. Module descriptions of the MCT & More smartphone app.

Modules and Their Descriptions

- Cognitive strategies
  - Techniques to better perceive and understand own thinking and to break through adverse behavior patterns (eg, cognitive reframing and motivational techniques)
- Positive activities
  - Behavior activation; interaction of body and emotion; and focus on positive things in life, own values and skills
- Communication and interaction
  - Dealing with feedback, ways to improve communication skills, setting boundaries, and approaching others
- Mindfulness and imagination
  - Conscious perception and experience of the current moment, acceptance, gratitude, focusing on positive things, and relaxation
- Metacognitive training
  - Modification of cognitive biases and dysfunctional beliefs and reducing stigmatization and shame
- Gambling
  - Modification of gambling-specific cognitive distortions, support with financial burdens, and dealing with relapses and gambling impulses
- My exercises
  - Own exercises can be created

Figure 2. Screenshots of exercises.



The contents and exercises are based on group metacognitive training (MCT) [50], cognitive behavioral therapy [51,52], and third wave techniques (eg, acceptance and mindfulness) [53,54]. The metacognitive training (MCT) was originally developed for people with psychosis [50]. Inspired by metacognitive training (MCT), a (group) training specifically for depression has evolved (metacognitive training for depression, D-MCT) [55]. Meta-analyses showed that metacognitive training (MCT) is effective in reducing anxiety, depression, and dysfunctional metacognitions (g=1.81-2.06) [56,57]. The app can be used as an add-on of metacognitive training for depression (D-MCT) but can also be used standalone. Metacognitive training for depression (D-MCT) focuses on the modification of cognitive biases and beliefs associated with the onset and maintenance

of mental disorders such as psychosis and depression [50,55]. The training seeks to enable individuals to recognize and correct automatic and unconscious thought patterns. It also targets dysfunctional assumptions about thought processes as well as dysfunctional coping strategies (eg, social withdrawal, thought suppression, and rumination). On the basis of the principle that taking care of personal psychological well-being is a bit like brushing one's teeth, the exercises should be performed regularly so that they become routine. Therefore, the app sends daily reminders via push messages. In addition, the *MCT & More* app contains gamification elements. Depending on the number of exercises completed, users can collect bronze, silver, or gold medals and obtain an open umbrella as a symbol for long-term protection. The app also offers the ability to create their own exercises. A learning algorithm, that is, an automatic adaptation to the user's behavior, was not integrated (the app does not fall under the Medical Devices Act). The app *MCT & More* is currently available in German, English, Arabic, Turkish, Persian, and Serbian and can be downloaded free of charge for both Android and iOS operating systems.

The app has been continuously developed (eg, gamification elements, design, program packages *gambling* and *metacognitive training*, additional exercises in the other program packages, and various language versions) since the last evaluation [37]. The self-help smartphone app did not undergo major changes during the evaluation process of this study.

Statistical Analyses

IBM Statistics 26 was used for statistical analysis. Independent samples *t* tests and chi-square tests were performed to compute group differences in baseline characteristics. Between-group differences over time (preintervention to postintervention) were calculated using ANCOVA with baseline scores as covariates. Pre-post differences were defined as within-group factors and groups as between-group factors. Paired samples *t* tests were used to analyze within-group differences. To determine the efficacy of the self-help smartphone app, intention-to-treat (ITT), PP, and complete-case (CC) analyses were performed. In the ITT analyses, all participants for whom baseline data were available were included in the evaluation. Missing data for the postvalues were calculated using expectation maximization. The PP analyses included only those participants who used the intervention as intended (at least once a week) and completed the postassessment. CC analyses included all participants who completed the postassessment (regardless of whether and how often the intervention was used). In the guidelines of the CONSORT (Consolidated Standards of Reporting Trials), it is recommended to perform both ITT and PP analyses in randomized controlled trials. With their conservative approach, ITT analyses comply with the guidelines of Good Clinical Practice and can be considered the gold standard for the evaluation of treatment effects [58,59]. In the PP analyses, the evaluation of the treatment effect is carried out under ideal conditions, so they provide an estimation of the actual efficacy. Furthermore, an explorative moderation analysis was carried out for the PP sample to identify possible moderators (included moderator variables were sociodemographic data, psychometric scales, and medication; Table 1) that affected differential symptom improvement (outcome measure: PHQ-9) using SPSS macro PROCESS by Hayes [60].

Table 1. Demographic description of the intention-to-treat sample (N=400).

Baseline characteristics	Intervention group (n=200)	Wait-list control group (n=200)	Chi-square test (*df*)	*t* test (*df*)	*P* value
Sociodemographic data					
Male, n (%)	24 (12)	19 (9.5)	1.0 (2)	N/A[a]	.60
Age (years), mean (SD)	23.13 (3.56)	22.84 (3.15)	N/A	−0.86 (398)	.39
German, n (%)	190 (95)	194 (97)	1.0 (1)	N/A	.31
School education (years), mean (SD)	12.39 (1)	12.48 (0.92)	N/A	0.91 (398)	.37
Marital status, n (%)			1.3 (3)	N/A	.73
Single	99 (49.5)	94 (47)			
Relationship	95 (47.5)	101 (50.5)			
Married	5 (2.5)	5 (2.5)			
Divorced	1 (0.5)	0 (0)			
Field of study, n (%)			5.1 (6)	N/A	.53
Engineering	4 (2)	3 (1.5)			
Natural sciences	9 (4.5)	13 (6.5)			
Medical science or health	61 (30.5)	59 (29.5)			
Legal sciences or economics	24 (12)	19 (9.5)			
Linguistics or culture	10 (5)	6 (3)			
Social sciences	88 (44)	90 (45)			
Others	4 (2)	10 (5)			
Semester, mean (SD)	5.6 (3.63)	5.57 (3.57)	N/A	−1.18 (398)	.24
Psychometric scales or psychiatric disorders, n (%)					
None	100 (50)	104 (52)	0.2 (1)	N/A	.69
Anxiety	54 (27)	40 (20)	2.7 (1)	N/A	.10
Depression	72 (36)	69 (34.5)	0.1 (1)	N/A	.75
PTSD[b]	11 (5.5)	11 (5.5)	0.0 (1)	N/A	.99
Alcohol or drug addiction	7 (3.5)	1 (0.5)	4.6 (1)	N/A	.03
OCD[c]	5 (2.5)	17 (8.5)	6.9 (1)	N/A	.008
Eating disorder	9 (4.5)	9 (4.5)	0.00 (1)	N/A	.99
Personality disorder	9 (4.5)	1 (0.5)	6.6 (1)	N/A	.01
ADD[d]	4 (2)	2 (1.0)	0.7 (1)	N/A	.41
Others	5 (2.5)	2 (1)	0.4 (1)	N/A	.56
Medication, n (%)					
None	155 (77.5)	157 (78.5)	0.1 (1)	N/A	.81
Antidepressants	16 (8.0)	17 (8.5)	0.0 (1)	N/A	.86
Measurements, mean (SD)					
PHQ-9[e]	11.13 (4.99)	10.98 (4.42)	N/A	−0.31 (398)	.76
WHOQOL-BREF[f]	3.64 (0.85)	3.7 (0.75)	N/A	0.69 (398)	.49
RSE[g]	25.73 (6.12)	26.62 (5.83)	N/A	1.49 (398)	.14
APOI[h]	50.27 (7.34)	49.92 (8.06)	N/A	−.46 (398)	.65
PATHEV[i]	35.69 (5.61)	36.35 (5.07)	N/A	1.15 (398)	.25
Psychotherapy experiences					

Baseline characteristics	Intervention group (n=200)	Wait-list control group (n=200)	Chi-square test (*df*)	*t* test (*df*)	*P* value
Previous treatments, n (%)			5.3 (3)	N/A	.15
None	107 (53.5)	109 (54.5)			
Short-term	39 (19.5)	49 (24.5)			
Long-term	27 (13.5)	28 (14)			
More than one	27 (13.5)	14 (7)			
Assessment, n (%)			10.7 (3)	N/A	.01
Positive	72 (36)	59 (29.5)			
Neutral	23 (11.5)	33 (16.5)			
Negative	20 (10)	7 (3.5)			
Others					
Fear of stigma, n (%)			4.5 (3)	N/A	.21
Yes or rather yes	127 (63.5)	121 (60.5)			
No	73 (36.5)	79 (39.5)			

[a]N/A: not applicable.

[b]PTSD: posttraumatic stress disorder.

[c]OCD: obsessive-compulsive disorder.

[d]ADD: attention deficit disorder.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]WHOQOL-BREF: World Health Organization Quality of Life-abbreviated version.

[g]RSE: Rosenberg Self-esteem Scale.

[h]APOI: Attitude Toward Psychological Online Interventions.

[i]PATHEV: Patient Questionnaire on Therapy Expectation and Evaluation.

Results

Sample Characteristics

A total of 400 participants (intervention group: 200 and wait-list control group: 200) were included in the analyses. Table 1 shows the demographic and psychopathological data of the participants at baseline.

The overall sample had an average age of 22.98 years (SD 3.36) and consisted of 10.8% (43/400) men and 88.5% (354/400) women. In addition, 0.8% (3/400) of participants stated diverse as their gender. The average PHQ-9 score was 11.1 (SD 4.71; moderate symptoms 10-14). Among the participants, 4.3% (17/400) met the criteria for severe depressive symptoms (PHQ-9 score>19), 20.3% (81/400) for moderately severe depressive symptoms (PHQ-9 score=15-19), 35.8% (143/400) for moderate depressive symptoms (PHQ-9 score=10-14), 32.3% (129/400) for mild depressive symptoms (PHQ-9 score=5-9), and 7.5% (30/400) for minimal depressive symptoms (PHQ-9 score=1-4). In addition, 46% (184/400) of participants stated that they had received psychotherapeutic treatment at least once.

The randomization was successful (Table 1). There were no significant differences between the groups in terms of age and gender or in primary and secondary outcome parameters (depressive symptoms, self-esteem, and quality of life). There were also no significant differences between the groups in terms of expected treatment outcomes and attitudes toward internet- and mobile-based interventions. However, the intervention group showed a significantly higher number of participants with an alcohol and drug addiction (intervention group: n=7; wait list control group: n=1) as well as a personality disorder (intervention group: n=9; wait list control group: n=1) and a significantly lower number of participants with an obsessive-compulsive disorder (intervention group: n=5; wait list control group: n=17). In addition, participants in the wait-list control group reported a neutral experience with psychotherapy more often (wait list control group: n=33; intervention group: n=23) and a positive experience (wait list control group: n=59; intervention group: n=72) and negative experience (wait list control group: n=7; intervention group: n=20) less often (Table 1).

Within-Group Differences

The results of paired samples *t* tests showed a significant reduction in depressive symptoms, both in the intervention group (t_{89}=4.88; *P*<.001; *d*=−0.38) and in the wait-list control group (t_{134}=2.7; *P*=.007; *d*=−0.21) from t0 to t1 (Table 2).

Results of paired samples *t* tests also indicated a significant increase in scores on the self-esteem scale (RSE) for the intervention group (t_{89}=−6.47; *P*<.001; *d*=0.38) and the wait-list control group (t_{132}=−3.46; *P*=.001; *d*=0.16). For both groups, the results of the paired samples *t* test did not show a significant increase in quality of life (WHOQOL-BREF).

Table 2. Outcome measures at each assessment time for per-protocol sample (used program at least once a week; n=225).

Measurements	Intervention group (n=90)				Wait-list control group (n=135)			
Questionnaires	Pre, mean (SD)	Post, mean (SD)	Cohen *d* (95% CI)	*P* value	Pre, mean (SD)	Post, mean (SD)	Cohen *d* (95% CI)	*P* value
PHQ-9[a]	11.27 (5.03)	9.30 (5.22)[b]	−0.38 (−0.8 to 0.03)	<.001	11.10 (4.42)	10.17 (4.32)[c]	−0.21 (−0.55 to 0.13)	.007
RSE[b]	25.56 (6.41)	28.00 (6.44)[b]	0.38 (−0.04 to 0.80)	<.001	26.59 (6.02)	27.57 (6.43)[b]	0.16 (−0.18 to 0.50)	.001
WHOQOL-BREF[c]	3.74 (0.80)	3.86 (0.82)	0.15 (−0.27 to 0.65)	.12	3.67 (0.76)	3.76 (0.72)	0.12 (−0.22 to 0.46)	.19

[a]PHQ-9: Patient Health Questionnaire-9.

[b]RSE: Rosenberg Self-esteem Scale.

[c]WHOQOL-BREF: World Health Organization Quality of Life-abbreviated version.

Between-Group Differences

For the primary outcome parameter (depressive symptoms, PHQ-9), the results of the ANCOVA were significant for ITT ($F_{1,398}$=3.94; *P*=.048), PP ($F_{1,223}$=3.98; *P*=.047), and CC ($F_{1,261}$=4.60; *P*=.03; Table 3).

For reducing depressive symptoms, a small effect size of η_p^2=0.018 (*d*=0.26) was found in the PP sample. Furthermore, the results of the ANCOVA showed statistical significance for the secondary outcome parameter self-esteem (RSE) for ITT ($F_{1,398}$=6.80; *P*=.009), PP ($F_{1,221}$=8.79; *P*=.003), and CC ($F_{1,259}$=7.26; *P*=.008). The analyses resulted in a small to medium effect size for the increase in self-esteem (η_p^2=0.038; *d*=0.40) in the PP sample across time. There was no significant improvement across time in quality of life (WHOQOL-BREF), as analyzed using an ANCOVA with baseline score as covariate in any of the samples: (ITT: $F_{1,398}$=0.56; *P*=.46; PP: $F_{1,223}$=0.41; *P*=.52; and CC: $F_{1,261}$=0.81; *P*=.37).

Table 3. Analysis of covariances with respective baseline values as covariates.

Measurements	CC[a] (n=263)			PP[b] (n=225)			ITT[c] (n=400)		
	F test (*df*)	*P* value	η_p^2	*F* test (*df*)	*P* value	η_p^2	*F* test (*df*)	*P* value	η_p^2
PHQ-9[d]	4.60 (1,398)	.03	0.017	3.98 (1,222)	.047	0.018	3.94 (1,261)	.048	0.010
RSE[e]	7.26 (1,398)	.008	0.027	8.79 (1,220)	.003	0.038	6.80 (1,259)	.009	0.017
WHOQOL-BREF[f]	0.81 (1,398)	.37	0.003	0.41 (1,223)	.52	0.002	0.56 (1,261)	.47	0.001

[a]CC: complete-cases.

[b]PP: per-protocol.

[c]ITT: intention-to-treat.

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

[e]RSE: Rosenberg Self-esteem Scale.

[f]WHOQOL-BREF: World Health Organization Quality of Life-abbreviated version.

Study Completion and App Use

Out of 400 participants, 263 (65.8%) completed the postassessment, 128 (64%) in the intervention group and 135 (67.5%) in the wait-list control group. Regarding study completion, there was no difference between the groups (χ^2_1=0.5; *P*=.46). Furthermore, participants who completed the study differed only in terms of their treatment expectations. Participants who completed the study expected a more positive treatment outcome (t0) than participants who did not complete the study (t_{398}=−2.12; *P*=.04).

In the intervention group, 60% (120/400) of participants reported how often they used the self-help smartphone app during the intervention period (completed the daily exercise). The self-help smartphone app was used by 23.3% (28/400) of participants daily, by 17.5% (21/400) of participants 4-6 times a week, by 25% (30/400) of participants 2-3 times a week, by 10% (12/400) of participants once a week, by 19.2% (23/400) of participants 1-3 times in total, and by 5% (6/400) of participants not at all. The improvement in symptoms (PHQ-9) did not correlate with use frequency (*r*=0.020; *P*=.83). However, use frequency correlated with the expected treatment outcomes (*r*=0.236; *P*=.009) and attitude toward internet- and mobile-based interventions (*r*=0.260; *P*=.004; 1=not at all to 6=daily). The more positive the attitude toward internet- and mobile-based interventions and the more positive the expected treatment outcomes, the more often the self-help smartphone app was used.

Attitude and Expectation

In total, of the 400 participants, 232 (58%) had a positive attitude toward internet- and mobile-based interventions, 30 (7.5%) had a neutral attitude toward internet- and mobile-based

interventions, and 138 (34.5%) had a negative attitude toward internet- and mobile-based interventions. Although 87.3% (349/400) of participants believed that internet- and mobile-based interventions are therapeutically effective (optimism regarding personal therapeutic goal clarification, emotional expectation of helpful efficacy, expectation of learning new skills, cognitive acceptance of the methodology), 31.8% (127/400) of participants also stated that they were skeptical about internet- and mobile-based interventions and that they perceived risks (regarding professionalism, side effects, and feasibility). In addition, 67.8% (271/400) of participants perceived difficulties caused by automation (poor crisis management, low learning success, poorer cognitive understanding of therapy contents, and lower motivation because of lack of personal contact). The advantages of anonymity were reported by 45.5% (182/400) of participants (increased discretion, personally increased self-autonomy, reduction of self-stigmatization, and stigmatization by other persons).

Of the 400 participants, 257 (64.3%) indicated a positive expectation and 117 (29.3%) indicated a negative expectation of treatment outcome regarding the self-help smartphone app. Approximately half of the participants (191/400, 47.8%) did not expect the self-help smartphone app to reduce their symptoms and indicated that this is not the right program for them (172/400, 43%). Only a few participants (38/400, 9.5%) were afraid of change as a result of the self-help smartphone app. The effectiveness of the app could not be predicted by attitudes toward internet- and mobile-based interventions (β=−.006; t_{275}=−0.17; P=.87; R^2<0.001; $F_{1,275}$=0.03; P=.87).

Side Effects

Of the 119 participants (intervention group) who completed the questionnaire on side effects (INEP), 51 (42.9%) reported at least one positive side effect. The most commonly reported positive side effect was that participants felt better when using the self-help smartphone app (43/119, 36.1%). Furthermore, 17.6% (21/119) of participants stated that they experienced less pain from events from the past, 14.3% (17/119) stated that they experienced fewer conflicts in their partnership, 13.6% (16/119) stated that they had a better relationship with their friends, 10.9% (13/119) stated that they had a better relationship with their family, and 8.4% (10/119) stated that trusting others was easier for them.

Overall, of the 119 participants 27 (22.7%) reported a negative side effect. Fear of stigmatization was the most common negative side effect (12/119, 10.1%). In addition, 6.7% (8/119) of participants reported that they had longer phases in which they felt bad, 5.9% (7/119) of participants reported that they had problems with insurance, 5% (6/119) of participants reported that they experienced more pain from events from the past, 1.7% (2/119) of participants reported that they felt worse, 1.7% (2/119) of participants reported that they were more concerned about financial issues, 0.8% (1/119) of participants reported that trusting others is more difficult for them, 0.8% (1/119) of participants reported that they had a worse relationship with their family, and 0.8% (1/119) of participants reported that they had a worse relationship with their friends. None of the participants stated that they had changed as a person to the negative, that they had suicidal thoughts or intentions for the first time, or that they experienced more conflicts in their partnership.

The most intense change was observed in the improvement (positive side effect; mean 2.62, SD 0.65) and deterioration (negative side effect; mean 3.00, SD 0) of the relationship with the families. In comparison with the negative side effects, positive side effects were mentioned 3 times more frequently (negative: 41/158, 25.9% and positive: 120/158, 75.9%).

Moderation Analysis

The results of the interaction effect of the explorative moderation analysis are shown in Table 4.

The analysis revealed that participants in the intervention group who had a higher expectation of treatment outcome (P=.02; PATHEV total score) and more hope (P=.049; PATHEV hope scale) showed a higher improvement in depressive symptoms (PHQ-9) than the wait-list control group. In addition, participants in the intervention group who were more worried that the app would not help them (P=.03) and participants in the intervention group who stated a higher reduced or excessive need to eat (P=.02) showed a less improved outcome (PHQ-9) than the wait-list control group.

Table 4. Moderators for Patient Health Questionnaire-9 improvement (dependent variable: Patient Health Questionnaire-9 total difference scores and independent variable: group, means are centered); results of per-protocol sample (N=225).

Moderator	B[a] (SE)	*t* test (*df*)	*P* value	LLCI[b]	ULCI[c]	*P* value[d] for −1 SD	*P* value[d] for 0	*P* value[d] for +1 SD
PATHEV[e] total scale	0.218 (0.090)	2.428 (117)	.02	0.041	0.396	.65	.04	.002
PATHEV hope scale	0.321 (0.162)	1.982 (117)	.049	0.002	0.641	.90	.07	.006
PATHEV item 1	−1.023 (0.464)	−2.204 (117)	.03	−1.937	−0.108	.004	.14	.70
PHQ-9[f] item 5	−1.058 (0.466)	−2.272 (223)	.02	−1.975	−0.140	.003	.01	.36

[a]B: interaction coefficient.

[b]LLCI: lower limit confidence interval.

[c]ULCI: upper limit confidence interval.

[d]The last 3 columns present the simple slopes.

[e]PATHEV: Patient Questionnaire on Therapy Expectation and Evaluation.

[f]PHQ-9: Patient Health Questionnaire-9.

Subjective Appraisal

In the intervention group, 119 participants completed the questionnaire on patient satisfaction (ZUF-8). Table 5 shows the users' subjective appraisal for each item.

The average total score was mean 20.28 (SD 5.36; 8=very dissatisfied to 32=very satisfied). The quality of the self-help smartphone app was rated positively by 64.7% (77/119) of participants. For each item, the positive evaluations outweighed the negative evaluations. In addition to ZUF-8, 3 further questions were asked regarding participant satisfaction. The majority of participants found the language 84.9% (101/119), text length 77.3% (92/119), and number of exercises 49.6% (59/119) in the self-help smartphone app to be just right.

Table 5. Subjective appraisal using *Fragebogen zur Patientenzufriedenheit* of *MCT & More* (n=119).

ZUF-8[a] item	Mean (SD)	Positive[b], n (%)
1. How do you rate the quality of the program? (excellent, good vs not that good, or not good)[c]	2.29 (0.61)	77 (64.7)
2. Did you receive the type of treatment you expected to receive? (absolutely, a lot vs a little, or not at all)	2.45 (0.78)	59 (49.6)
3. To what extent did the program help you cope with your problems? (absolutely, a lot vs a little, or not at all)[c]	2.63 (0.79)	56 (47.1)
4. Would you recommend the program to a friend with similar symptoms? (yes, probably yes vs probably not, or no)	2.61 (0.92)	52 (43.7)
5. How happy are you about the extent of the help you have received through using the program? (very satisfied, mostly satisfied vs somewhat dissatisfied, or dissatisfied)	2.50 (0.82)	66 (55.5)
6. Did the program help you to cope with your problems more successfully? (absolutely, a lot vs a little, or not at all)[c]	2.39 (0.63)	66 (55.5)
7. How satisfied are you with the program in general? (very satisfied, mostly satisfied vs somewhat unsatisfied, or unsatisfied)[c]	2.42 (0.85)	67 (56.3)
8. Would you use the program again? (Yes, probably yes vs probably not, or no)	2.44 (0.92)	59 (49.6)

[a]ZUF-8: Fragebogen zur Patientenzufriedenheit (German version of the Client Satisfaction Questionnaire-8).

[b]4-point rating scale: 1 and 2 were rated as negative and 3 and 4 as positive. In the table it is stated how often the question has been answered positively (rated 3 or 4).

[c]A lower score indicates a more positive response (inverted scores).

Discussion

Principal Findings

The study demonstrates the effectiveness of the self-help smartphone app *MCT & More* in students with depressive symptoms. As expected, the app led to a significant reduction in depressive symptoms and a significant increase in self-esteem in the intervention period of 4 weeks. In our study, a small effect size of d=0.26 (PHQ-9; PP sample) in reducing depressive symptoms was found, which is comparable with the effect size found in a meta-analysis of smartphone apps for depressive symptoms (g=0.22) [34] and is slightly higher than the effect size reported in a more recent meta-analysis in which the sample consisted of university students with depressive symptoms (g=0.18) [20]. In addition, a small to medium effect size of

d=0.40 (RSE; PP sample) was found for the increase in self-esteem, which corresponds to the findings of a study on a self-help web-based intervention, which was mainly focused on the treatment of depressive symptoms [61]. Contrary to our expectations, the use of the self-help smartphone app did not lead to a significant increase in quality of life (WHOQOL-BREF). The WHOQOL-BREF defines quality of life as the individual perception of one's own life situation [41]. It is possible that an improvement in the life situation only occurs after a longer period and would have been shown in follow-up examinations [61]. The quality of life was assessed using a single global item. It is possible that improvements would have been found on the subscales of the WHOQOL-BREF (eg, psychological quality of life and social relationships). In the intervention group and the wait-list control group, a significant reduction in depressive symptoms and a significant increase in self-esteem were observed after the intervention period of 4 weeks. Despite the improvements in the wait-list control group, significant group differences were found (significantly higher improvement in the intervention group). The improvement in the wait-list control group may be because of changes in external circumstances, the use of other services, or spontaneous remissions [62]. As the survey took place during the COVID-19 pandemic, depressive symptoms among students may have been more severe than usual [63]. In addition, the quality of life may have been reduced [64].

Adherence and Acceptance

Most participants used the self-help smartphone app regularly (91/129, 75.8%) at least once a week; self-assessment) and completed the study (263/400, 65.8%). Participants who completed the postassessment expected a more positive treatment outcome at the baseline assessment than the participants who dropped out of the study (t_{398}=−2.12; *P*=.04). This result was also found by Mira et al [36] and underlines the relevance of the expected treatment outcome at the beginning of the intervention for participants' study adherence. The reduction in symptoms (PHQ-9) did not correlate with the frequency of use of the self-help smartphone app. Other researchers who investigated the relationship between frequency of use and symptom reduction also concluded that using apps with a medium frequency only leads to a little additional benefit than using apps with low frequency [65]. It should be considered that the assumption of a linear relationship between frequency of use and symptom reduction might be too simplistic and that further variables need to be evaluated to better understand the relationship. A possible explanation may also be that, because the intervention was not linear or sequential, participants may have been more likely to use only the parts of the intervention they needed. For some, a low dose may have been sufficient for symptom improvement, whereas other users may have required to use the program more often.

Attitudes and Expectations

The students' overall expectation of treatment outcomes and their attitude toward internet- and mobile-based interventions was moderate. Almost half of the students (192/400, 48%) did not expect any improvement from the app, and about one-third of the students (140/400, 35%) had a negative attitude toward internet- and mobile-based interventions. These findings are consistent with the results of another German study, in which 41% of the respondents (patients with depressive symptoms in primary care) indicated a low acceptance of internet- and mobile-based interventions for depression [66]. However, other studies conducted in Germany found a more positive attitude toward internet- and mobile-based interventions for depressive symptoms (total scale: mean 55.86, baseline total scale intervention group: mean 56.13, and baseline total scale control group: mean 55.59) [67,68]. It is possible that the attitude toward internet- and mobile-based interventions was lower in this study (mean 50.09; higher scores indicate a more positive attitude) because of the young age of the sample. In Germany, younger people report a more negative attitude toward internet- and mobile-based interventions than older people [69]. The attitude could possibly be improved by showing information videos that address potential barriers of acceptance (eg, low expectations regarding efficacy and worries about data security) before use [66]. Furthermore, the more positive the attitude toward internet- and mobile-based interventions and the more positive the expectations of the treatment outcome, the more frequently the self-help smartphone app was used, which was also found in other studies [70,71].

Most of the students trusted the therapeutic efficacy (328/400, 82%), which seems contradictory at first, as about half of the participants (192/400, 48%) did not expect any improvement. It is possible that the participants were generally convinced of the efficacy of internet- and mobile-based interventions (APOI queries a general attitude) but did not expect any improvement for themselves (PATHEV refers to their own symptomatology). Individuals with depressive symptoms often do not believe their symptoms will improve because they often express feelings of hopelessness [72].

The moderation analysis revealed that participants in the intervention group who indicated a higher expectation of treatment outcome (PATHEV total score) and more hope (PATHEV scale hope) achieved a higher reduction in depressive symptoms (PHQ-9) than those in the wait-list control group. The results of the linear regression showed that the attitude toward internet- and mobile-based interventions did not allow a prediction of the effectiveness. In addition, participants who were more worried that the app would not help them (PATHEV, item 5) showed less improvement in symptoms (PHQ-9). These findings are in contrast with the results of Lüdtke et al [37], who found no impact (moderation analysis) of expected treatment outcome (University of Rhode Island Change Assessment) on symptom reduction. This could be because of the use of different measurements. However, Schröder et al [68] found that participants with a more positive attitude (APOI) at the beginning of an internet intervention experienced a stronger reduction in symptoms than participants with a more negative attitude. Another recent study on the effectiveness of a self-help smartphone app for depression showed that the expected treatment outcome was a predictor of symptom reduction [36].

Subjective Appraisal

The majority of the participants were satisfied with the quality of the self-help smartphone app (77/119, 64.7%) and the extent

of help (66/119, 55.5%). Another study evaluating a cognitive behavioral therapy–based self-help web-based program among students with depressive symptoms also found a moderate overall satisfaction (study by Santucci et al [73]: total satisfaction score (ZUF-8): mean 21.70 (SD 5.20) and total satisfaction score (ZUF-8) of this study: mean 20.28 (SD 5.36)). In a pilot study of the self-help smartphone app *MCT & More*, the participants reported a higher overall satisfaction (eg, 88.5% of the participants were satisfied with the quality of the self-help smartphone app [37]). This could be because of the different average ages of the samples (pilot study: 43 years and this study: 23 years). Older age is associated with greater intervention effects [74] and more positive attitudes toward internet-based interventions [69], which may lead to higher satisfaction.

Strength and Limitations

No psychiatric diagnosis was required to participate in the study, and participation was possible even with mild depressive symptoms, which led to a heterogeneity of depression levels. This has the advantage that a wider range of individuals with a desire for treatment was reached (regardless of whether they fulfilled the criteria of a diagnosis). On the other hand, it has been shown that individuals with severe depressive symptoms benefit more from low-threshold psychological interventions than mildly depressed individuals [75]. In contrast, a meta-analysis showed that self-guided internet-based interventions are effective regardless of symptom severity [17]. Furthermore, treatment adherence was rather high (91/120, 75.8% used the app at least once a week), which allowed the potential of the app to be well exploited. As the study was conducted on the web, the data collected were based on self-assessments of the participants. Therefore, it could not be eliminated that socially desirable or dishonest statements were made that could have distorted the results. In addition, despite the integrated control questions on studying, it could not be completely prevented that individuals who were not enrolled at a German university also took part in the study. An unambiguous verification of the student status (eg, via enrolment certificates) was not possible because of data privacy reasons. There was no structural equality between the sample and the general population (students in Germany) regarding gender and subject groups [13,76]. The higher proportion of women could be because women are more often affected by depression than men (women: 10.8% and men: 7.6%) and that this difference is particularly evident in young adulthood [77,78]. Furthermore, the study showed baseline differences regarding some comorbid self-reported diagnoses and the evaluation of previous therapy experiences. Nevertheless, randomization was largely considered successful. Follow-up investigations were not possible because of the time frame of the study. For this reason, no conclusions can be drawn regarding the medium- or long-term effects of the self-help smartphone app.

Conclusions

The effectiveness of the self-help smartphone app *MCT & More* was demonstrated in students with depressive symptoms, although the overall outcome expectation and attitude toward internet- and mobile-based interventions were only moderate. Despite the improvements in the wait-list control group, significant group differences were found. The use of the app led to a significantly higher reduction in depressive symptoms (d=0.26) and a significantly higher increase in self-esteem (d=0.40). The expected treatment outcome and the attitude toward internet- and mobile-based interventions were correlated with the frequency of use. The more positive the attitude and the more positive the result expectation, the more frequently the self-help smartphone app was used. Participants who indicated a higher expectation of treatment outcome and more hope achieved a higher reduction in depressive symptoms. In addition, participants who were more worried that the app would not help them and participants who stated a higher reduced or excessive need to eat showed less improvement in symptoms.

Future studies should investigate further variables (with respect to personal characteristics and app features) that positively influence the effectiveness of identifying ways of increasing efficacy. To make self-help smartphone apps as target group–specific as possible, further subgroups should be identified for which a particularly high or low effectiveness is shown. In addition, follow-up studies are required to determine the long-term effects. It should be investigated how attitudes toward internet- and mobile-based interventions and the expected treatment outcome can be improved to establish effective self-help smartphone apps as low-threshold offers at universities and to promote treatment adherence. The self-help smartphone app could be used regularly at German universities as a low-threshold program to enhance students' health.

Conflicts of Interest

The authors developed the app MCT & More.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 479 KB - mhealth_v9i7e26498_app1.pdf]

References

1. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
2. Grützmacher J, Gusy B, Lesener T, Sudheimer S, Willige J. Gesundheit Studierender in Deutschland 2017. Ein Kooperationsprojekt zwischen dem Deutschen Zentrum für Hochschul- und Wissenschaftsforschung, der Freien Universität

Berlin und der Techniker Krankenkasse. 2018. URL: https://www.tk.de/resource/blob/2050660/8bd39eab37ee133a2ec47e55e544abe7/gesundheit-studierender-in-deutschland-2017-studienband-data.pdf [accessed 2021-06-09]
3. Richardson T. Is student depression fuelling dropout rates? Trinity News. 2008. URL: https://researchportal.bath.ac.uk/en/publications/is-student-depression-fuelling-dropout-rates [accessed 2021-06-23]
4. Middendorff E, Apolinarski B, Becker K, Bornkessel P, Brandt T, Heißenberg S, et al. Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2016. Bundesministerium für Bildung und Forschung (BMBF). Berlin; 2017. URL: https://www.bmbf.de/upload_filestore/pub/21._Sozialerhebung_2016_Zusammenfassung.pdf [accessed 2021-06-09]
5. Meridew C, Athey A, Killgore W, Gehrels J, Alfonso-Miller P, Grandner M. 0190 Academic performance associated with sleep duration among student athletes: impact of insomnia, fatigue, and depression. Sleep 2018;41(suppl_1):74-75. [doi: 10.1093/sleep/zsy061.189]
6. Mihăilescu A, Diaconescu L, Ciobanu A, Donisan T, Mihailescu C. The impact of anxiety and depression on academic performance in undergraduate medical students. Eur Psychiatry 2020 Mar 23;33(S1):s284. [doi: 10.1016/j.eurpsy.2016.01.761]
7. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
8. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
9. Wallin E, Maathz P, Parling T, Hursti T. Self-stigma and the intention to seek psychological help online compared to face-to-face. J Clin Psychol 2018 Jul;74(7):1207-1218. [doi: 10.1002/jclp.22583] [Medline: 29315545]
10. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
11. Carey M, Jones K, Meadows G, Sanson-Fisher R, D'Este C, Inder K, et al. Accuracy of general practitioner unassisted detection of depression. Aust N Z J Psychiatry 2014 Jan 10;48(6):571-578 [FREE Full text] [doi: 10.1177/0004867413520047] [Medline: 24413807]
12. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
13. Deutsche Studierende und Studienanfänger/-innen im 1. Hochschulsemester nach Alter, Hochschularten und Geschlecht. Statistisches Bundesamt. 2019. URL: https://www.datenportal.bmbf.de/portal/de/K254.html [accessed 2021-06-09]
14. Loecherer R, Apolinário-Hagen J. Wirksamkeit und Akzeptanz von webbasierten Selbsthilfeprogrammen zur Förderung psychischer Gesundheit und zur Stressbewältigung im Studiumin Scoping-Review der aktuellen Forschungsliteratur. e-beratungsjournal. 2017. URL: https://www.e-beratungsjournal.net/ausgabe_0117/Loecherer_Apolinario-Hagen.pdf [accessed 2021-06-09]
15. Moessner M, Bauer S. E-mental-health und internetbasierte psychotherapie. Psychotherapeut 2017 May 5;62(3):251-266. [doi: 10.1007/s00278-017-0198-4]
16. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
17. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
18. Sander L, Rausch L, Baumeister H. Effectiveness of Internet-Based Interventions for the Prevention of Mental Disorders: A Systematic Review and Meta-Analysis. JMIR Ment Health 2016 Aug 17;3(3):e38. [doi: 10.2196/mental.6061] [Medline: 27535468]
19. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
20. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759. [doi: 10.1002/mpr.1759] [Medline: 30585363]
21. Bennett SD, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, et al. Practitioner Review: Unguided and guided self-help interventions for common mental health disorders in children and adolescents: a systematic review and meta-analysis. J Child Psychol Psychiatry 2019 Aug;60(8):828-847. [doi: 10.1111/jcpp.13010] [Medline: 30775782]
22. Klein JP, Gerlinger G, Knaevelsrud C, Bohus M, Meisenzahl E, Kersting A, et al. [Internet-based interventions in the treatment of mental disorders : overview, quality criteria, perspectives]. Nervenarzt 2016 Nov 20;87(11):1185-1193. [doi: 10.1007/s00115-016-0217-7] [Medline: 27649987]

23. Hintz S, Frazier PA, Meredith L. Evaluating an online stress management intervention for college students. J Couns Psychol 2015 Apr;62(2):137-147. [doi: 10.1037/cou0000014] [Medline: 24635586]
24. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: the role of therapist responsiveness. Couns Psychother Res 2013 Sep;13(3):184-193. [doi: 10.1080/14733145.2012.733715]
25. Ataş AH, Çelik B. Smartphone use of university students: patterns, purposes, and situations. Malays Online J Educ Technol 2019 Apr 01;7(2):54-70. [doi: 10.17220/mojet.2019.02.004]
26. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838. [doi: 10.3390/ijerph15122838] [Medline: 30551555]
27. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app biobase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
28. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
29. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
30. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
31. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013;8(7):e62873 [FREE Full text] [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]
32. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. Br Med J 2019 Mar 20;364:l920 [FREE Full text] [doi: 10.1136/bmj.l920] [Medline: 30894349]
33. Terhorst Y, Rathner E, Baumeister H, Sander L. Hilfe aus dem App-Store?: Eine systematische übersichtsarbeit und evaluation von apps zur anwendung bei depressionen. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
34. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
35. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
36. Mira A, Soler C, Alda M, Baños R, Castilla D, Castro A, et al. Exploring the relationship between the acceptability of an internet-based intervention for depression in primary care and clinical outcomes: secondary analysis of a randomized controlled trial. Front Psychiatry 2019;10:325 [FREE Full text] [doi: 10.3389/fpsyt.2019.00325] [Medline: 31133899]
37. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening psychischer Störungen mit dem “Gesundheitsfragebogen für Patienten (PHQ-D)”. Diagnostica 2004 Oct;50(4):171-181. [doi: 10.1026/0012-1924.50.4.171]
40. Rosenberg M. Society and the Adolescent Self-Image. Princeton, New Jersey: Princeton University Press; 1965:1-340.
41. Harper A, Power M, WHOQOL-Group X. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
42. Fu TS, Tuan Y, Yen M, Wu W, Huang C, Chen W, et al. Psychometric properties of the World Health Organization Quality Of Life Assessment-Brief in methadone patients: a validation study in northern Taiwan. Harm Reduct J 2013 Dec 10;10(1):37 [FREE Full text] [doi: 10.1186/1477-7517-10-37] [Medline: 24325611]
43. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
44. Ilić I, Šipetić S, Grujičić J, Mačužić I, Kocić S, Ilić M. Psychometric Properties of the World Health Organization's Quality of Life (WHOQOL-BREF) questionnaire in medical students. Medicina (Kaunas) 2019 Dec 04;55(12):772 [FREE Full text] [doi: 10.3390/medicina55120772] [Medline: 31817180]

45. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring Attitudes towards Psychological Online Interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
46. Schulte D. Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV). Zeitschrift für Klinische Psychologie und Psychotherapie 2005 Jun;34(3):176-187. [doi: 10.1026/1616-3443.34.3.176]
47. Juergen S, Nübling R. ZUF-8. Fragebogen zur Messung der Patientenzufriedenheit. Diagnostische Verfahren in der Psychotherapie. Göttingen, Germany: Hogrefe; 2002. URL: https://www.researchgate.net/publication/247161440_ZUF-8_Fragebogen_zur_Messung_der_Patientenzufriedenheit [accessed 2021-06-23]
48. Kriz D, Nübling R, Steffanowski A, Rieger J, Schmidt J. atientenzufriedenheit in der stationären Rehabilitation:Psychometrische Reanalyse des ZUF-8 auf der Basismultizentrischer Stichproben verschiedener Indikation. Z Med Psychol. 2008. URL: https://www.academia.edu/17331948/Patientenzufriedenheit_in_der_station%C3%A4ren_Rehabilitation_Psychometrische_Reanalyse_des_ZUF_8_auf_der_Basis_multizentrischer_Stichproben_verschiedener_Indikation_Patients_satisfaction_in_inpatient_rehabilitation_Psychometrical_evaluation_of_the_ZUF_8_based_on_a_multicenter_sample_of_different_i_ [accessed 2021-06-09]
49. Ladwig I, Rief W, Nestoriuc Y. Welche Risiken und Nebenwirkungen hat Psychotherapie? - Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). Verhaltenstherapie 2014 Oct 30;24(4):252-263. [doi: 10.1159/000367928]
50. Moritz S, Woodward TS. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. Curr Opin Psychiatry 2007 Nov;20(6):619-625. [doi: 10.1097/YCO.0b013e3282f0b8ed] [Medline: 17921766]
51. Beck AT. Thinking and depression. II. Theory and therapy. Arch Gen Psychiatry 1964 Jun 01;10(6):561-571. [doi: 10.1001/archpsyc.1964.01720240015003] [Medline: 14159256]
52. Ellis A. Rational psychotherapy and individual psychology. J Individ Psychol 1957;13(1):38-44 [FREE Full text]
53. Hayes S, Strosahl K, Wilson K. Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change. New York, United States: The Guilford Press; 1999:1-304.
54. Segal Z, Williams J, Teasdale J, Kabat-Zinn J. Die achtsamkeitsbasierte kognitive Therapie der Depression in neuer Ansatz zur Rückfallprävention. 2., erweiterte und überarbeitete Auflage mit MP3-CD Ed. Tübingen: dgvt-Verlag; 2015:A.
55. Jelinek L, Otte C, Arlt S, Hauschildt M. Denkverzerrungen erkennen und korrigieren: Eine Machbarkeitsstudie zum Metakognitiven Training bei Depression (D-MKT). Zeitschrift für Psychiatrie, Psychologie und Psychotherapie 2013 Oct;61(4):247-254. [doi: 10.1024/1661-4747/a000167]
56. Normann N, Morina N. The efficacy of metacognitive therapy: a systematic review and meta-analysis. Front Psychol 2018 Nov 14;9:2211 [FREE Full text] [doi: 10.3389/fpsyg.2018.02211] [Medline: 30487770]
57. Normann N, van Emmerik AA, Morina N. The efficacy of metacognitive therapy for anxiety and depression: a meta-analytic review. Depress Anxiety 2014 May;31(5):402-411. [doi: 10.1002/da.22273] [Medline: 24756930]
58. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018 Apr 27;97(4):380-387 [FREE Full text] [doi: 10.1111/aogs.13309] [Medline: 29377058]
59. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
60. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis : A Regression-Based Approach. New York, United States: Guilford Press; 2013:1-507.
61. Crisp D, Griffiths K, Mackinnon A, Bennett K, Christensen H. An online intervention for reducing depressive symptoms: secondary benefits for self-esteem, empowerment and quality of life. Psychiatry Res 2014 Apr 30;216(1):60-66. [doi: 10.1016/j.psychres.2014.01.041] [Medline: 24534125]
62. Bryman A. Social Research Methods, 4th Edition. Oxfordshire, United Kingdom: Oxford University Press; 2012:1-766.
63. Bäuerle A, Teufel M, Musche V, Weismüller B, Kohler H, Hetkamp M, et al. Increased generalized anxiety, depression and distress during the COVID-19 pandemic: a cross-sectional study in Germany. J Public Health (Oxf) 2020 Jul 13:A [FREE Full text] [doi: 10.1093/pubmed/fdaa106] [Medline: 32657323]
64. Riiser K, Helseth S, Haraldstad K, Torbjørnsen A, Richardsen KR. Adolescents' health literacy, health protective measures, and health-related quality of life during the Covid-19 pandemic. PLoS One 2020;15(8):e0238161 [FREE Full text] [doi: 10.1371/journal.pone.0238161] [Medline: 32857806]
65. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
66. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
67. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]

68. Schröder J, Berger T, Meyer B, Lutz W, Späth C, Michel P, et al. Impact and change of attitudes toward internet interventions within a randomized controlled trial on individuals with depression symptoms. Depress Anxiety 2018 May;35(5):421-430. [doi: 10.1002/da.22727] [Medline: 29489038]
69. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
70. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000 Jul 24;160(14):2101-2107. [doi: 10.1001/archinte.160.14.2101] [Medline: 10904452]
71. Hussein Z, Oon SW, Fikry A. Consumer attitude: does it influencing the intention to use mHealth? Procedia Comput Sci 2017;105:340-344. [doi: 10.1016/j.procs.2017.01.231]
72. Wolfersdorf M. Depressionen Verstehen Und Bewältigen. Berlin, Heidelberg: Springer Nature; 2011:1-154.
73. Santucci LC, McHugh RK, Elkins RM, Schechter B, Ross MS, Landa CE, et al. Pilot implementation of computerized cognitive behavioral therapy in a university health setting. Adm Policy Ment Health 2014 Jul;41(4):514-521. [doi: 10.1007/s10488-013-0488-2] [Medline: 23592231]
74. Reins JA, Buntrock C, Zimmermann J, Grund S, Harrer M, Lehr D, et al. Efficacy and Moderators of Internet-Based Interventions in Adults with Subthreshold Depression: An Individual Participant Data Meta-Analysis of Randomized Controlled Trials. Psychother Psychosom 2020 Jun 16:1-13 [FREE Full text] [doi: 10.1159/000507819] [Medline: 32544912]
75. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. Br Med J 2013 Feb 26;346:f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
76. Studierende und Studienanfänger/-innen im 1. Statistisches Bundesamt. 2020. URL: https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00056740/2110410207004a.pdf [accessed 2021-06-09]
77. Hapke U, Cohrdes C, Nübel J. Depressive Symptomatik im europäischen Vergleich – Ergebnisse des European Health Interview Survey (EHIS) 2. Robert Koch-Institut 2019:12-17. [doi: 10.25646/6221]
78. Thom J, Kuhnert R, Born S, Hapke U. 12-Monats-Prävalenz der selbstberichteten ärztlich diagnostizierten Depression in Deutschland. Robert Koch Institut 2017:9-15. [doi: 10.17886/RKI-GBE-2017-057]

Abbreviations

ANCOVA: analysis of covariance
APOI: Attitude Toward Psychological Online Interventions
CC: complete-case
CONSORT: Consolidated Standards of Reporting Trials
INEP: Inventory for Assessing Negative Effects of Psychotherapy
ITT: intention-to-treat
MCT: metacognitive training
PATHEV: Patient Questionnaire on Therapy Expectation and Evaluation
PHQ-9: Patient Health Questionnaire-9
PP: per-protocol
RSE: Rosenberg Self-esteem Scale
WHOQOL-BREF: World Health Organization Quality of Life-abbreviated version
ZUF-8: Fragebogen zur Patientenzufriedenheit (German version of the Client Satisfaction Questionnaire-8)

Edited by L Buis; submitted 18.12.20; peer-reviewed by P Tonn, A González-Robles, L Farrer; comments to author 16.02.21; revised version received 22.02.21; accepted 15.04.21; published 12.07.21.

Please cite as:
Bruhns A, Lüdtke T, Moritz S, Bücker L
A Mobile-Based Intervention to Increase Self-esteem in Students With Depressive Symptoms: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(7):e26498
URL: https://mhealth.jmir.org/2021/7/e26498
doi:10.2196/26498
PMID:34255711

Original Paper

A Mobile Phone–Based Life-Skills Training Program for Substance Use Prevention Among Adolescents: Cluster-Randomized Controlled Trial

Severin Haug[1], PhD; Raquel Paz Castro[1], PhD; Andreas Wenger[1], MSc; Michael Patrick Schaub[1], PhD

Swiss Research Institute for Public Health and Addiction at Zurich University, Zurich, Switzerland

Corresponding Author:
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction at Zurich University
Konradstrasse 32
Zurich, 8005
Switzerland
Phone: 41 444481174
Email: severin.haug@isgf.uzh.ch

Abstract

Background: Life skills are abilities for adaptive and positive behavior that enable individuals to deal effectively with the demands and challenges of everyday life. Life-skills training programs conducted within the school curriculum are effective in preventing the onset and escalation of substance use among adolescents. However, their dissemination is impeded due to their large resource requirements. Life-skills training provided via mobile phones may provide a more economic and scalable approach.

Objective: The goal of this study was to test the appropriateness (ie, acceptance, use, and evaluation) and short-term efficacy of a mobile phone–based life-skills training program to prevent substance use among adolescents within a controlled trial.

Methods: The study design was a two-arm, parallel-group, cluster-randomized controlled trial with assessments at baseline and follow-up assessments after 6 and 18 months. This report includes outcomes measured up to the 6-month follow-up. The efficacy of the intervention was tested in comparison to an assessment-only control group. The automated intervention program *SmartCoach* included online feedback and individually tailored text messages provided over 22 weeks. The contents were based on social cognitive theory and addressed self-management skills, social skills, and substance use resistance skills. Linear mixed models and generalized linear mixed models, as well as logistic or linear regressions, were used to investigate changes between baseline and 6-month follow-up in the following outcomes: 30-day prevalence rates of problem drinking, tobacco use, and cannabis use as well as quantity of alcohol use, quantity of cigarettes smoked, cannabis use days, perceived stress, well-being, and social skills.

Results: A total of 1759 students from 89 Swiss secondary and upper secondary school classes were invited to participate in the study. Of these, 1473 (83.7%) students participated in the study; the mean age was 15.4 years (SD 1.0) and 55.2% (813/1473) were female. Follow-up assessments at 6 months were completed by 1233 (83.7%) study participants. On average, program participants responded to half (23.6 out of 50) of the prompted activities. Program evaluations underlined its appropriateness for the target group of secondary school students, with the majority rating the program as helpful and individually tailored. The results concerning the initial effectiveness of this program based on 6-month follow-up data are promising, with three of nine outcomes of the intention-to-treat analyses showing beneficial developments of statistical significance (ie, quantity of alcohol use, quantity of tobacco use, and perceived stress; $P<.05$) and another three outcomes (ie, problem drinking prevalence, cannabis use days, and social skills) showing beneficial developments of borderline significance ($P<.10$).

Conclusions: The results showed good acceptance of this intervention program that could be easily and economically implemented in school classes. Initial results on program efficacy indicate that it might be effective in both preventing or reducing substance use and fostering life skills; however, data from the final 18-month follow-up assessments will be more conclusive.

Trial Registration: ISRCTN Registry ISRCTN41347061; https://doi.org/10.1186/ISRCTN41347061

(JMIR Mhealth Uhealth 2021;9(7):e26951) doi:10.2196/26951

KEYWORDS

life skills; substance use; prevention; adolescents; mobile phone

Introduction

During adolescence numerous biological, psychological, and social transitions take place, which determine a young person's development and future [1,2]. These transitions allow adolescents to develop skills in order to achieve greater autonomy, build relationships with peers, develop a positive body image, and find one's identity. However, they are also accompanied by an increased willingness to take risks during a time when the cognitive functions of the brain (eg, to regulate emotions) are not yet fully developed [3]. Shifts in emotional regulation as well as increased risk behavior increase the susceptibility of an individual to develop mental and substance use disorders. These disorders are largely responsible for the health burden of 10- to 24-year-old individuals [4]. Substance use, as well as the development of substance use disorders, co-occur with mental disorders and typically first arise throughout the adolescent years [1].

The prevalence of lifetime and recent alcohol and tobacco use increases sharply in both genders from 11 to 15 years of age [5]. In Switzerland, the lifetime prevalence of alcohol use increased from 22% among 11-year-old boys to 70% among 15-year-old boys, and from 11% among 11-year-old girls to 69% among 15-year-old girls [6]. The proportion of pupils who reported having smoked cigarettes at least once in their life increased from 6% among 11-year-old boys to 35% among 15-year-old boys, and from 2% among 11-year-old girls to 30% among 15-year-old girls. This age range reflects a critical time when substance prevention programs should be implemented.

A systematic review of studies, which examined the efficacy of prevention, early intervention, and harm reduction in adolescents for tobacco, alcohol, and illicit drugs, illustrated the effectiveness of taxation, public consumption bans, advertising restrictions, and minimum legal age. Additionally, promising effectiveness of preventive interventions, which provide life-skills training in educational settings, was shown [7]. Using schools as a medium to reach adolescents with preventive interventions is particularly suitable, as it facilitates the delivery and access to adolescents within compulsory secondary education. A Cochrane review on school-based programs for the prevention of tobacco smoking demonstrated a significant intervention effect from the combination of social competence and social influence interventions [8]. Another Cochrane review on school-based prevention programs for alcohol misuse among young people concluded that certain generic psychosocial and developmental prevention programs can be effective [9]. A large proportion of generic programs tending to social competencies and social influences, which were referenced in the reviews mentioned above, are defined as life-skills training and are based on Bandura's social learning theory [10]. This theory explains that children and adolescents discover substance use by modeling, imitation, and reinforcement, which is influenced by individual cognitions and attitudes. Moreover, in light of the social influences approach [11], substance use susceptibility increases as a result of a lack of personal and social skills, and adolescents begin drug use because of pressure from friends, family, and the media. According to the definition from the World Health Organization, life skills are "abilities for adaptive and positive behavior that enable individuals to deal effectively with the demands and challenges of everyday life" [11].

Life-skills intervention programs to prevent substance use [12-14] primarily combine training in self-management skills (eg, adapting to stress, emotional self-regulation, and goal setting), social skills (eg, assertiveness and communication skills), and skills facilitating the resistance to substance use (eg, opposing peer pressure to drink alcohol, identifying and resisting media influences that promote cigarette smoking, and correcting normative misperceptions of substance use). In spite of the fact that these life-skills training programs were compelling at preventing the onset [8,14,15] of using an explicit substance or at reducing problematic substance use [9], their implementation and dispersal in schools present genuine difficulties [16]: teachers or other professionals need time, training, knowledge, and skills to prepare and administer such programs [17].

Digital interventions have great potential to overcome the above-mentioned obstacles that hinder successful program implementation and larger-scale dissemination of life-skills training in schools. These programs have a large reach at low cost and offer the ability to deliver uniquely personalized content automatically, which can be accessed anytime and anywhere. Furthermore, digital interventions might be more appealing for adolescents because they can better ensure privacy and tailor contents to their needs. A systematic review of digital alcohol and other drug prevention programs [18] identified nine trials, six of which demonstrated significant, but modest, effects for alcohol and/or other drug use outcomes. The programs were delivered in the United States, Australia, and the Netherlands and provided between 1 and 12 online curriculum-based standard lessons or tailored feedback. All programs were universal (ie, delivered interventions to all students regardless of their level of risk) and were primarily based on principles of the social learning theory [10], the social influences approach [19], and the social cognitive theory [20,21].

A promising way of delivering preventive services, besides conventional personal computers, is to do so remotely by using mobile technologies. Almost all (99%) adolescents between the ages of 12 and 19 years in Switzerland, as in most other developed countries, own a mobile phone. Compared to services that can only be accessed at particular times or places, they provide a targeted and confidential means of intervention delivery [22]. Mobile phone–based interventions can provide almost constant support to users, in comparison to interventions that can only be accessed at specific times or locations, and they provide a discrete and confidential means of intervention delivery [23]. Mobile phone text messaging, in particular, is a suitable means of delivering individually tailored messages via mobile phones. This interactive service allows cost-effective, instantaneous, direct delivery of messages to individuals. Recent reviews underline the potential efficacy of text messaging–based interventions to reduce alcohol and tobacco use for different at-risk target groups, including adolescents and young adults [24,25].

Ready4life is a mobile phone–based life-skills training program for substance use prevention. Its acceptance and potential

effectiveness was tested within a pre-post study in Switzerland [26]. Program participants were vocational school students with a mean age of 17 years, who received up to three weekly text messages over 6 months. The *ready4life* program was based on social cognitive theory and addressed self-management skills, social skills, and substance use resistance skills. Active program engagement was encouraged through interactive features, such as quiz questions, message and picture contests, and integration of a friendly competition with prizes, in which program users collected credits with each interaction. A total of 4 out of 5 eligible students participated in the program and the associated study. Pre-post comparisons, between baseline and follow-up assessments, revealed decreased perceived stress and increases in several life skills that were addressed. The proportion of adolescents with at-risk alcohol use significantly declined from 20% at baseline to 16% at follow-up.

Based on these promising findings [26], a similar universal prevention program was developed for the target group of secondary school students; these students are slightly younger than vocational school students and their substance use is not yet fully advanced [5]. Our main hypothesis concerning the final follow-up at month 18 is that the individually tailored intervention program will be more effective than assessment only in preventing the onset and escalation of problematic alcohol and tobacco use.

This study presents (1) the results on appropriateness (acceptance, use, and evaluation of duration, intensity, tailoring, helpfulness, comprehensibility, etc) of this program as well as (2) initial results on its efficacy considering 6-month follow-up assessments of this controlled trial.

Methods

Objectives and Study Design

This study aimed at testing the acceptance and short-term efficacy of *SmartCoach*, a mobile phone–based life-skills training program to prevent substance use among secondary school students. The efficacy of the intervention was tested in comparison to an assessment-only control group, considering data from the first follow-up assessment after 6 months.

Participants, Setting, and Procedure

We tested the intervention program in secondary and upper secondary school students, typically aged between 14 and 17 years. Secondary schools in the German-speaking part of Switzerland were invited to participate in the study by cooperating regional centers for addiction prevention. Employees of these centers arranged 60-minute information sessions in participating secondary school classes during regular school lessons reserved for health education. These information sessions were led by junior scientists from the Swiss Research Institute for Public Health and Addiction, who were experienced in work with young people, experienced in the provision of preventive interventions, and trained on the study and the program to be delivered.

The parents of secondary school students below the age of 16 years were informed at least one week in advance of this session. They received a letter including information about the study and the intervention program and were asked to give written informed consent regarding their child's participation in the study. Adolescents aged 16 years or older gave their own informed consent.

Within the first half of the information sessions in the school classes, the junior scientists raised awareness about the importance of life skills to effectively cope with the demands and challenges of everyday life. For this purpose, they used video sequences demonstrating typical stressors and demands for this age group (eg, search for an apprenticeship, exam stress, and peer pressure for substance use) and different strategies to cope with them. The importance of emotional regulation skills and social skills to effectively cope with these stressors were discussed based on case vignettes. Subsequently, the students were informed about, and invited to participate in, a study testing innovative channels for the provision of life-skills training. To ensure adherence to the study protocol and representativeness of the sample [27], a reward of CHF 10 (US $10.90) for participation in each of the two follow-up assessments was announced.

Students were invited to participate in the study if they met the following criteria: (1) were a minimum age of 14 years, (2) owned a mobile phone, and (3) provided parental informed consent if they were under 15 years of age. Informed consent was obtained online from all study participants. Subsequently, they were invited to choose a username, provide their mobile phone number, and fill in the baseline assessment directly on their mobile phone.

Participants of the intervention group received additional questions, which were necessary for the tailoring of the intervention content. Furthermore, for participants of the intervention group, the mobile phone–based intervention program and its association with a friendly competition was described in detail. Subsequently, participants of the intervention group received individually tailored web-based feedback directly on their mobile phone (see Intervention Program section). During the subsequent 6 months, participants of the intervention group received individually tailored mobile phone–based life-skills training.

Participants of the assessment-only control group were thanked for their study participation and were informed about their group assignment and their reward for participation in the follow-up assessment.

Follow-up assessments in both study groups were conducted using a similar procedure: participants were invited to the online follow-up assessments via SMS text messaging, which included a link to the follow-up survey. Nonresponders were additionally addressed via computer-assisted telephone interviews conducted by research assistants. Study participants were recruited between March 2019 and March 2020. The 6-month follow-up assessments were conducted between August 2019 and September 2020.

Ethical Review and Trial Registration

The study protocol was approved on June 21, 2018, by the Ethics Committee of the Faculty of Arts and Sciences at the University of Zurich, Switzerland (approval No. 18.6.5). The trial was

executed in compliance with the Helsinki Declaration. The trial was registered on July 21, 2018, at the ISRCTN Registry (ISRCTN41347061).

Randomization and Allocation Concealment

To avoid spillover effects within school classes, we conducted a cluster-randomized controlled trial using a school class as a randomization unit. Due to the heterogeneity of students in the different secondary schools, we used a separate randomization list for each school (ie, stratified randomization). Furthermore, to approximate equality of sample sizes in the study groups, we used block randomization with computer-generated randomly permuted blocks of 4 cases [28].

Junior scientists supervising the baseline assessment were blinded to the group allocation of school classes. In addition, group allocation was not revealed to participants until they had provided their informed consent, username, mobile phone number, and baseline data. Furthermore, the research assistants who performed the computer-assisted follow-up assessments for primary and secondary outcomes were blinded to the group allocation.

Intervention Program

Theoretical Background and Intervention Contents

The intervention elements of the program were based on social cognitive theory [20,21]. The key concepts of this theory, which were addressed within *SmartCoach*, were (1) outcome expectations, (2) self-efficacy, (3) observational learning, (4) facilitation, and (5) self-regulation. Their implementation within the *SmartCoach* program is described in more detail within the study protocol [29].

Technological Background

The intervention program was developed using the MobileCoach system. Technical details of the system are described elsewhere [30,31]. The MobileCoach system is available as an open-source project. Password protection and Secure Sockets Layer encoding were used to ensure the privacy and safety of data transfer.

Individually Tailored Feedback

Individually tailored web-based feedback was provided to study participants of the intervention group immediately after completion of the online baseline assessment within the school classes. It comprised seven screens, including textual and graphical feedback on stress in general, individual level of stress in various domains, individual applied and suggested coping strategies, as well as individual level of social skills. Instruments for the assessment of stress and coping strategies were derived from the Juvenir 4.0 study, a national survey on stress in adolescents with more than 1500 participants [32]. Data from this survey were also used to provide age- and gender-specific feedback on individual stress level.

Text Messages

For a period of 22 weeks, program participants received between two and four individualized text messages per week on their mobile phone. These messages were generated and sent by the fully automated system. Within the first 7 weeks, the messages focused on self-management skills (eg, coping with stress, emotional self-regulation, or management of feelings of anger and frustration). In weeks 8 to 17, the messages focused on social skills (eg, making requests, refusing unreasonable requests, and meeting new people). In weeks 18 to 22, the text messages focused on substance use resistance skills (eg, recognizing and resisting media influences, correcting normative misperceptions of substance use, or understanding the associations of self-management skills and social skills with substance use). The messages were tailored according to the individual data from the baseline assessment and were based on text messaging assessments during the program runtime (eg, on substance use or on the individual's emotional state).

Several interactive features, such as quiz questions, tasks to create individually tailored if-then behavior plans based on implementation intentions, and message contests, were implemented within the program. Due to the wide dissemination of smartphones among adolescents [22], several messages also included hyperlinks to audio files (eg, audio testimonials and motivational podcasts) as well as to thematically appropriate video clips, pictures, and related websites. Table 1 displays the sequence and content of the text messages.

Figures 1 and 2 show a selection of intervention elements from the *SmartCoach* program.

Table 1. Sequence and content of text messages within the *SmartCoach* program.

Week No.	Content	Required activities
1	Introduction to self-management skills Origin and function of stress	Reply to quiz question Click on video link
2	Quiz on common stressors	Reply to quiz question Click on video link
3	Tailored stress reduction strategies for individual stressors	Reply to text message with options Click on video or website link
4	Self-challenge on general stress reduction strategies	Reply to text message with options Reply to text message on successful application of chosen strategy
5	Quiz on eustress versus distress	Reply to quiz question Click on video link
6	Tailored stress reduction strategies for individual stressors	Reply to text message with options Click on video or website link
7	Group contest on preferred stress management strategy	Post a picture and text message on individually preferred strategy Voting of others' posts Viewing of most-voted posts
8	Introduction to socials skills Quiz on social skills	Click on link to an overview picture Reply to quiz question Click on link to picture
9	Tailored strategies for improving personal social skills	Reply to text message with options Click on video or website link
10	Quiz on use of body language in different situations	Reply to quiz question Click on video link
11	Tailored strategies for improving personal social skills	Reply to text message with options Click on video or website link
12	Self-challenge on strategies to improve social skills in different areas	Reply to text message with options Reply to text message on successful application of chosen strategy
13	Origin of smartphone addiction	Reply to quiz question Click on video link
14	Quiz on associations between smartphone use and stress, tailored to gender	Reply to quiz question Click on video links
15	Self-challenge on smartphone detox	Reply to text message with options Reply to text message on successful detox in chosen situation Click on video link
16	Quiz on recognition of peer pressure	Click on link to the first part of the video Reply to quiz question Click on link to the second part of the video
17	Group contest on favorite social situation	Post a picture and text message on favorite social situation Voting of others' posts Viewing of others' posts
18	Introduction to substance use resistance skills Quiz on substance use prevalence (alcohol and tobacco) in reference group and normative feedback	Click on link to an overview picture Reply to quiz question
19	Quiz on the presence of tobacco advertisements directed to adolescents in everyday life	Reply to quiz question Click on video link
20	Quiz on risks of alcohol use	Reply to quiz question Click on website link
21	Tailored information on social consequences of alcohol use	Click on video link

Week No.	Content	Required activities
22	Group contest on motivation for abstinence or low-risk alcohol use	Post a motivational text message Voting of others' posts Viewing of others' posts

Figure 1. Screenshots (translated into English) from the *SmartCoach* program: baseline assessment (left) and feedback on social skills (right).

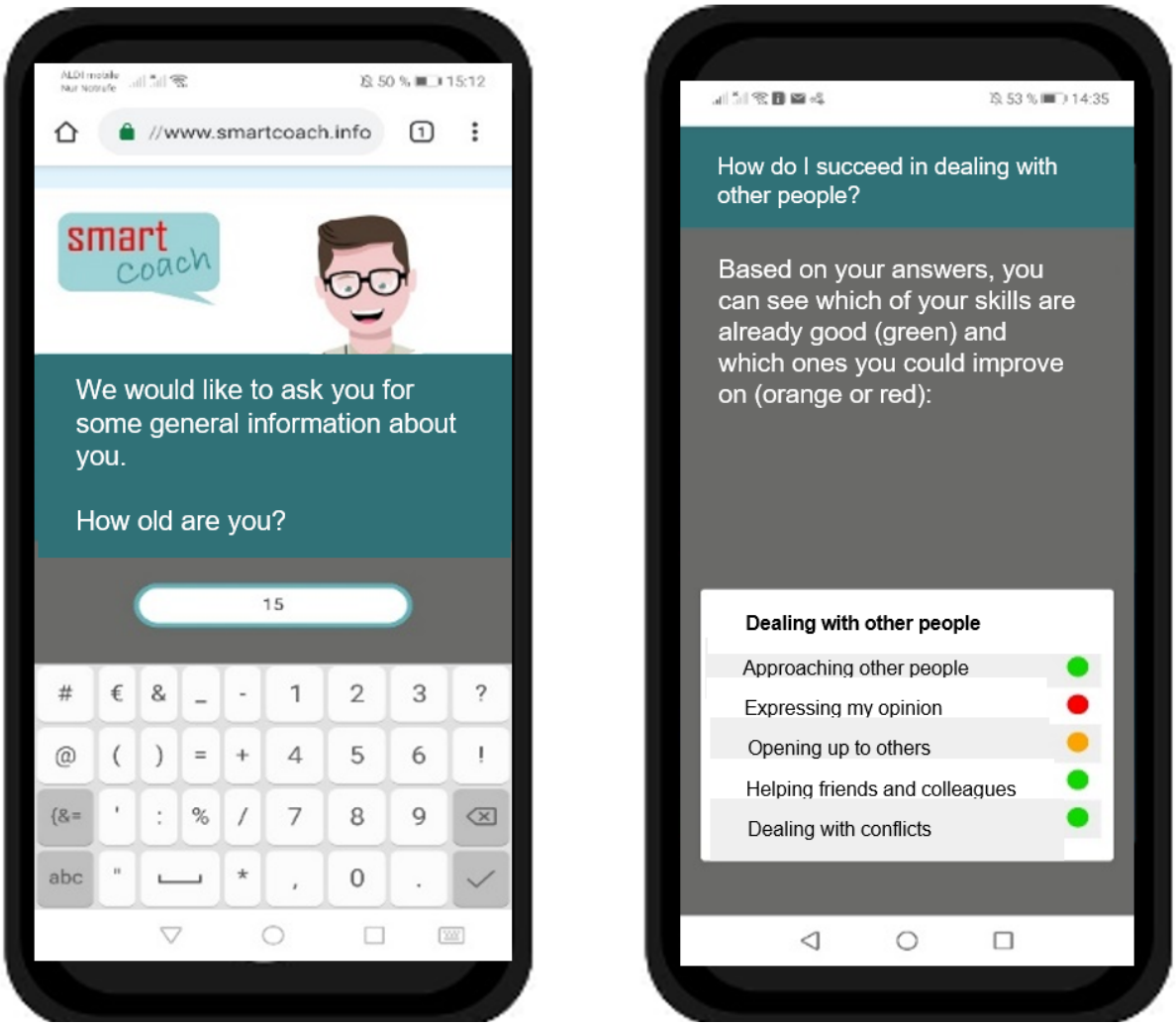
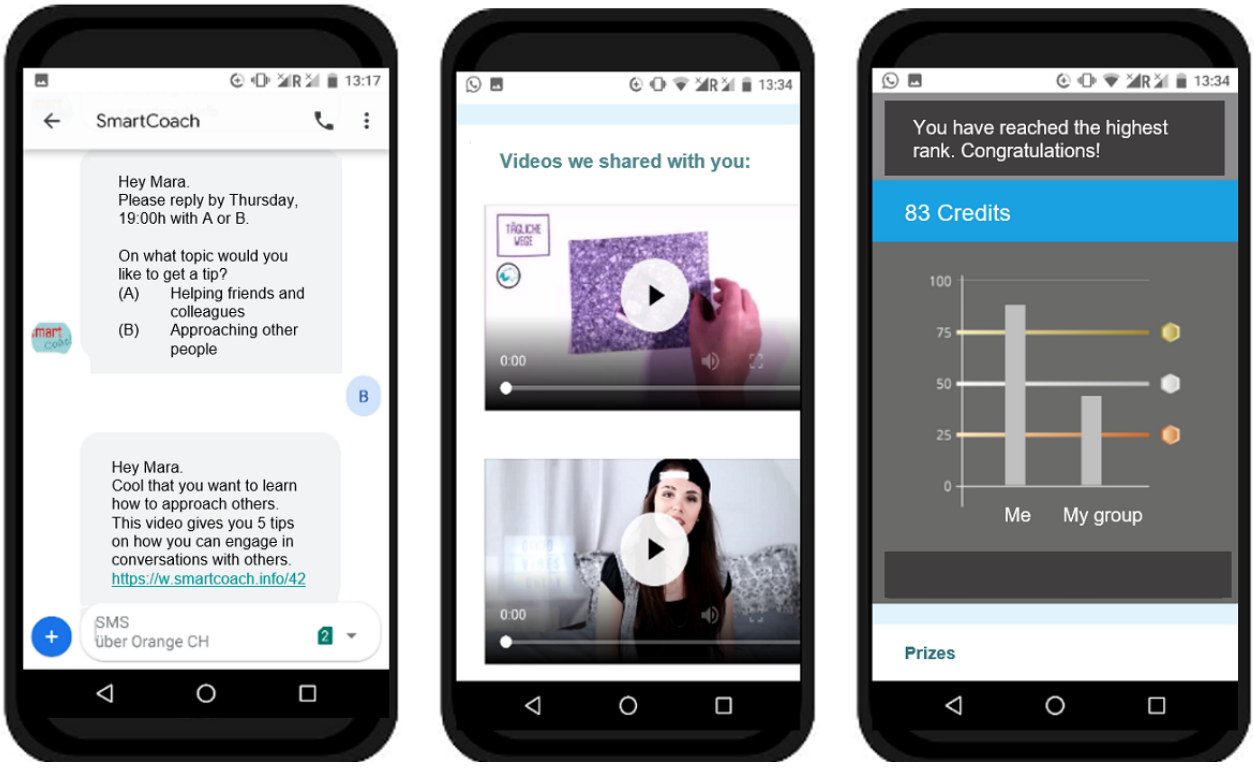


Figure 2. Screenshots (translated into English) from the *SmartCoach* program: text messages (left), video clips (middle), and friendly competition (right).



Prize Draw

To stimulate active program engagement, program use was associated with a friendly competition, which allowed program users to collect credits for each interaction (eg, answering monitoring text messages, participating in quizzes, creating messages or pictures within contests, and accessing video links integrated in text messages). The more credits participants collected, the higher their chances of winning one of 10 prizes, which were part of a prize draw (10 cash prizes of CHF 50 [US $54.50] each) after program completion. Participants were able to compare their number of credits with that of other program participants in their group (ie, similar starting date) at any time from an individual profile page. As can be seen in Figure 2 on the right, this was a mixture of feedback on the absolute number of credits for a participant and the relative score compared to the average number of credits for the reference group. However, the absolute individual score was ultimately decisive for the chances of winning a prize, so a win was possible when reaching the bronze level (25 credits), the chance was twice as high when reaching the silver level (50 credits), and there was a three-fold chance of winning when reaching the gold level (75 credits).

Assessments and Outcomes

Demographics

At baseline, demographic variables (ie, age, sex, and immigration background) as well as type of school (ie, secondary or upper secondary school) were assessed.

Program Use and Evaluation

To obtain the number of program participants who unsubscribed from the program within the program runtime of 6 months, we analyzed the log files of the MobileCoach system, in which all incoming and outgoing text messages were recorded. Using these log files, we also assessed the number of activities performed (eg, replies to text messaging prompts, accessing weblinks within text messages, and participating in contests) during the program. At follow-up, we assessed another aspect of SMS usage by asking the participants whether they usually (1) read through the text messages thoroughly, (2) took only a short look at them, or (3) did not read the text messages.

Furthermore, we evaluated whether (1) the number of received text messages was felt to be appropriate, (2) the duration of the program was adequate, (3) the participants would recommend the program to others, (4) the text messages were comprehensible, (5) the text messages were helpful, and (6) the text messages were perceived as individually tailored. Finally, program participants were asked to rate the program and different program elements, using the response categories *very good*, *good*, *less than good*, *bad*, and *don't know*.

Outcomes

Baseline and follow-up assessments included the following:

1. Problem drinking and alcohol use in the preceding 30 days, assessed by the short form of the Alcohol Use Disorders Identification Test–Consumption Items (AUDIT-C) [33]. This test is comprised of three items: (1) frequency of alcohol consumption, (2) quantity of alcohol consumption, and (3) binge drinking. Pictures were used to illustrate the quantity of a standard drink, which corresponded to 12 g to 14 g of pure alcohol. Based on a validation study of a large German sample, a cutoff score of ≥5 was used [34].
2. The 30-day point prevalence rate for smoking abstinence (ie, *not having smoked a puff* within the past 30 days

according to the criteria of the Society for Nicotine and Tobacco Research [35]).

3. Quantity of cigarettes smoked in the preceding 30 days, assessing by the number of smoking days and the typical number of cigarettes smoked per smoking day.
4. Cannabis use in the preceding 30 days, assessed by an item of the HBSC (Health Behaviour in School-aged Children) study [36] addressing the number of cannabis consumption days.
5. Perceived stress, assessed by a single item from the Swiss Juvenir study [32]—"How often have you had the feeling of being overstressed or overwhelmed in the last month?"—with answer options ranging from 1 (never) to 5 (all the time).
6. Well-being, assessed by the 5-item World Health Organization Well-Being Index (WHO-5) [37], with the final score ranging from 0, representing the worst imaginable well-being, to 100, representing the best imaginable well-being.
7. Social skills, assessed by the brief version of the 10-item Interpersonal Competence Questionnaire (ICQ-10) [38] addressing the following domains: (1) initiation of relationships, (2) negative assertion, (3) disclosure of personal information, (4) emotional support, and (5) conflict management.

The primary outcomes, according to the study protocol [29], are (1) prevalence of problem drinking in the preceding 30 days according to the AUDIT-C and (2) prevalence of cigarette smoking in the preceding 30 days (ie, having smoked at least a puff, according to the criteria of the Society for Nicotine and Tobacco Research [35]). Secondary outcomes were (1) prevalence of cannabis use in the preceding 30 days (ie, having used cannabis at least once), (2) quantity of alcohol use in the preceding 30 days, (3) quantity of cigarettes smoked in the previous 30 days, (4) frequency of cannabis use in the preceding 30 days, (5) perceived stress, (6) well-being, and (7) social skills.

Data Analyses

Descriptive statistics were used to present indicators of program use and evaluation. In order to examine baseline differences between participants of the intervention and control groups, we performed chi-square tests for categorical variables as well as *t* tests and Mann-Whitney *U* tests for continuous variables. The same tests were applied to assess whether participants who were lost to follow-up differed from those who responded, as a function of the study group.

We analyzed data according to the intention-to-treat (ITT) principle. For the ITT analyses, we used multiple imputation procedures as described elsewhere [39]. We imputed for each group separately to preserve homogeneity within the groups and potential interventional effects. Overall predictors of missing data at follow-up were gender, immigration background, education, and number of students within a school class. Differential predictors of missing data at follow-up were problem drinking, tobacco smoking, and use of the program. Thus, these predictors were part of all imputation models for the study's primary and secondary outcomes. The remaining study outcome predictors were variables that correlated at least weakly with these (r>0.20). Binary variables were imputed using logistic regression, categorical variables using multinomial logit models, and continuous variables using predictive mean matching. We examined 50 data sets and no systematic bias in convergence was revealed; thus, the final inferences were derived from this solution.

Next, we calculated the intraclass correlation (ICC) for primary and secondary outcomes. In our study, the ICC determines the extent to which study outcomes vary across classrooms. If an ICC is close to 0, standard regressions provide unbiased coefficients, whereas an ICC higher than 0 indicates that hierarchical models are needed to avoid a type I statistical error. In previous studies, ICCs between 0.05 and 0.10 were considered negligible [40,41]. However, it is an open debate as to how well the ICC performs depending on the underlying data [42]. Thus, we opted for a conservative approach and conducted linear mixed models (LMMs) and generalized linear mixed models (GLMMs) where the ICC was higher than 5%, and logistic or linear regressions where the ICC was below 5%.

Within LMMs and GLMMs, we modeled a random intercept for school class, while predictors and covariates were identical to logistic or linear regressions. Analyses of binary outcomes focused on follow-up values. Independent variables included baseline values for the binary variables of interest, group as a predictor, and variables for which baseline differences were observed as covariates. Analyses of continuous outcomes included change in score from baseline to follow-up as the dependent variable. Independent variables included group as a predictor and variables for which baseline differences were observed. We included in all models a covariate that modeled the possible effect of the lockdown measures undertaken in Switzerland between February 28 and June 22, 2020, because of the COVID-19 pandemic. During this period, several parts of students' lives were affected (eg, schools and/or bars were closed), which may have had an effect on our outcomes. The results from the imputed data set were cross-checked with the nonimputed data set. Results with a type I error rate of P<.05 on two-sided tests were considered statistically significant. Analyses were performed using SPSS, version 25 (IBM Corp), and R, version 3.6.1 (The R Foundation). Multiple imputation was conducted with the mice (multivariate imputation by chained equations) package in R [43], and LMM and GLMM were conducted with the lme4 (linear mixed-effects 4) package in R [44].

Results

Study Participants

Figure 3 depicts participants' progression through the trial. At the online screening assessment, 1759 students were present in 89 classes. Of these, 1623 (92.3%) students received parental approval to participate, and 1473 (83.7%) students ultimately participated in the study. A total of 44 classes containing 750 students in total were randomly assigned to the intervention group, and 45 classes containing 723 students in total were assigned to the control group. Follow-up assessments at 6 months were completed by 597 out of 750 (79.6%) participants

in the intervention group and 636 out of 723 (88.0%) participants in the control group.

Baseline characteristics for the study sample are shown in Table 2. The mean age was 15.4 years (SD 1.0), and 55.2% (813/1473) of the participants were female.

Baseline differences between the intervention and control groups were revealed for immigration background, education, prevalence of tobacco smoking and problem drinking, quantity of alcohol use, and perceived stress.

Concerning attrition bias, the analysis revealed that intervention group participants who were lost to follow-up reported more frequent tobacco use (Wald=6.38, *df*=1; *P*=.01) and problem drinking (Wald=8.97, *df*=1; *P*=.003) at baseline than controls.

Figure 3. Participants' progression through the trial.

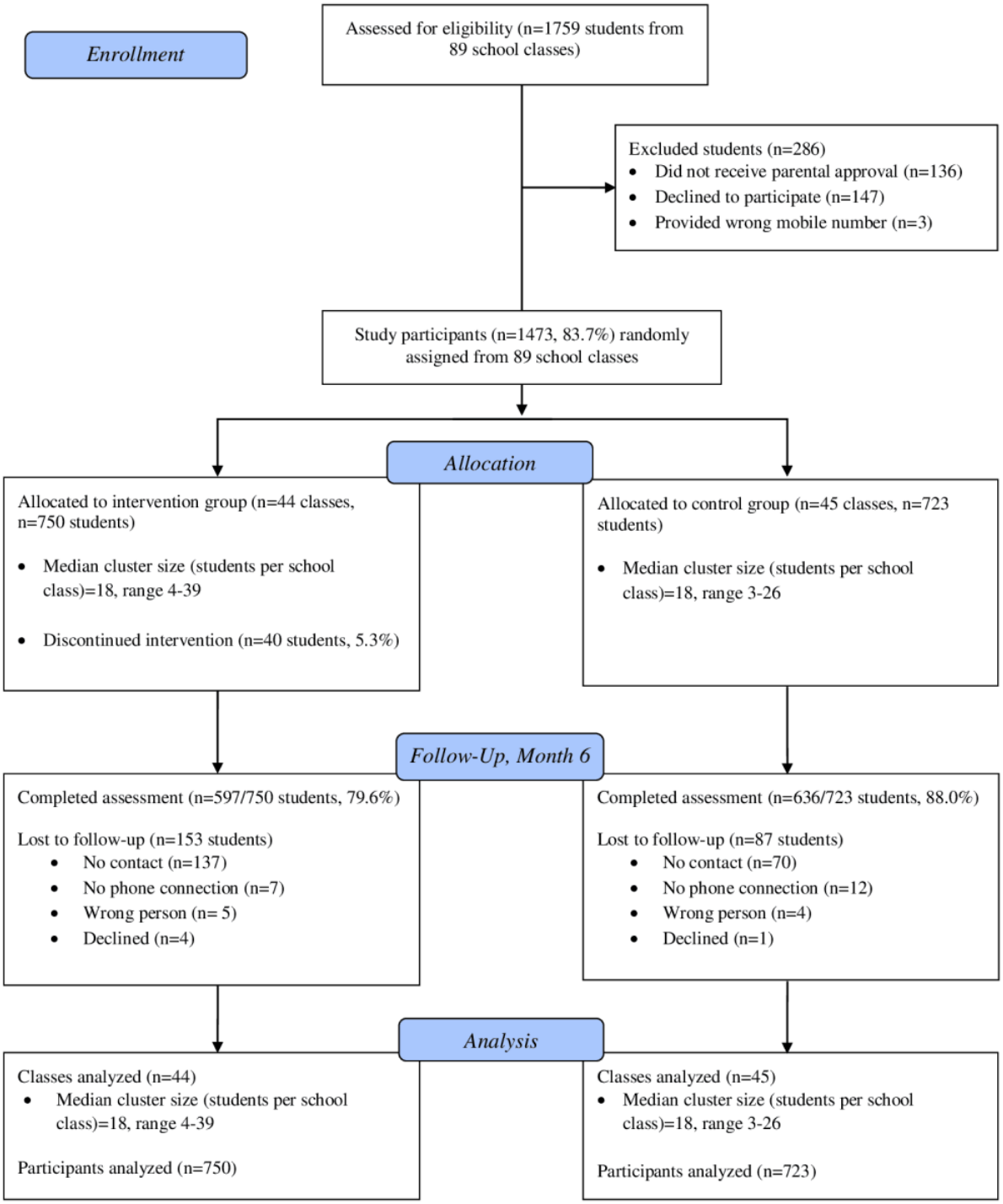


Table 2. Baseline characteristics of the study sample.

Variable	Intervention group (n=750)	Control group (n=723)	Total (N=1473)	*P* value[a]
Sex, n (%)				.34[b]
Male	327 (43.6)	333 (46.1)	660 (44.8)	
Female	423 (56.4)	390 (53.9)	813 (55.2)	
Age (years), mean (SD)	15.4 (1.0)	15.5 (1.0)	15.4 (1.0)	.052[c]
Immigration background, n (%)				<.001[b]
No immigration background	389 (51.9)	309 (42.7)	698 (47.4)	
One parent born outside Switzerland	173 (23.1)	168 (23.2)	341 (23.2)	
Both parents born outside Switzerland	188 (25.1)	246 (34.0)	434 (29.5)	
Type of school, n (%)				.007[b]
Secondary school	165 (22.0)	203 (28.1)	368 (25.0)	
Upper secondary school	585 (78.0)	520 (71.9)	1105 (75.0)	
Tobacco smoking, preceding 30 days, n (%)				.09[b]
No	659 (87.9)	614 (84.9)	1273 (86.4)	
Yes	91 (12.1)	109 (15.1)	200 (13.6)	
Quantity of cigarettes smoked, preceding 30 days, mean (SD)	5.3 (1.5)	7.9 (1.8)	6.6 (1.2)	.10[d]
Problem drinking, preceding 30 days, n (%)				
No	636 (84.8)	573 (79.3)	1209 (82.1)	.006[b]
Yes	114 (15.2)	150 (20.7)	264 (17.9)	
Quantity of alcohol use, preceding 30 days, mean (SD)	5.9 (16.3)	7.5 (16.4)	6.7 (16.4)	.003[d]
Cannabis use, preceding 30 days, n (%)				.95[b]
No	644 (85.9)	620 (85.8)	1264 (85.8)	
Yes	106 (14.1)	103 (14.2)	209 (14.2)	
Cannabis use days, preceding 30 days, mean (SD)	0.77 (3.4)	0.78 (3.5)	0.78 (3.5)	.95[b]
Perceived stress score[e], mean (SD)	2.9 (0.9)	3.0 (0.9)	2.99 (0.9)	.04[c]
Well-being score (WHO-5[f]), mean (SD)	52.9 (17.3)	51.6 (17.3)	52.3 (17.3)	.11[c]
Social skills score (ICQ-10[g]), mean (SD)	14.9 (2.2)	14.9 (2.2)	14.9 (2.2)	.45[c]

[a]*P* values for the comparison of the intervention and control groups.

[b]*P* value calculated from chi-square test.

[c]*P* value calculated from *t* test.

[d]*P* value calculated from Mann-Whitney *U* test.

[e]Perceived stress scores range from 1 (never) to 5 (all the time).

[f]WHO-5: 5-item World Health Organization Well-Being Index; final scores range from 0 (worst imaginable well-being) to 100 (best imaginable well-being).

[g]ICQ-10: 10-item Interpersonal Competence Questionnaire; final scores range from 5 (always poor/unable to handle social situations) to 20 (always good/able to handle social situations).

Program Use and Evaluation

During the intervention program, which lasted for 22 weeks, 40 of the 750 (5.3%) program participants withdrew their participation. A total of 50 activities (eg, replying to text messaging prompts, accessing weblinks within text messages, and participating in contests) were prompted over the 6-month program. The mean number of activities carried out by participants was 23.6 (SD 15.9). A total of 9.6% (72/750) of participants did not take part in any of the activities prompted by the program. Low engagement with the program was established for 149 out of 750 (19.9%) students, who interacted with it only 1 to 10 times. A total of 14.0% (105/750) of the participants interacted 11 to 20 times, 13.3% (100/750)

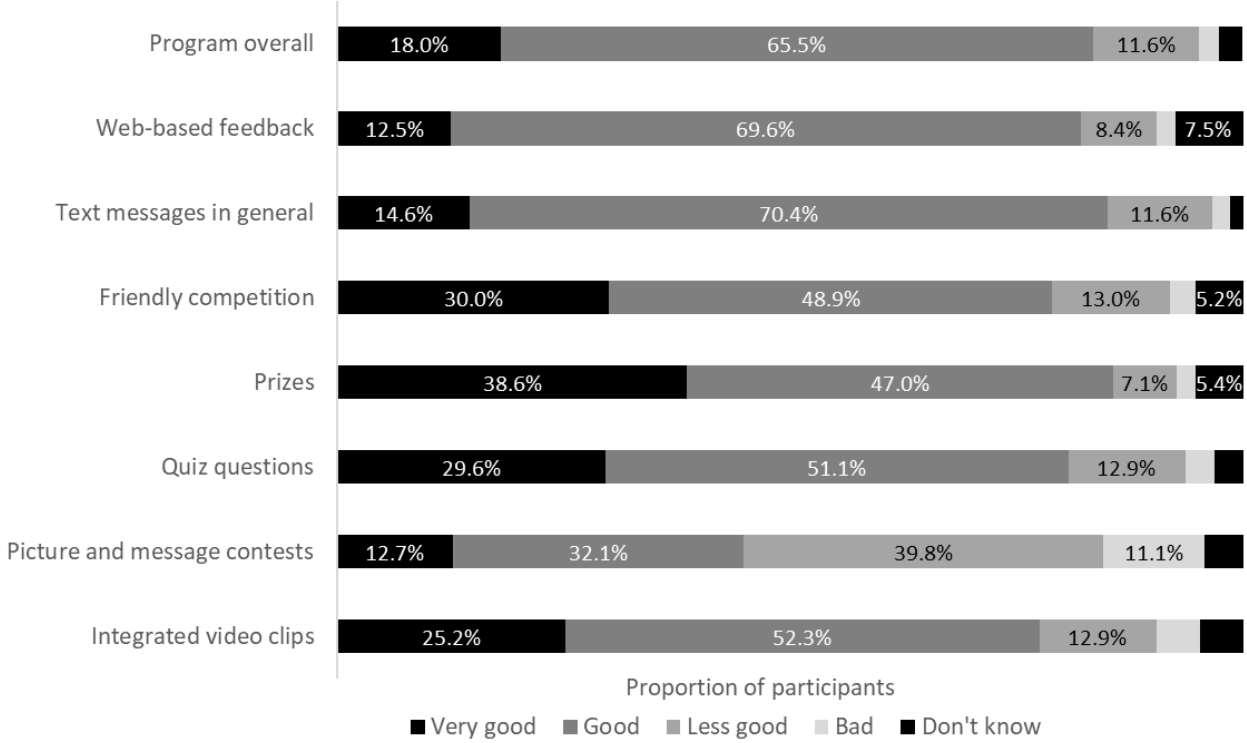
interacted 21 to 30 times, 24.9% (187/750) interacted 31 to 40 times, and 18.3% (137/750) interacted 41 to 50 times.

Of 597 participants with valid follow-up data, 563 (94.3%) answered the question regarding whether they had read the text messages. Of these, 70.9% (399/563) indicated that they *read the SMS messages thoroughly*, 27.0% (152/563) reported that they *took a quick look at the SMS messages*, and 1.6% (12/563) chose the predefined response category *I did not read the SMS messages*. The number of text messages received was rated as appropriate by 78.6% (442/562) of participants; 12.0% (90/562) would have preferred fewer messages, and 5.3% (30/562) would have preferred more text messages. Three-quarters of the participants rated the total length of the program as adequate (424/561, 75.6%); 7.5% (42/561) would shorten the program, and 12.7% (95/561) would extend the program. Over half of the participants (378/560, 67.5%) would recommend the program to others, while 32.5% (182/560) would not.

Almost all participants reported that the text messages were comprehensible (544/550, 98.9%). Participants were also asked if the text messages were helpful, and 384 out of 550 (69.8%) agreed that they were. A majority (327/550, 59.5%) indicated that they perceived the text messages as individually tailored to them.

Figure 4 presents additional evaluations of the program and specific program elements. The program, overall, was evaluated as *very good* or *good* by 83.6% (469/561) of the participants. Out of the specific program elements, the prizes, the text messages in general, the web-based feedback, and the quiz questions received the best evaluations, with more than 82.2% (461/561) of participants rating them as *good* or *very good*. The picture and message contests received the poorest ratings: 44.9% (252/561) of the participants rated them as *good* or *very good*.

Figure 4. Evaluations of the program and specific program elements by program participants (n=560). Values are presented for percentages >5.0%.



Initial Efficacy Based on 6-Month Follow-Up

The results of the complete-case (CC) and ITT analyses examining prevalence of problem drinking, tobacco smoking, and cannabis use are displayed in Table 3.

In the 30 days preceding the 6-month follow-up assessment, prevalence of problem drinking increased by 2.5% (from 15.2% to 17.7%) in the intervention group and by 3.4% (from 20.7% to 24.1%) in the control group, relative to that observed at baseline. This group effect was significant in the CC analysis (odds ratio [OR] 0.64, 95% CI 0.44-0.91; *P*=.01) but not in the ITT analysis (OR 0.71, 95% CI 0.49-1.03; *P*=.07). The prevalence of tobacco smoking also showed a steeper increase for controls (from 15.1% to 18.5%) compared to those who received the intervention (from 12.1% to 14.5%) from baseline to the 6-month follow-up, but this effect was not significant, neither in the ITT analysis (OR 0.83, 95% CI 0.50-1.36; *P*=.46) nor in the CC analysis (OR 0.82, 95% CI 0.46-1.43; *P*=.48). No significant group effect was observed for the pre-post difference in prevalence of cannabis smoking (+2.3% vs +1.3%; P_{ITT}=.21; P_{CC}=.60).

Results for continuous outcomes are summarized in Table 4.

Table 3. Intervention effects for dichotomous outcomes.

Outcome	Intervention group (n=750)			Control group (n=723)			*z*	*P* value	Odds ratio (95% CI)
	Baseline, n (%)	Follow-up, n (%)	Diff[a], %	Baseline, n (%)	Follow-up, n (%)	Diff, %			
Complete-case analysis									
Problem drinking in past 30 days[b]	114 (15.2)	88/597 (14.7)	–0.5	150 (20.7)	149/635 (23.5)	2.8	–0.45	.01	0.64 (0.44-0.91)
Tobacco smoking in past 30 days[b]	91 (12.1)	75/597 (12.6)	0.5	109 (15.1)	114/635 (18.0)	2.9	–0.20	.48	0.82 (0.46-1.43)
Cannabis use in past 30 days[c]	106 (14.1)	79/596 (13.3)	–0.8	103 (14.2)	101/635 (15.9)	1.7	–0.11	.60	0.90 (0.60-1.34)
Intention-to-treat analysis									
Problem drinking in past 30 days[b]	114 (15.2)	133 (17.7)	2.5	150 (20.7)	174 (24.1)	3.4	–0.34	.07	0.71 (0.49-1.03)
Tobacco smoking in past 30 days[b]	91 (12.1)	109 (14.5)	2.4	109 (15.1)	134 (18.5)	3.4	–0.19	.46	0.83 (0.50-1.36)
Cannabis use in past 30 days[c]	106 (14.1)	123 (16.4)	2.3	103 (14.2)	112 (15.5)	1.3	0.21	.21	1.24 (0.89-1.73)

[a]Diff: difference.

[b]Based on generalized linear mixed models with a random effect for school classes, group as fixed factor, follow-up scores as outcomes, and baseline scores, lockdown experience, immigration background, school type, perceived stress, and problematic alcohol use at baseline as covariates.

[c]Based on generalized linear model with the follow-up scores as outcomes, group as predictor, and the baseline score, lockdown experience, immigration background, school type, perceived stress, and problematic alcohol use at baseline as covariates.

Quantity of alcohol consumed per month decreased by 0.6 standard drinks in the intervention group and increased by 0.7 standard drinks in the control group (P_{ITT}=.03; P_{CC}=.06) from baseline to the follow-up assessment. Further, a significant group effect was observed for pre-post differences in cigarette smoking (–1.7 cigarettes per month in the intervention group and +5.0 cigarettes per month in the control group; P_{ITT}=.01; P_{CC}=.07) and reported stress (–0.2 in the intervention group and no change in the control group; P_{ITT}=.02; P_{CC}=.03). No significant group effect was observed for frequency of cannabis smoking (+0.01 days in the intervention group and +0.39 days in the control group; P_{ITT}=.053; P_{CC}=.02). Pre-post differences in well-being (+4.7 in the intervention group and +3.3 in the control group; P_{ITT}=.16; P_{CC}=.24) and social skills (+0.5 in the intervention group and +0.3 in the control group; P_{ITT}=.07; P_{CC}=.10) also did not differ significantly between groups.

Table 4. Intervention effects for continuous outcomes.

Outcome	Intervention group (n=750)			Control group (n=723)			*t* test (*df*)	*P* value	Effect size *d* (95% CI)
	Baseline	Follow-up	Diff[a], mean	Baseline	Follow-up	Diff, mean			
Complete-case analysis, mean (SD)									
Quantity of alcohol use in past 30 days[b]	5.9 (16.3)	4.7 (11.9)	–1.2	7.5 (16.4)	8.1 (17.1)	0.6	–1.64 (1224)	.06	–0.05 (–0.16 to 0.06)
Quantity of cigarettes smoked in past 30 days[b]	5.3 (41.8)	3.3 (20.7)	–2.0	7.9 (47.8)	12.5 (69.9)	4.6	–5.01 (1223)	.07	–0.11 (–0.22 to 0.006)
Cannabis smoking days in past 30 days[b]	0.8 (3.4)	0.7 (2.9)	–0.9	0.8 (3.5)	1.2 (4.5)	0.4	–0.45 (1209)	.02	–0.15 (–0.27 to –0.04)
Perceived stress score[c] in past 30 days[d]	2.9 (0.9)	2.8 (0.9)	–0.1	3.0 (0.9)	3.0 (1.0)	0	–0.19 (1209)	.03	–0.13 (–0.25 to –0.02)
Well-being score (WHO-5[e])[d]	52.9 (17.3)	56.9 (16.8)	4.0	51.6 (17.3)	54.3 (18.0)	2.7	1.64 (1211)	.24	0.05 (–0.06 to 0.17)
Social skills score (ICQ-10[f])[b]	14.9 (2.2)	15.4 (2.1)	0.5	14.9 (2.2)	15.2 (2.2)	0.3	0.20 (1184)	.10	0.08 (–0.03 to 0.19)
Intention-to-treat-analysis, mean (SD)									
Quantity of alcohol use in past 30 days[b]	5.9 (16.3)	5.3 (13.1)	–0.6	7.5 (16.4)	8.1 (16.8)	0.7	–1.74 (1465)	.03	–0.08 (–0.18 to 0.02)
Quantity of cigarettes smoked in past 30 days[b]	5.3 (41.8)	3.6 (21.1)	–1.7	7.9 (47.8)	12.9 (68.6)	5.0	–6.66 (1466)	.01	–0.13 (–0.23 to –0.03)
Cannabis use days in past 30 days[b]	0.8 (3.4)	0.8 (2.9)	0.0	0.8 (3.5)	1.2 (4.6)	0.4	–0.33 (1466)	.053	–0.12 (–0.22 to –0.01)
Perceived stress score in past 30 days[d]	2.9 (0.9)	2.7 (0.9)	–0.2	3.0 (0.9)	3.0 (1.0)	0	–0.21 (1473)	.02	–0.15 (–0.25 to –0.05)
Well-being score (WHO-5)[d]	52.9 (17.3)	57.6 (17.1)	4.7	51.6 (17.3)	54.9 (18.6)	3.3	2.04 (1473)	.16	0.07 (–0.04 to 0.17)
Social skills score (ICQ-10)[b]	14.9 (2.2)	15.4 (2.1)	0.5	14.9 (2.2)	15.2 (2.2)	0.3	0.20 (1465)	.07	0.08 (–0.02 to 0.18)

[a]Diff: difference.

[b]Based on linear models with the change scores from baseline to follow-up as outcomes, group as predictor, and lockdown experience, immigration background, school type, perceived stress, and problematic alcohol use at baseline as covariates.

[c]Perceived stress scores range from 1 (never) to 5 (all the time).

[d]Based on linear mixed models with a random effect for school classes, group as fixed factor, change scores from baseline to follow-up as outcomes, and lockdown experience, immigration background, school type, perceived stress, and problematic alcohol use at baseline as covariates.

[e]WHO-5: 5-item World Health Organization Well-Being Index; final scores range from 0 (worst imaginable well-being) to 100 (best imaginable well-being).

[f]ICQ-10: 10-item Interpersonal Competence Questionnaire; final scores range from 5 (always poor/unable to handle social situations) to 20 (always good/able to handle social situations).

Discussion

Principal Findings

This study tested the appropriateness and initial effectiveness of *SmartCoach*, a mobile phone–based life-skills training program for substance use prevention in a sample of proactively recruited secondary school students in Switzerland. Three main findings were revealed: (1) 4 out of 5 secondary school students (84%) participated in the study, showing a high interest in this interventional approach; (2) overall program use and engagement was good; and (3) initial results on program efficacy showed a significant intervention effect for some of the considered outcomes, including quantity of alcohol consumed per month, quantity of cigarettes smoked per month, and reported stress.

The proactive invitation for program and study participation in secondary and upper secondary schools, in combination with the offer of a low-threshold mobile phone–based intervention,

permitted us to reach 4 out of 5 adolescents for participation in the *SmartCoach* program and/or the associated study. Given the program duration of 22 weeks and that program participants needed to indicate their mobile phone number, this high participation rate of 84% is remarkable. Compared to the related *ready4life* program for life-skills training among vocational school students, the participation rate was similar, with 82% of students present within the school classes that could be recruited [26]. In contrast, substance-specific mobile phone–based programs, conducted in the same setting and using similar recruitment procedures, achieved slightly lower participation rates, with 50% to 75% of smokers participating in comparable programs to support smoking cessation [45-47] and around 75% in comparable programs for the prevention of problem drinking [48,49]. This preference for general life-skills training programs compared to substance-related programs might reflect the higher attractiveness of life skills–related topics, like stress management or social-skills training, but might also be associated with stigma and prejudice about substance-related disorders [50,51].

Concerning program use and engagement, the overall results were positive, with the majority of students (95%) remaining registered for the program for the total duration of 6 months and, on average, with program participants responding to half (mean of 23.6 out of 50) of the prompted activities. Program evaluations underlined its appropriateness for the target group of secondary school students, with the majority rating the program as helpful and individually tailored. However, 10% failed to engage in any of the 50 program activities, and another 20% showed low engagement, with less than 10 program interactions. Based on these findings, there is clearly room for improvement in terms of active program engagement, particularly concerning the picture and message contests, which received the poorest ratings among all program elements. The poor rating for this highly interactive element might be due to the limitations of mobile phone text messaging to receive and send pictures, which could be implemented more elegantly within a chat-based native smartphone app; however, direct comparisons of coaching programs based on SMS text messaging and smartphone apps concerning engagement and efficacy are still pending [52]. Compared to other text messaging–based prevention programs for adolescents, program engagement with *SmartCoach* was similar: the mean number of activities carried out by participants in the *ready4life* life-skills training program among vocational school students was 15.5 out of 39 possible activities [26]; within a smoking cessation program for vocational and upper secondary school students, participants answered a mean of 6.6 (SD 3.5) out of 11 text message prompts [53].

Further measures to increase program engagement based on the recommendations from a recent review [54] might be customizable features to provide a tailored experience and promote a sense of agency. For the *SmartCoach* program, this could include more flexibility concerning timing and extent of the intervention (eg, by the provision of fixed content at certain points in time and optional content, which the user can request flexibly). Furthermore, the provision of the right type of support at the right time by adapting to an individual's changing internal and contextual state, as conceptualized in just-in-time adaptive interventions (JITAIs) [55], has the potential to increase program engagement and effectiveness. Although JITAIs are typically provided via smartphone apps, a recently published study demonstrated that a text messaging–based just-in-time planning intervention was effective in reducing alcohol use among adolescents [56].

The results concerning the initial effectiveness of this program based on 6-month follow-up data are promising, with three of nine outcomes of the ITT analyses showing beneficial developments of statistical significance (ie, stress, quantity of alcohol use, and quantity of tobacco use; $P<.05$) and another three outcomes (ie, problem drinking prevalence, cannabis use days, and social skills) showing beneficial developments of borderline significance ($P<.10$). As the metric measures of substance use are more sensitive to change than the binary prevalence measures, which present the main outcomes of this study and provided the basis for the power calculations, these initial results are not conclusive for the effects of the primary outcomes of this study.

Limitations

The main limitations of this study are as follows:

1. Power calculations were based on the 18-month follow-up assessment [29]; therefore, all results concerning efficacy of the program should be considered as preliminary.
2. All data relied on self-report and the associated possibility that results may have been influenced by social desirability and a potential recall bias. Measures used to avoid under- or overreporting of substance use included assurance of confidentiality and anonymous assessments conducted via online survey and without personal contact, which may have increased the reliability of self-reported data.
3. Cluster randomization according to school class did not result in a balancing of all baseline characteristics.
4. There was selective attrition in the intervention group for persons with higher tobacco use and problem drinking at baseline. Although multiple imputations were used to compensate for this imbalance as much as possible, it would be interesting to investigate the reasons for this selective attrition. It is possible that the program reinforced cognitive dissonance and, associated with this, created a reactance toward the program. For future programs, this would mean that content should be chosen very carefully in this respect.
5. Some of the follow-up assessments were conducted during the lockdown restrictions due to the COVID-19 pandemic. This might have affected the generalizability of the results; however, this potential effect was addressed by the inclusion of a corresponding dummy covariate within all outcome analyses.
6. The results could not be generalized to secondary and upper secondary schools in Switzerland, as we recruited a convenience sample of school classes willing to participate in the study. However, the comparison of substance use prevalence rates among a representative sample of 15-year-old students in Switzerland [6] and the baseline characteristics of the study sample did not reveal major deviations. The 30-day point prevalence rates for tobacco

smoking were 14% in this study and 16% in the representative survey. For cannabis use, 30-day point prevalence was also 14% in this study and 11% in the survey. Concerning alcohol, the figures are not directly comparable, with 18% of the study sample showing problem drinking in the previous 30 days according to the AUDIT-C [33] and 25% practicing binge drinking in the representative sample [6].

Conclusions and Outlook

This is the first study that tested the appropriateness and efficacy of a mobile phone–delivered life-skills training program for substance use prevention among adolescents within a controlled trial. Our results suggest that this program, which delivers individualized messages and interactive activities integrated within a friendly competition, is both appropriate and promising in its effectiveness. Given that the program could be presented and introduced by research workers to students within one school lesson, it could be easily and economically implemented.

Our initial results indicate that the program might be effective in both preventing or reducing substance use and fostering life skills, such as coping with stress. However, data from the final 18-month follow-up will provide more robust results, including regarding potential moderators and mediators of program efficacy. Concerning moderators, it would be of particular interest to examine whether individuals with higher levels of substance use could also benefit from life-skills training programs. It would also be of particular interest to test which of the life skills addressed and successfully modified might prevent or decrease substance use.

Acknowledgments

Funding for this project was provided by the Swiss National Science Foundation (grant 10001C_179222/1). The funding institution did not influence the design and conduct of the study; the management, analysis, or interpretation of data; or the preparation, review, or approval of the manuscript. We thank Andreas Filler and his team from Pathmate Technologies for setting up the server for the program.

Conflicts of Interest

SH, AW, and RPC were involved in the development of the intervention program *SmartCoach*.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 107 KB - mhealth_v9i7e26951_app1.pdf]

References

1. Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry 2011 Oct;52(10):1015-1025 [FREE Full text] [doi: 10.1111/j.1469-7610.2011.02446.x] [Medline: 21815892]
2. Hall WD, Patton G, Stockings E, Weier M, Lynskey M, Morley KI, et al. Why young people's substance use matters for global health. Lancet Psychiatry 2016 Mar;3(3):265-279. [doi: 10.1016/S2215-0366(16)00013-4] [Medline: 26905482]
3. White AM. Understanding adolescent brain development and its implications for the clinician. Adolesc Med State Art Rev 2009 Apr;20(1):73-90, viii. [Medline: 19492692]
4. Gore FM, Bloem PJN, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: A systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
5. Inchley J, Currie D, Budisavljevic S, Torsheim T, Jåstad A, Cosma A, et al. Spotlight on Adolescent Health and Well-being. Findings From the 2017/2018 Health Behaviour in School-aged Children (HBSC) Survey in Europe and Canada. International Report. Volume 1. Key Findings. Copenhagen, Denmark: WHO Regional Office for Europe; 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/332091/9789289055000-eng.pdf [accessed 2021-06-30]
6. Delgrande Jordan M, Schneider E, Eichenberger Y, Kretschmann A. La Consommation de Substances Psychoactives des 11 à 15 Ans en Suisse – Situation en 2018 et Évolutions Depuis 1986 - Résultats de l'Étude Health Behaviour in School-aged Children (HBSC). Lausanne, Switzerland: Addiction Suisse; 2019. URL: https://www.addictionsuisse.ch/fileadmin/user_upload/DocUpload/HBSC-2018-rr-100.pdf [accessed 2021-06-30]
7. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
8. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. Cochrane Database Syst Rev 2013 Apr 30(4):CD001293. [doi: 10.1002/14651858.CD001293.pub3] [Medline: 23633306]
9. Foxcroft DR, Tsertsvadze A. Universal school-based prevention programs for alcohol misuse in young people. Cochrane Database Syst Rev 2011 May 11(5):CD009113. [doi: 10.1002/14651858.CD009113] [Medline: 21563171]

10. Bandura A. Social learning theory of aggression. J Commun 1978;28(3):12-29. [doi: 10.1111/j.1460-2466.1978.tb01621.x] [Medline: 690254]
11. World Health Organization. Skills for Health. Skills-Based Health Education Including Life Skills: An Important Component of a Child-Friendly/Health-Promoting School. Geneva, Switzerland: World Health Organization; 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/42818/924159103X.pdf?sequence=1&isAllowed=y [accessed 2021-06-30]
12. Botvin GJ, Griffin KW. Life skills training: Preventing substance misuse by enhancing individual and social competence. New Dir Youth Dev 2014;2014(141):57-65. [doi: 10.1002/yd.20086] [Medline: 24753278]
13. Ellickson PL, McCaffrey DF, Ghosh-Dastidar B, Longshore DL. New inroads in preventing adolescent drug use: Results from a large-scale trial of project ALERT in middle schools. Am J Public Health 2003 Nov;93(11):1830-1836. [doi: 10.2105/ajph.93.11.1830] [Medline: 14600049]
14. Weichold K, Blumenthal A. Long-term effects of the life skills program IPSY on substance use: Results of a 4.5-year longitudinal study. Prev Sci 2016 Jan;17(1):13-23. [doi: 10.1007/s11121-015-0576-5] [Medline: 26202801]
15. Velasco V, Griffin KW, Botvin GJ, Corrado Celata and Gruppo LST Lombardia. Preventing adolescent substance use through an evidence-based program: Effects of the Italian adaptation of Life Skills Training. Prev Sci 2017 May;18(4):394-405. [doi: 10.1007/s11121-017-0776-2] [Medline: 28353126]
16. Domitrovich CE, Durlak JA, Staley KC, Weissberg RP. Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. Child Dev 2017 Mar;88(2):408-416. [doi: 10.1111/cdev.12739] [Medline: 28213889]
17. Ellickson PL. You've shown the program model is effective. Now what? New Dir Youth Dev 2014;2014(141):95-105. [doi: 10.1002/yd.20089] [Medline: 24753281]
18. Champion KE, Newton NC, Teesson M. Prevention of alcohol and other drug use and related harm in the digital age: What does the evidence tell us? Curr Opin Psychiatry 2016 Jul;29(4):242-249. [doi: 10.1097/YCO.0000000000000258] [Medline: 27153124]
19. Evans RI, Dratt LM, Raines BE, Rosenberg SS. Social influences on smoking initiation: Importance of distinguishing descriptive versus mediating process variables. J Appl Soc Psychol 1988 Sep;18(11):925-943. [doi: 10.1111/j.1559-1816.1988.tb01184.x]
20. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
21. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviors interact: Social cognitive theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008:169-188.
22. Suter L, Waller G, Bernath J, Külling C, Willemse I, Süss D. JAMES: Jugend, Aktivitäten, Medien - Erhebung Schweiz [JAMES: Youth, Activity, Media - Data Switzerland]. Zurich, Switzerland: Zürcher Hochschule für angewandte Wissenschaften; 2018. URL: https://digitalcollection.zhaw.ch/bitstream/11475/12884/1/Ergebnisbericht_JAMES_2018.pdf [accessed 2021-06-30]
23. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": A telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
24. Ondersma SJ, Ellis JD, Resko SM, Grekin E. Technology-delivered interventions for substance use among adolescents. Pediatr Clin North Am 2019 Dec;66(6):1203-1215. [doi: 10.1016/j.pcl.2019.08.009] [Medline: 31679608]
25. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
26. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub MP. A mobile phone-based life skills training program for substance use prevention among adolescents: Pre-post study on the acceptance and potential effectiveness of the program, ready4life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
27. Edwards P, Cooper R, Roberts I, Frost C. Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires. J Epidemiol Community Health 2005 Nov;59(11):987-999 [FREE Full text] [doi: 10.1136/jech.2005.034397] [Medline: 16234429]
28. Pocock SJ. Clinical Trials: A Practical Approach. Chichester, UK: John Wiley & Sons; 1994.
29. Haug S, Paz Castro R, Wenger A, Schaub MP. Efficacy of a mobile phone-based life-skills training program for substance use prevention among adolescents: Study protocol of a cluster-randomised controlled trial. BMC Public Health 2018 Sep 10;18(1):1102 [FREE Full text] [doi: 10.1186/s12889-018-5969-5] [Medline: 30200928]
30. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. In: Proceedings of the 2015 Wireless Telecommunications Symposium (WTS). 2015 Presented at: 2015 Wireless Telecommunications Symposium (WTS); April 15-17, 2015; New York, NY p. 1-6. [doi: 10.1109/wts.2015.7117255]

31. Haug S, Kowatsch T, Castro RP, Filler A, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in young people: Study protocol of a cluster-randomised controlled trial. BMC Public Health 2014 Aug 07;14:809 [FREE Full text] [doi: 10.1186/1471-2458-14-809] [Medline: 25099872]
32. Jacobs Foundation. Juvenir Studie 4.0. Zuviel Stress – Zuviel Druck! Wie Schweizer Jugendliche mit Stress und Leistungsdruck umgehen [Juvenir Study 4.0. Too Much Stress – Too Much Pressure! How Swiss Adolescents Cope With Stress and Pressure]. Basel, Switzerland: Jacobs Foundation; 2015. URL: https://jacobsfoundation.org/app/uploads/2017/07/Juvenir-4.0_Kurzfassung_DE_final.pdf [accessed 2021-06-30]
33. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
34. Rumpf H, Wohlert T, Freyer-Adam J, Grothues J, Bischof G. Screening questionnaires for problem drinking in adolescents: Performance of AUDIT, AUDIT-C, CRAFFT and POSIT. Eur Addict Res 2013;19(3):121-127. [doi: 10.1159/000342331] [Medline: 23183686]
35. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: Issues and recommendations. Nicotine Tob Res 2003 Mar;5(1):13-25. [Medline: 12745503]
36. Marmet S, Archimi A, Windlin B, Delgrande Jordan M. Substanzkonsum bei Schülerinnen und Schülern in der Schweiz im Jahr 2014 und Trend seit 1986 - Resultate der Studie “Health Behaviour in School-aged Children” (HBSC) [Substance Use Among Students in Switzerland in 2014 and Trends Since 1986 - Results of the Study "Health Behaviour in School-aged Children” (HBSC)]. Lausanne, Switzerland: Sucht Schweiz; 2015. URL: https://www.hbsc.ch/pdf/hbsc_bibliographie_292.pdf [accessed 2021-06-30]
37. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: A systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
38. Coroiu A, Meyer A, Gomez-Garibello CA, Brähler E, Hessel A, Körner A. Brief form of the Interpersonal Competence Questionnaire (ICQ-15). Eur J Psychol Assess 2015 Oct;31(4):272-279. [doi: 10.1027/1015-5759/a000234]
39. van Buuren S. Flexible Imputation of Missing Data. Boca Raton, FL: CRC Press; 2012.
40. Mingo MA, Bell SM, McCallum RS, Walpitage DL. Relative efficacy of teacher rankings and curriculum-based measures as predictors of performance on high-stakes tests. J Psychoeduc Assess 2019 Feb 21;38(2):147-167. [doi: 10.1177/0734282919831103]
41. Alemu YM, Habtewold TD, Alemu SM. Mother's knowledge on prevention of mother-to-child transmission of HIV, Ethiopia: A cross sectional study. PLoS One 2018;13(9):e0203043 [FREE Full text] [doi: 10.1371/journal.pone.0203043] [Medline: 30204752]
42. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemp Clin Trials 2012 Sep;33(5):869-880 [FREE Full text] [doi: 10.1016/j.cct.2012.05.004] [Medline: 22627076]
43. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011 Dec;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
44. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015 Oct;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
45. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: A cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
46. Haug S, Paz Castro R, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: Results of a cluster-randomised controlled trial. J Subst Abuse Treat 2017 Nov;82:55-66. [doi: 10.1016/j.jsat.2017.09.008] [Medline: 29021116]
47. Haug S, Bitter G, Hanke M, Ulbricht S, Meyer C, John U. Brief text messaging (SMS)-delivered smoking cessation intervention for apprentices in vocational schools: A longitudinal intervention study for process and outcome evaluation [Article in German]. Gesundheitswesen 2013 Oct;75(10):625-631. [doi: 10.1055/s-0032-1331728] [Medline: 23292965]
48. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013 Sep 02;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
49. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Feb;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
50. Schomerus G, Lucht M, Holzinger A, Matschinger H, Carta MG, Angermeyer MC. The stigma of alcohol dependence compared with other mental disorders: A review of population studies. Alcohol Alcohol 2011;46(2):105-112. [doi: 10.1093/alcalc/agq089] [Medline: 21169612]
51. Corrigan PW, Lurie BD, Goldman HH, Slopen N, Medasani K, Phelan S. How adolescents perceive the stigma of mental illness and alcohol abuse. Psychiatr Serv 2005 May;56(5):544-550. [doi: 10.1176/appi.ps.56.5.544] [Medline: 15872162]

52. Suffoletto B. Text message behavioral interventions: From here to where? Curr Opin Psychol 2016 Mar 01;9:16-21 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.012] [Medline: 26665157]
53. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone-based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Nov 01;19(11):e356 [FREE Full text] [doi: 10.2196/jmir.7928] [Medline: 29092811]
54. Achilles MR, Anderson M, Li SH, Subotic-Kerry M, Parker B, O'Dea B. Adherence to e-mental health among youth: Considerations for intervention development and research design. Digit Health 2020;6:1-9 [FREE Full text] [doi: 10.1177/2055207620926064] [Medline: 32547775]
55. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
56. Haug S, Paz Castro R, Scholz U, Kowatsch T, Schaub MP, Radtke T. Assessment of the efficacy of a mobile phone-delivered just-in-time planning intervention to reduce alcohol use in adolescents: Randomized controlled crossover trial. JMIR Mhealth Uhealth 2020 May 26;8(5):e16937 [FREE Full text] [doi: 10.2196/16937] [Medline: 32452818]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test–Consumption Items
CC: complete case
GLMM: generalized linear mixed model
HBSC: Health Behaviour in School-aged Children
ICC: intraclass correlation
ICQ-10: 10-item Interpersonal Competence Questionnaire
ITT: intention to treat
JITAI: just-in-time adaptive intervention
lme4: linear mixed-effects 4
LMM: linear mixed model
mice: multivariate imputation by chained equations
OR: odds ratio
WHO-5: 5-item World Health Organization Well-Being-Index

Edited by L Buis; submitted 05.01.21; peer-reviewed by R Schwarzer, J Keller, D Szinay; comments to author 27.02.21; revised version received 10.03.21; accepted 20.04.21; published 13.07.21.

Please cite as:
Haug S, Paz Castro R, Wenger A, Schaub MP
A Mobile Phone–Based Life-Skills Training Program for Substance Use Prevention Among Adolescents: Cluster-Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(7):e26951
URL: https://mhealth.jmir.org/2021/7/e26951
doi:10.2196/26951
PMID:34255703

Original Paper

Assessing the Contribution of Self-Monitoring Through a Commercial Weight Loss App: Mediation and Predictive Modeling Study

Gregory Farage[1*], PhD; Courtney Simmons[1*], MPH; Mehmet Kocak[1], PhD; Robert C Klesges[1,2], PhD; G Wayne Talcott[1,2], PhD; Phyllis Richey[1], PhD; Marion Hare[1], MS, MD; Karen C Johnson[1], MPH, MD; Saunak Sen[1], PhD; Rebecca Krukowski[1], PhD

[1]Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Center for Addiction Prevention Research, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

[*]these authors contributed equally

Corresponding Author:
Rebecca Krukowski, PhD
Department of Preventive Medicine
College of Medicine
University of Tennessee Health Science Center
66 N Pauline St
Memphis, TN, 38163
United States
Phone: 1 901 448 2426
Email: rkrukows@uthsc.edu

Abstract

Background: Electronic self-monitoring technology has the potential to provide unique insights into important behaviors for inducing weight loss.

Objective: The aim of this study is to investigate the effects of electronic self-monitoring behavior (using the commercial *Lose It!* app) and weight loss interventions (with differing amounts of counselor feedback and support) on 4- and 12-month weight loss.

Methods: In this secondary analysis of the Fit Blue study, we compared the results of two interventions of a randomized controlled trial. Counselor-initiated participants received consistent support from the interventionists, and self-paced participants received assistance upon request. The participants (N=191), who were active duty military personnel, were encouraged to self-monitor their diet and exercise with the Lose It! app or website. We examined the associations between intervention assignment and self-monitoring behaviors. We conducted a mediation analysis of the intervention assignment for weight loss through multiple mediators—app use (calculated from the first principal component [PC] of electronically collected variables), number of weigh-ins, and 4-month weight change. We used linear regression to predict weight loss at 4 and 12 months, and the accuracy was measured using cross-validation.

Results: On average, the counselor-initiated–treatment participants used the app more frequently than the self-paced–treatment participants. The first PC represented app use frequencies, the second represented calories recorded, and the third represented reported exercise frequency and exercise caloric expenditure. We found that 4-month weight loss was partially mediated through app use (ie, the first PC; 60.3%) and the number of weigh-ins (55.8%). However, the 12-month weight loss was almost fully mediated by 4-month weight loss (94.8%). Linear regression using app data from the first 8 weeks, the number of self–weigh-ins at 8 weeks, and baseline data explained approximately 30% of the variance in 4-month weight loss. App use frequency (first PC; P=.001), self-monitored caloric intake (second PC; P=.001), and the frequency of self-weighing at 8 weeks (P=.008) were important predictors of 4-month weight loss. Predictions for 12-month weight with the same variables produced an R^2 value of 5%; only the number of self–weigh-ins was a significant predictor of 12-month weight loss. The R^2 value using 4-month weight loss as a predictor was 31%. Self-reported exercise did not contribute to either model (4 months: P=.77; 12 months: P=.15).

Conclusions: We found that app use and daily reported caloric intake had a substantial impact on weight loss prediction at 4 months. Our analysis did not find evidence of an association between participant self-monitoring exercise information and weight

loss. As 12-month weight loss was completely mediated by 4-month weight loss, intervention targets should focus on promoting early and frequent dietary intake self-monitoring and self-weighing to promote early weight loss, which leads to long-term success.

Trial Registration: ClinicalTrials.gov NCT02063178; https://clinicaltrials.gov/ct2/show/NCT02063178

(JMIR Mhealth Uhealth 2021;9(7):e18741) doi:10.2196/18741

KEYWORDS

weight loss; self-monitoring; obesity; apps; behavioral intervention

Introduction

Background

Consistent dietary and physical activity self-monitoring is an important component of successful weight loss in both traditional in-person behavioral weight loss programs [1,2] and technology-based programs [2-5]. Technology may also increase self-monitoring adherence [2,6-8]. Perhaps because technology-based dietary and physical activity self-monitoring requires minimal effort [9], commercial technology-based dietary and physical activity monitoring programs have grown in quantity and popularity with apps such as *Lose It!*, which has reported more than 30 million downloads to date, and MyFitnessPal, which had 225 million users in 2018 [10,11].

Self-weighing is another form of self-monitoring that facilitates weight loss [12-14], perhaps because of the behavioral changes in diet or exercise that occur as participants become more aware of their weight trajectories [15,16]. Until recently, many studies involving self-weighing relied on questionnaires in which participants could specify their weighing frequency [12]; now, it is possible to directly measure adherence to self-weighing through smart scales that record a participant's weight and self-weighing habits and transmit this information through wireless cellular technology.

Early performance (ie, self-monitoring adherence and weight loss) in a weight loss program is an indicator of long-term weight loss [17-20]. Tsai et al [17] showed that more detailed food records before randomization led to greater weight loss at 1 year. Similarly, Krukowski et al [19] found that early dietary and physical activity self-monitoring is a predictor of weight loss success. In a study by Unick et al [20], weight change at 2 months was predictive of weight change 8 years later. Early identification of participants who are not self-monitoring and who may be at most risk of not losing weight gives clinicians and researchers the opportunity to target those falling behind with additional resources.

Objectives

The first aim of our investigation was to increase our understanding of self-monitoring behaviors that lead to successful weight loss by using a commercial weight loss app, Lose It!. The second aim was to study if app use predicted weight loss at 4 and 12 months using participant data at an early stage (eg, 4 weeks, 8 weeks, and 4 months) to better identify individuals who might need early attention. As the app use data were composed of many interrelated variables, we summarized the variables using principal component analysis (PCA), which also addresses multicollinearity. We implemented a mediation analysis under the counterfactual framework to understand the mechanism of action of intervention assignment on weight loss at 4 and 12 months. We performed linear regression models with variables from the PCA analysis from the first 4 and 8 weeks of this study to predict weight loss at 4 and 12 months in the participants.

The prediction analysis allowed us to statistically model the relationship between the weight loss variables (at 4 and 12 months) and our set of independent predictors. It revealed the relationships among the variables; however, it did not indicate whether these relationships are causal. The complementary mediation analysis permitted us to make causal inferences about the effect of treatment assignment on weight loss through the mediators.

Methods

Study Design

This is a secondary analysis of the Fit Blue study, which was adapted from the Look AHEAD (Action for Health in Diabetes) Intensive Lifestyle Intervention [21-23] for a military lifestyle. The participants were randomly assigned to 1 of 2 treatment groups using a randomized block design with block size 4 [24]. The 2 treatment groups, counselor-initiated treatment and self-paced treatment, differed in the amount of self-initiation required to receive treatment. The counselor-initiated group received 28 phone calls over 12 months with counselors, regular feedback through email on their self-monitored entries on the same schedule as their phone sessions, 28 lesson materials, meal replacements, individualized detailed exercise and meal plans, and access to materials such as food scales, exercise videos, resource books, and cookbooks. They were also encouraged to participate in four challenges to boost their motivation. The self-paced participants could receive the same number of phone sessions and email feedback as the counselor-initiated participants upon request. The self-paced participants also had access to lesson materials and exercise and meal plans, although they had to initiate the request for assistance. Further details about the study design, meal replacements, and study website can be found elsewhere [24]. The main outcome as well as the treatment engagement outcomes have been previously published [25].

The behavioral change goals were standard across the two conditions. All participants were asked to record their daily dietary intake and physical activity for 12 months on the Lose It! app or website. Lose It! premium accounts were created for all participants. They were encouraged to track their diet, caloric

intake, and exercise using the app. The Lose It! app gave the participants access to a database with more than 7 million foods and corresponding nutritional information [26]. They were encouraged to lose 10% or more of their initial body weight at a rate of 1-2 pounds per week. Participants with a starting weight of less than 79 kg, between 79 kg and 97.5 kg, and more than 97.5 kg were given daily calorie goals of 1200-1300 kcal, 1500-1600 kcal, and 1800-1900 kcal, respectively. All participants were given a daily goal of consuming no more than 30% of calories from fat. In addition, the participants received a personalized exercise plan based on their self-reported physical activity at the baseline visit. They were asked to gradually increase aerobic exercise from their current level reported at baseline until reaching 225-250 minutes weekly, at which point they were to maintain this amount of exercise. A BodyTrace e-scale was provided to each participant, and they were requested to self-weigh daily. In addition, they were asked to attend two in-person follow-up data collection visits at 4 and 12 months.

Participants

A total of 248 active duty military personnel at Joint Base San Antonio (previously named Lackland Air Force Base) in San Antonio, Texas, participated in the Fit Blue study [24]. To be eligible to participate, participants had to meet the following criteria: aged 18 years or older, BMI of 25 kg/m^2 or more, access to a computer and email, and clearance from a health care provider. In addition, 1 week of monitoring dietary intake and physical activity on Lose It! was required for eligibility [24,27].

Participant recruitment began in December 2013 and ended in March 2016. The Fit Blue study was approved by the institutional review board of the Wilford Hall Ambulatory Surgical Center and acknowledged by the institutional review board of the University of Tennessee Health Science Center. The study approval was maintained over the course of the study, and a data and safety monitoring officer reviewed the accumulated data.

Measures

Physical Measurements

The primary outcome was weight change (in kilograms) measured on a calibrated scale (Tanita BWB-800S) in street clothes and without shoes. Height (in centimeters) was measured without shoes using a wall-mounted stadiometer. These measures were recorded at the baseline, 4-month weigh-in, and 12-month data collection visits. Our outcome variable, that is, weight loss, was modeled as the log ratio of the final weight at 4 and 12 months over the baseline weight. We log-transformed the weight change to stabilize variance and address the skewness of the distribution. As we are mostly interested in the participant's achievement, the weight loss ratio was classified as success (≥5% loss), some loss (2.3%-5%), or no loss (<2.3%). A 5% weight loss represents a benchmark at which point clinical benefits are observed [28]. Weight loss of less than 2.3% (or approximately 2.3 kg) has been used in previous research to denote weight stability [29].

Weight Self-Monitoring Behaviors

The participants were asked to monitor their weight daily using the BodyTrace e-scale. The time-stamped weights were transmitted to the study team over cellular technology, and they were also uploaded to each participant's personalized website to view their progress. The self-weighing variable represents the number of days the participants weighed themselves using the BodyTrace e-scale. We included only the plausible values filtered by the True Profile Finder algorithm [30].

Sociodemographic Characteristics

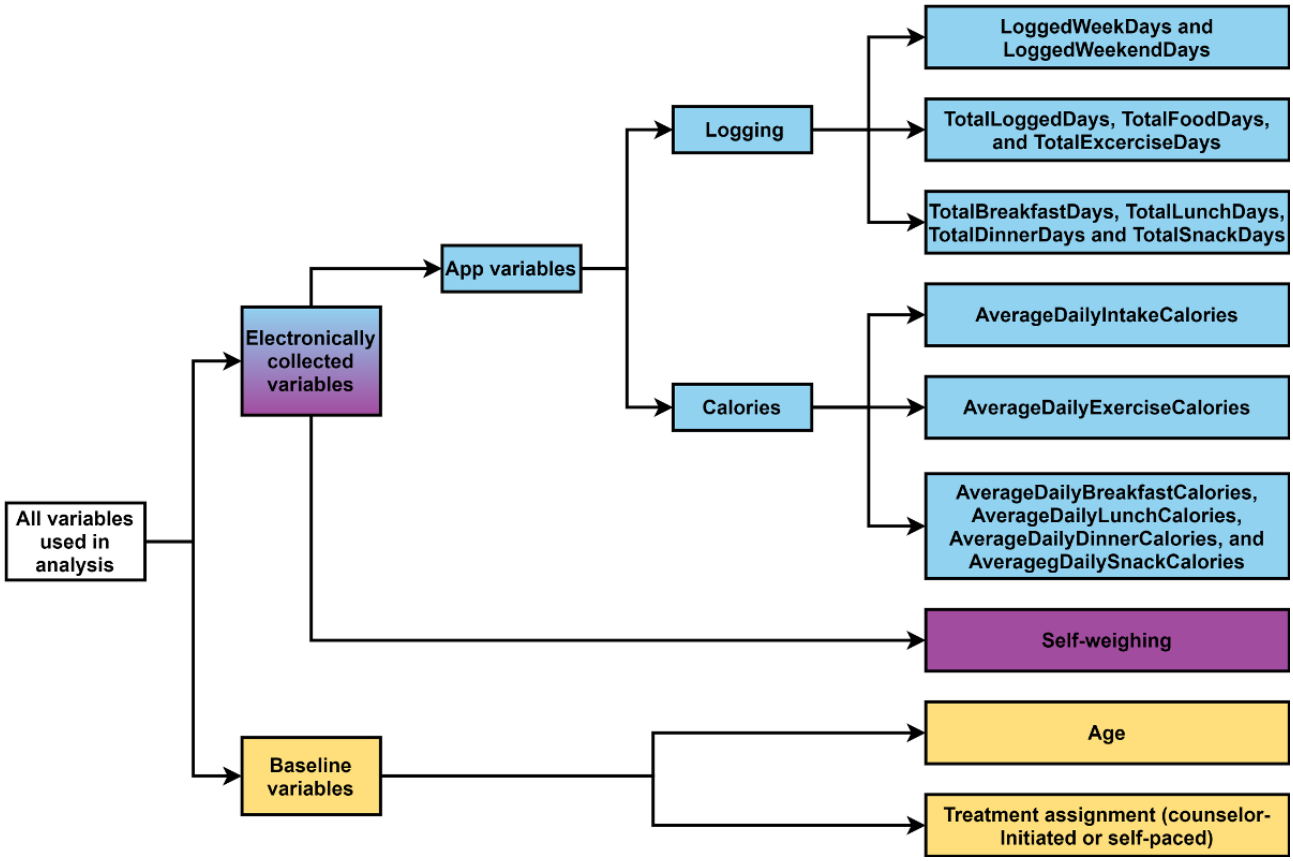
Age, education level, race, and gender were collected through a baseline questionnaire.

Dietary and Physical Activity Monitoring Behaviors

The participants recorded their daily consumption (ie, food and beverage items along with their calories) and exercise (ie, type and duration) using Lose It!. The participants were able to log food items for meals, snacks, and beverages, as well as exercise type and volume. They could also use the app to log their weight, which was counted as logging but not as the self-weighing variable. A total of 9 logging-specific measures were calculated to quantify logging behavior, and six caloric measures were estimated to assess caloric intake. We measured the average of the total number of days each week that participants logged at least 1 entry for food, beverage, exercise, or weight.

The independent variables included in the analysis (Figure 1) were categorized into two groups: baseline variables and electronically collected variables. The baseline variables consisted of age and treatment assignment. The electronically collected variables consisted of the frequency of 16 self-weighing and Lose It! app variables (Textbox 1). Each of the measures was calculated for a specific time period (ie, the first 4 weeks, the first 8 weeks, 4 months, and 12 months) depending on the analysis. The time periods used are mentioned in the analysis sections.

Figure 1. Diagram of variables used in prediction models and mediation analysis.



Textbox 1. Description of the electronically collected variables.

Logging measures

- *LoggedWeekDays* and *LoggedWeekendDays* are the number of weekdays and weekend days, respectively, that include at least 1 entry for food, beverage, exercise, or weight.
- *TotalLoggedDays* is the total number of days with at least 1 entry for food, beverage, exercise, or weight.
- *TotalFoodDays* is the total number of days with at least 1 food or beverage entry.
- *TotalExerciseDays* is the total number of days with at least 1 entry for exercise.
- *TotalBreakfastDays*, *TotalLunchDays*, *TotalDinnerDays*, and *TotalSnackDays* are the total number of days with at least 1 entry for breakfast, lunch, dinner, and snack, respectively.

Caloric measures

- *AverageDailyIntakeCalories* is the average daily intake of calories from food over the total number of days with at least 1 food entry. The entries for daily calories that exceeded an upper bound (ie, 4000 kcal for women and 5000 kcal for men) and lower bound (ie, 600 kcal for women and 800 kcal for men) were excluded for implausibility [31].
- *AverageDailyExerciseCalories* is the average daily calories burned through exercise over the total number of days with at least 1 entry for exercise.
- *AverageDailyBreakfastCalories*, *AverageDailyLunchCalories*, *AverageDailyDinnerCalories*, and *AverageDailySnackCalories* are the average daily caloric intake values over the total number of days with at least 1 entry for breakfast, lunch, dinner, and snack, respectively.

Statistical Analysis

Overview

The analysis was conducted using R version 4.0.4 [32]. We summarized the data and created graphics for descriptive analysis using the tidyverse package (version 1.3.0) [33]. Descriptive statistics were calculated to examine changes in logging frequency over the course of the study. We considered all sociodemographic variables collected for analysis and included them only if they were associated with a *P* value of .20 or less with the weight loss outcome. Using this criterion, only age was included in the final analysis.

Loss to Follow-up

As this is a longitudinal study, not all individuals' weights were observed at the 4- and 12-month visits. Our main analysis included only individuals with complete weight data. However, we also investigated the sensitivity of the effect of the missing

data using three different analyses. In the last observation carried forward analysis, we assumed that there was no change in the participants' weight after missing a visit. With a more conservative approach, the baseline observation carried forward analysis assumption is that the participants who missed a visit returned to their baseline weight. Neither method showed a substantial difference in the outcome. The results are detailed in the supplemental material available on a GitHub repository [34].

Principal Component Analysis

We performed a PCA on the Lose It! app variables to reduce the number of variables considered and still capture the useful information in these variables. The PCA also had the additional advantage of addressing multicollinearity in the variables [35]. Each principal component (PC) was a linear combination of the original variables and was uncorrelated with the other PCs. The PC that captured the largest variance was called the first PC, and the PC that captured the second largest variance was called the second PC, and so on [36,37]. In our case, we had 15 electronically collected variables, of which some were highly correlated with each other. The PCA reduced the number of uncorrelated variables to a much smaller number. We used the packages FactoMineR (version 2.4) [38] and factoextra (version 1.0.7) [39] to obtain the PCs and the new uncorrelated variables. We used a scree plot to determine the number of PCs to include in the analysis. The PCs that jointly explained 70% of the variance in the data were included in the analysis. Depending on the analysis of interest, PCA was applied to the variables measured over a specific period. Textbox 2 specifies these periods and the data used in the corresponding analyses.

Textbox 2. Time period for the electronically collected variables according to the analysis.

Mediation for 4-month weight loss

- Baseline to 4 months

Mediation for 12-month weight loss

- Baseline to 12 months

Prediction of the 4-month weight loss

- Baseline to 4 weeks or baseline to 8 weeks data

Prediction of the 12-month weight loss

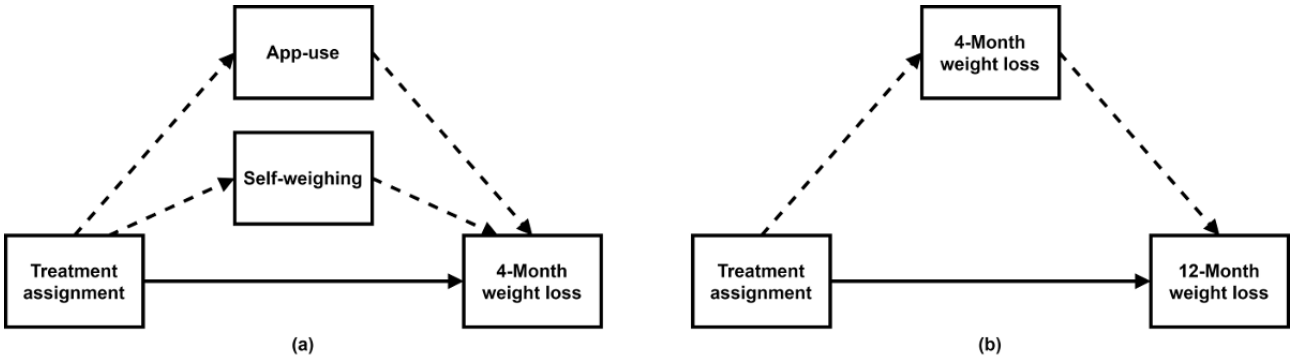
- Baseline to 4 weeks or baseline to 8 weeks data

To identify the most important variables that explain the variations in our data, we studied the quality of the representation and the contribution of the variables. The quality of representation of a variable is determined by the square of the correlation coefficient between a variable and a PC. The contribution of a variable for a given PC is estimated by the ratio of its squared correlation coefficient over the sum of all squared correlation coefficients between each variable and the given component [39].

Mediation

We conducted a causal mediation analysis to understand whether, and how much of, the intervention assignment effect on weight loss operates through intermediate variables such as *app use* and *self-weighing* frequency. Figure 2 depicts the hypothesized causal framework underlying our mediation analysis. The analysis was based on a counterfactual framework using a linear regression analytic approach [40]. Simply put, this framework partitions the total effect (TE) of the intervention assignment into the sum of the average causal mediation effect (ACME) and average direct effect. It can be shown that this approach can be used with both linear and nonlinear models and is equivalent to the traditional approach of MacKinnon et al [41] when some conditions hold [42]. In the counterfactual framework, we need to assume two statements to make valid inferences about the causal mediation analysis. This assumption is known as the sequential ignorability assumption. First, the treatment variable is statistically independent of the outcome and mediator variable; second, the mediator is independent of the outcome, given the observed exposure and pretreatment confounders. The counterfactual framework introduced by Imai et al [42] allows us to assess the robustness of our causal conclusions with respect to the violation of the sequential ignorability assumption. To establish if any association existed between the intervention assignment variable and the outcome of weight loss (at 4 and 12 months) and each mediator, we used the two-tailed t statistic. Significance was established using 10,000 permutations. In the causal mediation analysis [42], if the treatment assignment has no effect on the mediator, then the causal mediation effect is zero. The randomized exposure variable (ie, intervention assignment) has two states: self-paced or counselor-initiated. We conducted two mediation analyses for two outcomes: the 4-month weight loss and the 12-month weight loss. In the first mediation analysis, we considered two mediators of the 4-month weight loss: app use (calculated from the first PC of the electronically collected variables) and the frequency of self-weighing. In the second mediation analysis, we considered the mediation of the 12-month weight loss by the 4-month weight loss itself (Figure 2).

Figure 2. Diagram representing our causal model for treatment assignment, mediators, and outcomes. Arrows show the direction of causation. Dashed arrows represent mediated effects, and solid arrows represent direct effects. a: Causal model for the 4-month weight loss outcome. b: Causal model for the 12-month weight loss outcome.



The mediation analysis included all the data accumulated during the entire study, up to 4 months and 12 months, according to the period of interest. In all models for mediation, we included baseline covariates (age). To estimate the ACME (ie, indirect effect) and the average direct effect from the model-based causal mediation analysis, we used the package mediation (version 4.5.0) [43]. The uncertainty estimates were computed using a nonparametric bootstrap [43,44] with 10,000 simulations. The main motivation is not only to investigate whether a mediation exists, but also to determine if it is partially or fully mediated. It is fully mediated only if the direct effect is zero and partially mediated otherwise. As we had multiple mediators, we were interested in identifying the mediator that conveyed the greatest effect on weight loss [40,45-47].

Weight Loss Predictions Through Linear Regression

We used the mediation results to guide our variable selection process for the prediction. Five models were considered for 4-month weight loss, and six models were considered for 12-month weight loss. These models used a combination of baseline and electronically collected variables (ie, frequency of self-weighing and app variables summarized by PCA) as predictor variables. We predicted 4-month weight loss and 12-month weight loss using three different sets of predictor variables: (1) only baseline variables (ie, age and treatment assignment), (2) only electronically collected variables from the first 4 weeks, and (3) baseline and electronically collected variables from the first 4 weeks combined. The latter two models were repeated using electronically collected variables from the first 8 weeks. Finally, we considered one more model for 12-month weight loss using 4-month weight loss as the only predictor variable.

We used five-fold cross-validation to evaluate the linear regression models. The PCA was conducted using four out of five folds, including samples with a response value and incomplete samples without response values [48]. The resulting PCs were used to create a linear regression model. The model predicted log weight loss for the fifth fold composed of complete samples. This process was repeated until all participants' weight loss was predicted. We categorized actual and predicted weight loss values as successful weight loss, some loss, or no loss. Prediction accuracy was evaluated using R^2, the Spearman rank correlation coefficient (ρ), and the multi-class area under the receiver operating characteristic curve (AUC) value. The folds and cross-validation were performed using the caret package (version 6.0-80).

Results

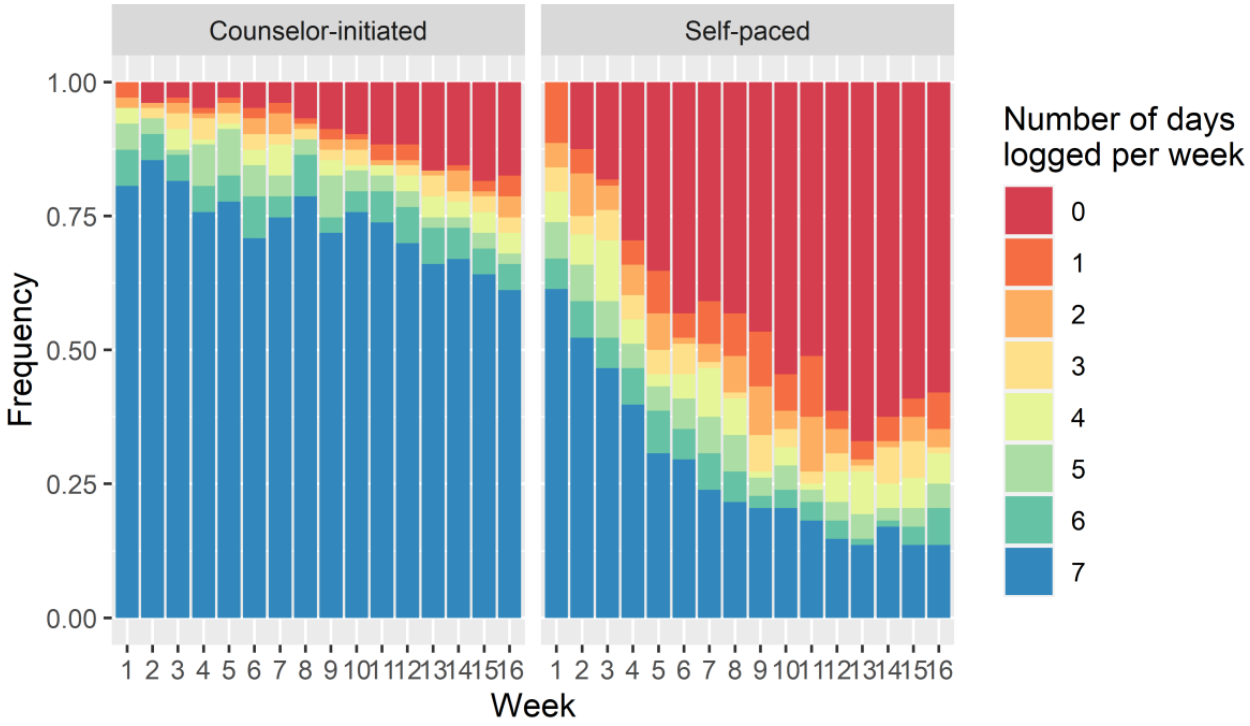
Overview

Owing to technical problems when downloading the data for 57 of the participants, 191 of the 248 randomized subjects were included in the analysis (including 103 in the counselor-initiated treatment and 88 in the self-paced treatment). Approximately 80.6% (154/191) of the participants attended the 4-month data collection visit, and approximately 80.1% (153/191) participated in the 12-month data collection visit. The study demographics are summarized in Table 1. At 4 months, the counselor-initiated treatment lost an average of 3.7 kg (SD 3.6), and the self-paced treatment lost 0.6 kg (SD 3.1). At 12 months, the counselor-initiated treatment lost 2.4 kg (SD 5.0) on average, and the self-paced treatment gained 0.2 kg (SD 5.1). Figure 3 shows the differences in the number of days with at least 1 item logged per week between the treatment groups. The counselor-initiated treatment 's average number of days with at least 1 item logged (ie, food, beverage, exercise, or weight entries) decreased from 6.48 days per week (week 1) to 5.03 days per week (week 16) over the first 4 months of the study and dropped to 0.70 days per week (week 52) by 12 months. The logging instances for the self-paced treatment decreased from 5.55 days per week (week 1) to 1.99 days per week (week 16) at 4 months and finally to 0.01 days per week (week 52) days by 12 months.

Table 1. Demographics of counselor-initiated and self-paced participants (N=191).

Characteristics	Counselor-initiated participants (n=103)	Self-paced participants (n=88)	Total (N=191)	*P* value
Gender, n (%)				.99
Female	53 (51.5)	46 (52.3)	99 (51.8)	
Male	50 (48.5)	42 (47.7)	92 (48.2)	
Race, n (%)				.66
Black or African American	22 (21.4)	16 (18.2)	38 (19.9)	
White	65 (63.1)	61 (69.3)	126 (65.9)	
Other	16 (15.5)	11 (12.5)	27 (14.1)	
Education, n (%)				.55
Less than college	46 (44.7)	44 (50)	90 (47.1)	
College or above	57 (55.3)	44 (50)	101 (52.9)	
BMI (kg/m^2), n (%)				.37
Overweight (BMI 25-29.9)	51 (49.5)	37 (42)	88 (46.1)	
Obese (BMI≥30)	52 (50.5)	51 (58)	103 (53.9)	
Age (years), mean (SD)	35.5 (8.2)	33.9 (6.8)	34.8 (7.6)	.14

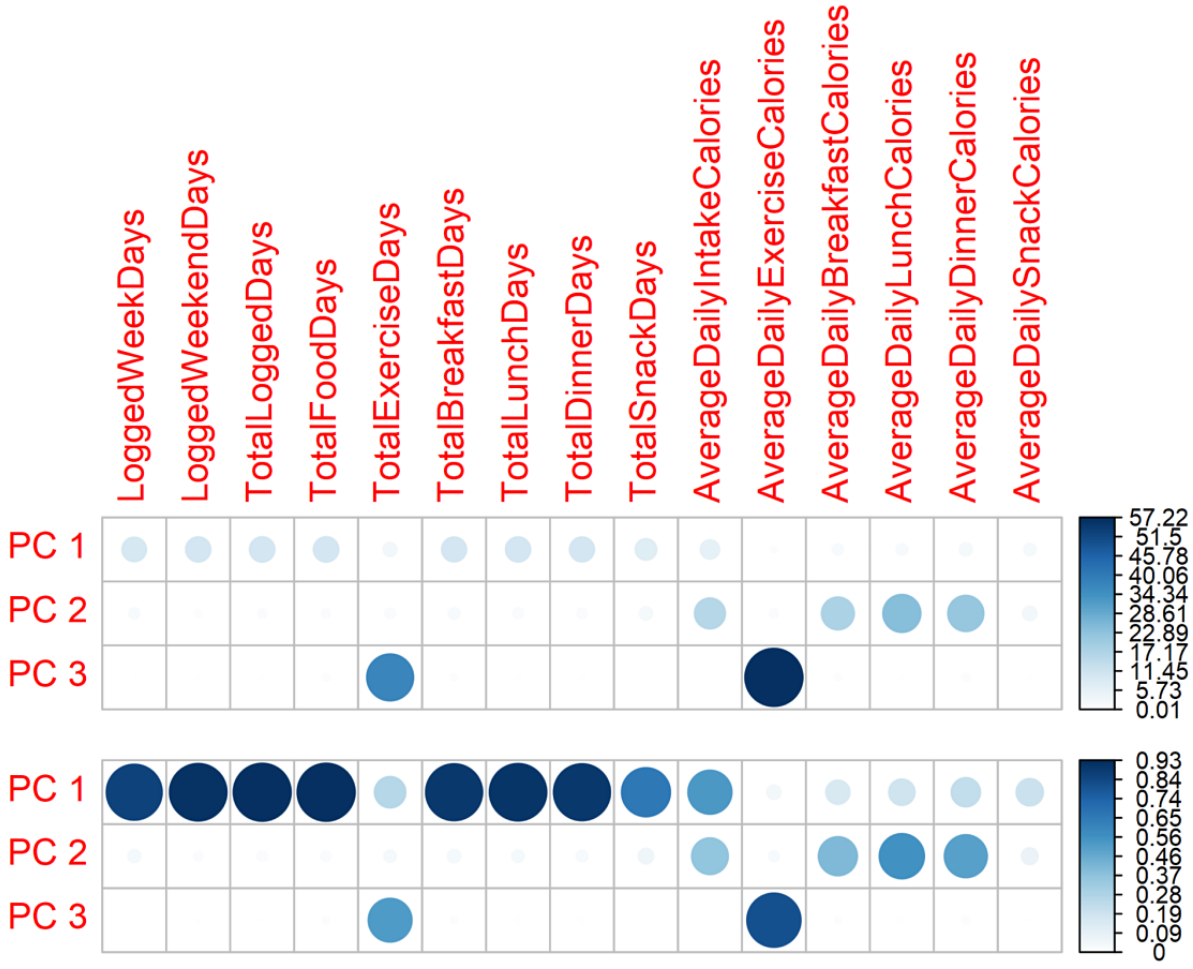
Figure 3. Number of days per week with at least one logging (by week and treatment groups) during the first 4 months.



The app variables were summarized using PCA with data from the first 4 and 8 weeks. A total of 72.8% of the variation in the 8-week data was captured in the first two components.

Although the PCs are linear combinations of the original variables, they have some interpretive value, that is, we still preserved some cohesive interpretations of the three first PCs. Figure 4 shows that PC_1 mainly comprises app variables describing the frequency with which the participants used the app; therefore, we labeled PC_1 *app-use*, and it explains 57.5% of the variance. PC_2 mostly describes daily caloric intake that was self-monitored by the participants, and it explains 15.3% of the variance. We labeled PC_2 *app-calories*. The inclusion of PC_3 explains 9.4% of the variance (82.2% total), and because this component represents the participants' exercise self-monitoring, it is labeled *app-exercise*.

Figure 4. The contribution of variable results (top) for the first 3 PCs from the 8-week PC analysis runs on a scale of 0 to 100. The darker and larger a circle, the more it contributed to a PC. The quality of the representation of variable results (bottom) are on a scale of 0 to 1. The darker and larger a circle, the more it is represented by a PC compared with other PCs. PC: principal component.



Mediation

We estimated the significance of the relationship between the intervention assignment variable (self-paced or counselor-initiated) and weight loss for two different periods (ie, baseline to 4 months and baseline to 12 months). There was a significant relationship between the treatment assignment variable (self-paced or counselor-initiated) and weight loss at 4 months (P<.001) and 12-months (P=.002).

For the two periods of interest (4 months and 12 months), treatment assignment was significantly associated (P<.001) with each mediator: the *app-use* component (PC_1), the *app-calories* component (PC_2), self-weighing frequency, and 4-month weight loss. However, treatment assignment was not associated with the *app-exercise* component (PC_3; P=.73). We considered only *app-use* (PC_1) as a mediator because it explained more than 50% of the variance by itself in comparison with *app-calories* (PC_2; <20%) and *app-exercise* (PC_3; <10%), which explained a much smaller proportion of the variance.

During the first 4 months (Table 2), we found that the indirect intervention's effect on weight loss transmitted through *app-use* (PC_1) and self-weighing frequency roughly accounted for 60% and 55% of the TE, respectively. For weight loss at 12 months, the indirect effect (ACME: 0.030, 90% CI 0.0017-0.046) through the 4-month weight loss accounted for 94% of the TE. The TE denotes how much weight loss would change overall if treatment assignment was changed from the self-paced to the counselor-initiated intervention. Table S7 in Multimedia Appendix 1 [44] contains the results of the sensitivity analysis; more details are provided in the GitHub repository [34].

Table 2. The 4- and 12-month mediation estimates the total effect, average direct effect (ADE), average causal mediation effect (ACME), and proportion mediated effect.

Mediator	Total effect (90% CI)	ADE[a] (90% CI)	ACME[b] (90% CI)	Proportion mediated, %
4-month mediation				
App-use (PC_1[a])	0.034 (0.023 to 0.045)	0.013 (0.002 to 0.025)	0.021 (0.012 to 0.029)	60.3
Self-weighing frequency	0.034 (0.023 to 0.045)	0.015 (0.004 to 0.026)	0.019 (0.012 to 0.027)	55.8
12-month mediation				
4-month weight loss	0.032 (0.015 to 0.048)	0.002 (−0.012 to 0.031)	0.03 (0.019 to 0.043)	94.8

[a]PC_1: principal component 1.

Weight Loss Predictions Through Linear Regression

The cross-validated linear models built on baseline variables and electronically collected variables (ie, *app-use* [PC_1], *app-calories* [PC_2], *app-exercise* [PC_3], and self-weighing frequency) were first compared for weight loss prediction accuracy using 4-week data. For 4-month weight loss, the 8-week model explained approximately 4% more variance than the model built on 4-week data (R^2=0.30 and 0.26, respectively). For 12-month weight loss, R^2 also slightly increased when 8-week data were used compared with using 4-week data (R^2=0.08 and 0.06, respectively).

As results with 8-week data showed marginal improvement when compared with those using 4-week data, this study primarily focuses on analysis with 8-week data. The results from the 8-week model performance are shown in Table 3. More information on the 4-week results and scatterplots of the models can be found in Multimedia Appendix 1.

Table 3. Accuracy of 8-week models for predicting 4-month and 12-month weight losses.

Model	4-month prediction			12-month prediction		
	R^2	ρ	mAUC[a]	R^2	ρ	mAUC
Baseline variables (ie, age and treatment assignment)	0.16	0.41	0.65	0.06	0.25	0.58
App-use (PC_1[b])+app-calories (PC_2[c])+self-weighing frequency	0.31	0.58	0.74	0.07	0.29	0.57
Baseline variables+app-use (PC_1)+app-calories (PC_2)+self-weighing frequency	0.30	0.58	0.73	0.08	0.29	0.58
4-month weight loss	N/A[d]	N/A	N/A	0.31	0.52	0.66

[a]mAUC: multi-class area under the receiver operating characteristic curve.

[b]PC_1: principal component 1.

[c]PC_2: principal component 2.

[d]N/A: not applicable.

For 4-month predictions, cross-validation of only the baseline variables using a linear model resulted in modest R^2 and ρ values (R^2=0.16; ρ=0.41; AUC=0.65). Adding the electronically collected variables (ie, *app-use* [PC_1], *app-calories* [PC_2], and self-weighing frequency at 8 weeks) to the model explained approximately 15% more variance than the baseline variables alone (R^2=0.30; ρ=0.58; AUC=0.73). The 12-month predictions were less successful than the 4-month predictions when considering a combination of baseline variables and electronically collected variables (Table 3). Using all analysis variables only accounted for 8% of the variance in the data (ρ=0.29; AUC=0.58) for the 12-month prediction. For both the 4-month and 12-month results, predictions based on only electronically collected variables tended to be slightly better than the predictions from the models that included baseline variables. This suggests that treatment assignment and age do not improve out-of-sample prediction when technological variables are already in the model.

The final linear model predicting 4-month weight loss, summarized in Table 4, was generated with baseline variables, the *app-use* component, the *app-calories* component, and the frequency of self-weighing at 8 weeks. The model revealed that the components of *app-use* and *app-calories* as well as the frequency of weighing were significant at the P=.01 level. After accounting for these variables, treatment assignment and age were found to be not significant. The adjusted R^2 value for the model was 0.32. If the *app-exercise* component (PC_3) was included, it was found to be not significant.

Table 4. Linear regression model summary predicting 4-month weight loss with 8-week data.

Coefficients	b (SE)	*t* test (*df*)	*P* value
Intercept	−1.89E−02 (1.67E−02)	−1.14 (148)	.25
App-use (PC_1[a])	4.51E−03 (1.35E−03)	3.34 (148)	.001
App-calories (PC_2[b])	6.20E−03 (2.06E−03)	3.01 (148)	.003
Treatment assignment	8.71E−03 (7.30E−03)	1.19 (148)	.23
Age	3.98E−04 (3.97E−04)	1 (148)	.32
Self-weighing frequency at 8 weeks	6.58E−04 (2.41E−04)	2.73 (148)	.007

[a]PC_1: principal component 1.

[b]PC_2: principal component 2.

For 12-month predictions (Table 5), the components of *app-use* and *app-calories* were not statistically significant predictors. Treatment assignment and the component of *app-exercise* were also not significant. However, age and the frequency of self-weighing were significant predictors.

Table 5. Linear regression results predicting 12-month weight loss with 8-week data.

Coefficients	b (SE)	*t* test (*df*)	*P* value
Intercept	−6.69E−02 (2.62E−01)	−2.55 (147)	.01
App-use (PC_1[a])	2.96E−03 (2.05E−03)	1.45 (147)	.15
App-calories (PC_2[b])	4.01E−03 (3.17E−03)	1.26 (147)	.21
Treatment assignment	6.27E−03 (1.12E−02)	0.56 (147)	.58
Age	1.34E−03 (6.08E−04)	2.21 (147)	.03
Self-weighing frequency at 8 weeks	7.84E−04 (3.77E−04)	2.08 (147)	.04

[a]PC_1: principal component 1.

[b]PC_2: principal component 2.

Although predicting 12-month weight loss with baseline and electronic data from the first 8 weeks resulted in low accuracy, using only 4-month weight loss as a predictor resulted in an R^2 of 0.31 (ρ=.52; multi-class AUC=.66).

Discussion

Principal Findings

Our results suggest that early study self-monitoring data, specifically PCs representing app use and self-monitoring of caloric intake and frequency of self-weighing, predict weight loss at 4 months, consistent with previous research indicating early self-monitoring predictors of treatment success [17,19]. Data from the first 8 weeks generated slightly more accurate weight loss predictions than data from the first 4 weeks; these results are consistent with those of Unick et al [20], who found that weight loss at 1 and 2 months was associated with 8-year weight loss. Predicting 12-month weight loss using early study data proved to be more challenging; however, 4-month weight loss was predictive of 12-month weight loss, which follows the results of previous research that indicated that early weight change can predict long-term weight change [18,20,49].

Our mediation results (Table 2) showed that *app-use* and the frequency of self-weighing partially mediated the relationship between treatment assignment and weight loss during the first 4 months. This suggests that the intervention not only directly affected 4-month weight loss but also indirectly affected weight change through weighing and app use behavior. However, the intervention mainly had a short-term effect because the results demonstrate a full mediation effect on the association between treatment assignment and 12-month weight loss through the 4-month weight loss.

Predictive modeling results revealed consistent self-monitoring to be an important aspect of 4-month weight loss, which reflects the findings of many previous studies [1,2]. The decrease in app use over time also followed a similar pattern to previous findings in web-based and traditional self-monitoring studies [19,50,51]. Our descriptive results on the differences in logging trends between the treatment groups provide preliminary evidence that self-monitoring with regular feedback may improve self-monitoring consistency, which differs from the results of some previous studies that show that there is no difference in adherence between a treatment group that receives feedback and a group that receives no feedback [3].

The regression results also demonstrated that the PC representing exercise self-monitoring was not a significant predictor of short-term weight loss. Previous research has shown that exercise is more crucial in maintaining weight than losing weight [52-54]. Although we did not measure exercise (ie, the behavior) in the study, self-monitoring exercise seemed to add

little benefit to short-term weight loss. Nonetheless, weight loss interventions should continue to encourage participants to increase their physical activity because of its role in weight maintenance; however, interventionists could focus on stressing the importance of self-monitoring caloric intake and self-weighing over monitoring exercise during the early intense weight loss period.

Overall, our results indicate that short-term weight loss leads to long-term weight loss, which is consistent with previous research. It is likely that the intervention establishes certain behaviors that induce weight loss at 4 months, some of which are continued at 12 months. The counselor-initiated condition seemed to establish more of these behaviors by 4 months, and the counselor-initiated–treatment participants were more successful at self-monitoring.

Strengths and Limitations

The strengths of the study include that it was a randomized clinical trial and one of the few studies that used data from a popular commercial app. In addition, the context of the randomized intervention study allowed us to conduct a causal mediation analysis to understand whether the intervention assignment directly and indirectly affected weight loss. Further, this is one of few weight control studies in the military.

By including a behavioral run-in period of dietary self-monitoring (as in previous studies [17]) in addition to requiring a medical clearance letter, it is possible that the participants may have been more motivated and well-informed about the study activities than those who did not complete these tasks. However, an examination of the characteristics of the randomized individuals compared with those of the individuals who were not randomized [27] showed that higher educational status was the only independent predictor of randomization. It is also interesting to note that it was initially expected that most of the participants would be motivated to join the study to assist them in passing the military fitness test. However, when the motivators for weight loss in this sample were examined, it was found that the most frequently endorsed motivators were improved physical health, improved fitness, improved quality of life, and a desire for longevity [55].

A limitation of using a commercial app in analysis is that the researchers have little control over the format in which they collect and receive data. In this study, a large portion of the nutrient information was not available because of the methods that the participants chose to log their calories. For instance, some participants logged calories directly without food description or nutrient details (eg, a lump sum of 2000 calories for the full day without further details); therefore, we were unable to use the nutrient reports in the analysis. We were also unable to include data from 57 participants because of technical errors in the data retrieval process because of a clerical mistake. This differential missingness between the 2 intervention conditions could have introduced bias; however, we note that the cause of the missing data was the data acquisition process and unrelated to the identity or behavior of the participants.

Conclusions

We found that long-term weight loss was completely mediated through short-term weight loss, reiterating the importance of interventions that produce strong successes quickly. Approximately one-third of the 12-month weight loss was explained by weight loss at 4 months. More than half of the effect of the behavioral intervention on weight loss at 4 months was mediated through self-monitoring app use and self-weighing frequency, indicating the potency of these self-regulatory behaviors. As we did not find evidence of an effect of self-monitored exercise, it suggests that diet self-monitoring should be prioritized for successful weight loss.

Acknowledgments

This research represents a collaborative research and development agreement with the United States Air Force (CRADA #13-168-SG-C13001). This study was also partially funded by the National Institute of Diabetes and Digestive and Kidney Diseases (RO1 DK097158; principal investigators: RK and RCK). The opinions expressed in this document are solely those of the authors and do not represent an endorsement by, or the views of, the United States Air Force, the Department of Defense, or the United States Government. We would like to thank the participants, the leadership at Second Air Force, and the research staff for their dedication to the research. We gratefully acknowledge the partnership with Lose It! and Body Trace.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental analysis tables.
[DOCX File , 50 KB - mhealth_v9i7e18741_app1.docx]

References

1. Boutelle K, Kirschenbaum D. Further support for consistent self-monitoring as a vital component of successful weight control. Obes Res 1998 May;6(3):219-224 [FREE Full text] [doi: 10.1002/j.1550-8528.1998.tb00340.x] [Medline: 9618126]
2. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behav Ther 1993;24(3):377-394. [doi: 10.1016/S0005-7894(05)80212-6]

3. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
4. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. Obesity 2008 May;16(5):1052-1060 [FREE Full text] [doi: 10.1038/oby.2008.19] [Medline: 18309301]
5. Spring B, Duncan JM, Janke EA, Kozak AT, McFadden HG, DeMott A, et al. Integrating technology into standard weight loss treatment: a randomized controlled trial. JAMA Intern Med 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1221] [Medline: 23229890]
6. Allen JK, Stephens J, Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
7. 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]
8. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
9. Harvey J, Krukowski R, Priest J, West D. Log often, lose more: electronic dietary self-monitoring for weight loss. Obesity 2019 Mar;27(3):380-384 [FREE Full text] [doi: 10.1002/oby.22382] [Medline: 30801989]
10. How Lose It! Is Helping Real People. Lose It. 2019. URL: http://www.webcitation.org/77f0LbIKt [accessed 2019-04-15]
11. Baer SK. 150 Million Users Of The MyFitnessPal App Had Their Data Stolen By A Third Party. BuzzFeed News. 2018. URL: https://www.buzzfeednews.com/article/skbaer/myfitnesspal-under-armour-data-stolen [accessed 2019-04-14]
12. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
13. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
14. Vanwormer JJ, French SA, Pereira MA, Welsh EM. The impact of regular self-weighing on weight management: a systematic literature review. Int J Behav Nutr Phys Act 2008 Nov 4;5:54. [doi: 10.1186/1479-5868-5-54] [Medline: 18983667]
15. Steinberg DM, Bennett GG, Askew S, Tate DF. Weighing every day matters: daily weighing improves weight loss and adoption of weight control behaviors. J Acad Nutr Diet 2015 Apr;115(4):511-518 [FREE Full text] [doi: 10.1016/j.jand.2014.12.011] [Medline: 25683820]
16. Zheng Y, Sereika SM, Ewing LJ, Danford CA, Terry MA, Burke LE. Association between self-weighing and percent weight change: mediation effects of adherence to energy intake and expenditure goals. J Acad Nutr Diet 2016 Apr;116(4):660-666 [FREE Full text] [doi: 10.1016/j.jand.2015.10.014] [Medline: 26727241]
17. Tsai AG, Fabricatore AN, Wadden TA, Higginbotham AJ, Anderson A, Foreyt J, et al. Readiness redefined: a behavioral task during screening predicted 1-year weight loss in the look AHEAD study. Obesity 2014 Apr;22(4):1016-1023 [FREE Full text] [doi: 10.1002/oby.20648] [Medline: 24151217]
18. West DS, Harvey-Berino J, Krukowski RA, Skelly JM. Pretreatment weight change is associated with obesity treatment outcomes. Obesity 2011 Sep;19(9):1791-1795 [FREE Full text] [doi: 10.1038/oby.2011.22] [Medline: 21331064]
19. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170 [FREE Full text] [doi: 10.1037/a0028135] [Medline: 22545978]
20. Unick JL, Neiberg RH, Hogan PE, Cheskin LJ, Dutton GR, Jeffery R, Look AHEAD Research Group. Weight change in the first 2 months of a lifestyle intervention predicts weight changes 8 years later. Obesity 2015 Jul;23(7):1353-1356 [FREE Full text] [doi: 10.1002/oby.21112] [Medline: 26110890]
21. Johnston CA, Moreno JP, Foreyt JP. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. Curr Atheroscler Rep 2014 Dec;16(12):457 [FREE Full text] [doi: 10.1007/s11883-014-0457-6] [Medline: 25288176]
22. Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, Williamson D, Look AHEAD Research Group. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]
23. Look AHEAD Research Group, Pi-Sunyer X, Blackburn G, Brancati FL, Bray GA, Bright R, et al. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the look AHEAD trial. Diabetes Care 2007 Jun;30(6):1374-1383 [FREE Full text] [doi: 10.2337/dc07-0048] [Medline: 17363746]
24. Krukowski RA, Hare ME, Talcott GW, Johnson KC, Richey PA, Kocak M, et al. Dissemination of the look AHEAD intensive lifestyle intervention in the United States Air Force: study rationale, design and methods. Contemp Clin Trials 2015 Jan;40:232-239 [FREE Full text] [doi: 10.1016/j.cct.2014.12.014] [Medline: 25545025]

25. Krukowski RA, Hare ME, Talcott GW, Gladney LA, Johnson KC, Richey PA, et al. Dissemination of the look AHEAD intensive lifestyle intervention in the United States Military: a randomized controlled trial. Obesity 2018 Oct;26(10):1558-1565 [FREE Full text] [doi: 10.1002/oby.22293] [Medline: 30277030]
26. Weight Loss That Fits. Lose It. 2018. URL: http://www.webcitation.org/73rfgQDPz [accessed 2018-11-12]
27. Fahey M, Hare ME, Talcott GW, Kocak M, Hryshko-Mullen A, Klesges RC, et al. Characteristics associated with participation in a behavioral weight loss randomized control trial in the U.S. Military. Mil Med 2019 Mar 1;184(3-4):e120-e126 [FREE Full text] [doi: 10.1093/milmed/usy199] [Medline: 30125001]
28. Williamson DA, Bray GA, Ryan DH. Is 5% weight loss a satisfactory criterion to define clinically significant weight loss? Obesity (Silver Spring) 2015 Dec;23(12):2319-2320 [FREE Full text] [doi: 10.1002/oby.21358] [Medline: 26523739]
29. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/NEJMoa061883] [Medline: 17035649]
30. Kocak M, Krukowski R, Talcott G. Processing, Cleaning, and Streaming data in SAS. PharmaSUG. 2018. URL: http://web.archive.org/web/20201124054222/www.pharmasug.org/proceedings/2018/DV/PharmaSUG-2018-DV11.pdf [accessed 2019-08-20]
31. Jeon KJ, Lee O, Kim H, Han SN. Comparison of the dietary intake and clinical characteristics of obese and normal weight adults. Nutr Res Pract 2011 Aug;5(4):329-336 [FREE Full text] [doi: 10.4162/nrp.2011.5.4.329] [Medline: 21994528]
32. The R Project for Statistical Computing. URL: https://www.R-project.org/ [accessed 2019-07-31]
33. Tidyverse: Easily Install and Load the Tidyverse. 2017. URL: https://tidyverse.tidyverse.org/ [accessed 2019-08-01]
34. Farage G, Gale1 C, Kocak M, Klesges RC, Talcott GW, Richey P, et al. Assessing the contribution of self-monitoring through a commercial weight loss app: mediation and predictive modeling. GitHub, Inc. 2021. URL: https://github.com/senresearch/dietappstudy [accessed 2021-06-23]
35. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale) 2016 Apr;6(2):- [FREE Full text] [doi: 10.4172/2161-1165.1000227] [Medline: 27274911]
36. Jolliffe I, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 2016 Apr 13;374(2065):20150202 [FREE Full text] [doi: 10.1098/rsta.2015.0202] [Medline: 26953178]
37. Ringnér M. What is principal component analysis? Nat Biotechnol 2008 Mar;26(3):303-304. [doi: 10.1038/nbt0308-303] [Medline: 18327243]
38. Lê S, Josse J, Husson F. FactoMinerR a package for multivariate analysis. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
39. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses Analysis. The Comprehensive R Archive Network. 2017. URL: https://cran.r-project.org/web/packages/factoextra [accessed 2019-08-16]
40. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. Polit Anal 2017 Jan 4;21(2):141-171. [doi: 10.1093/pan/mps040]
41. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593-614 [FREE Full text] [doi: 10.1146/annurev.psych.58.110405.085542] [Medline: 16968208]
42. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010 Dec;15(4):309-334. [doi: 10.1037/a0020761] [Medline: 20954780]
43. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. J Stat Soft 2014;59(5):1-38. [doi: 10.18637/jss.v059.i05]
44. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010 Dec;15(4):309-334. [doi: 10.1037/a0020761] [Medline: 20954780]
45. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiol Methods 2014 Jan;2(1):95-115 [FREE Full text] [doi: 10.1515/em-2012-0010] [Medline: 25580377]
46. Daniel RM, de Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics 2015 Mar 28;71(1):1-14 [FREE Full text] [doi: 10.1111/biom.12248] [Medline: 25351114]
47. Jerolon A, Baglietto L, Birmele E, Perduca V, Alarcon F. Causal Mediation Analysis in Presence of Multiple Mediators Uncausally Related. arXivstat. 2018. URL: https://arxiv.org/abs/1809.08018 [accessed 2020-06-12]
48. Jed WM, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T. Caret: Classification and Regression Training. The Comprehensive R Archive Network. 2018. URL: https://CRAN.R-project.org/package=caret [accessed 2020-09-19]
49. Varkevisser RD, van Stralen MM, Kroeze W, Ket JC, Steenhuis IH. Determinants of weight loss maintenance: a systematic review. Obes Rev 2019 Feb 16;20(2):171-211 [FREE Full text] [doi: 10.1111/obr.12772] [Medline: 30324651]
50. Peterson ND, Middleton KR, Nackers LM, Medina KE, Milsom VA, Perri MG. Dietary self-monitoring and long-term success with weight management. Obesity (Silver Spring) 2014 Sep 13;22(9):1962-1967 [FREE Full text] [doi: 10.1002/oby.20807] [Medline: 24931055]
51. Pellegrini CA, Conroy DE, Phillips SM, Pfammatter AF, McFadden H, Spring B. Daily and seasonal influences on dietary self-monitoring using a smartphone application. J Nutr Educ Behav 2018 Jan;50(1):56-61.e1 [FREE Full text] [doi: 10.1016/j.jneb.2016.12.004] [Medline: 29325663]

52. Swift DL, Johannsen NM, Lavie CJ, Earnest CP, Church TS. The role of exercise and physical activity in weight loss and maintenance. Prog Cardiovasc Dis 2014;56(4):441-447 [FREE Full text] [doi: 10.1016/j.pcad.2013.09.012] [Medline: 24438736]
53. Miller WC, Koceja DM, Hamilton EJ. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. Int J Obes Relat Metab Disord 1997 Oct;21(10):941-947. [doi: 10.1038/sj.ijo.0800499] [Medline: 9347414]
54. Pronk NP, Wing RR. Physical activity and long-term maintenance of weight loss. Obes Res 1994 Nov;2(6):587-599 [FREE Full text] [doi: 10.1002/j.1550-8528.1994.tb00110.x] [Medline: 16358405]
55. Maclin-Akinyemi C, Krukowski RA, Kocak M, Talcott GW, Beauvais A, Klesges RC. Motivations for weight loss among active duty military personnel. Military Medicine 2017 Sep;182(9):e1816-e1823 [FREE Full text] [doi: 10.7205/milmed-d-16-00380]

Abbreviations

ACME: average causal mediation effect
AHEAD: Action for Health in Diabetes
AUC: area under the receiver operating characteristic curve
PC: principal component
PCA: principal component analysis
TE: total effect

Edited by G Eysenbach, Q Zeng; submitted 16.03.20; peer-reviewed by A Beleigoli, A Andrade, BT Nezami; comments to author 28.08.20; revised version received 22.12.20; accepted 15.04.21; published 14.07.21.

Please cite as:
Farage G, Simmons C, Kocak M, Klesges RC, Talcott GW, Richey P, Hare M, Johnson KC, Sen S, Krukowski R
Assessing the Contribution of Self-Monitoring Through a Commercial Weight Loss App: Mediation and Predictive Modeling Study
JMIR Mhealth Uhealth 2021;9(7):e18741
URL: https://mhealth.jmir.org/2021/7/e18741
doi:10.2196/18741
PMID:34259635

Original Paper

A School-Based Mobile App Intervention for Enhancing Emotion Regulation in Children: Exploratory Trial

Bettina Moltrecht[1,2], PhD; Praveetha Patalay[3], PhD; Jessica Deighton[1], PhD; Julian Edbrooke-Childs[1], PhD

[1]Evidence-based Practice Unit, University College London & Anna Freud National Centre, London, United Kingdom

[2]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[3]MRC Unit for Lifelong Health and Ageing, University College London, London, United Kingdom

Corresponding Author:
Julian Edbrooke-Childs, PhD
Evidence-based Practice Unit
University College London & Anna Freud National Centre
4-8 Rodney Street
London, N1 9JH
United Kingdom
Phone: 44 020 7794 ext 2313
Email: Julian.childs@annafreud.org

Abstract

Background: Most mental health disorders are first experienced in childhood. The rising rates of mental health difficulties in children highlight the need for innovative approaches to supporting children and preventing these difficulties. School-based digital interventions that address shared risk factors and symptoms, such as emotion dysregulation, present exciting opportunities to enhance mental health support for children on a larger scale.

Objective: This study investigates the use of a new app-based intervention designed to support children's emotion regulation in schools. The aim is to optimize the usability, acceptability, and utility of the app and explore its scope for implementation with the target user in the school context.

Methods: As part of an interdisciplinary development framework, the app is being evaluated in a 3-month trial across 4 primary schools. In total, 144 children (aged 10-12 years) took part and accessed the intervention app in the classroom or at home. Outcomes regarding usability, acceptability, and implementation opportunities were assessed through digital user data, self-report questionnaires (132/144, 91.6%), and semistructured interviews with children (19/144, 13.2%) and teachers (6/8, 75%).

Results: The app usage data showed that 30% (128/426) of the users were returning users. Self-report data indicated that 40.1% (53/132) of the children had not used the app, whereas 57.5% (76/132) had used it once or more. Of the children who had used the app, 67% (51/76) reported that the app was helpful. Interviews with children and teachers suggested positive experiences with the app and that it helped them to calm down and relax. Children reported that they perceived the app as acceptable, usable, and helpful. In terms of the intervention's usability, most features functioned well; however, certain technical issues were reported, which may have led to reduced engagement levels. Teachers not only reported overall positive experiences but also discussed access difficulties and reported a lack of content as one of the main barriers to implementing the app. Having a web-based app significantly enhanced accessibility across devices and settings and provided teachers with more opportunities to use it. We identified the need for new, activating app features in addition to the existing, primarily relaxing ones. The findings indicated that it is possible to use and evaluate an app intervention in the school context and that the app could help enhance children's emotion regulation. We discuss areas for improvement regarding the app, study design, and future implementation strategies.

Conclusions: We share important insights with regard to the development, implementation, and evaluation of a new app for supporting children's emotion regulation in schools. Our results demonstrate that mental health apps represent a promising means to facilitate effective mental health service provision in and outside of the school context. Important lessons learned are shared to support other researchers and clinicians on similar journeys.

(JMIR Mhealth Uhealth 2021;9(7):e21837) doi:10.2196/21837

KEYWORDS

emotion regulation; digital mental health; mhealth; school intervention; child mental health; mobile phone

Introduction

Background

Approximately 10%-20% of children and young people worldwide experience mental health problems, making it one of the leading causes of disability in this population [1-3]. Considering the significant impact of mental health problems on other developmental outcomes (eg, academic achievement and physical health) [4,5], human and economic costs are substantial, calling for new, innovative approaches to support children and to prevent this problem. To address this challenge, this study explores the use of a newly developed web-based app as a universal school intervention to support children's mental health and well-being by strengthening their emotion regulation skills.

Schools have long been identified as an ideal setting to provide youth mental health support [6,7] because of the considerable amount of time children spend at school. Moreover, it has been suggested that implementing mental health interventions in schools can help overcome important social and environmental barriers to accessing mental health services, such as costs, family demographic factors, transport, and social stigma [8,9]. Evidence from 52 systematic reviews and meta-analyses has suggested that schools are key facilitators for the development and implementation of effective mental health interventions [10]. However, environment-specific challenges, such as increased pressure on teachers to meet academic targets, are pertinent to the implementation of new interventions in schools. Thus, interventions that require a significant amount of time and effort from teachers are often discarded [11]. Digital interventions provide an effective means to overcome such challenges, as they minimize costs, time, and personal resources in comparison with face-to-face interventions [11].

Digital Mental Health Interventions for Children and Young People

Digital mental health interventions have received increasing attention in recent years [12], with growing evidence supporting their effectiveness in clinical and community settings [13,14]. Owing to the rapid and consistent progress of technology, children and young people are increasingly adopting mobile apps, and their usage should therefore be considered in today's youth mental health provision. We built on these developments and designed a new web-based app to support children in the school context, which remains widely underexplored. To date, only a few app interventions have been developed and evaluated, specifically for children and young people [15,16], and even less so in the school setting, with some exceptions [17]. Early findings from this research indicated that school mental health apps were perceived as potentially useful. However, they also highlighted that future studies need to focus more on refining and adjusting digital interventions to meet young users' needs and make them suitable for the context in which they are being delivered [18]. In line with this, it was highlighted that an iterative approach of testing, evaluating, and refining is integral to the development and design of digital interventions.

Newer evaluation guidelines for digital interventions consistently emphasize that any early evaluation attempts should focus on optimizing a digital intervention to ensure adequate levels of usability, acceptability, and engagement before any feasibility or efficacy testing [19,20]. Given that effectiveness trials are highly resource demanding, it is recommended that researchers follow a staged and iterative evaluation process [21], particularly with digital interventions. Digital interventions come with an added layer of complexity due to the underlying technology, which needs to be tested for its stability and usability first, as any issue with the technology would otherwise have a tremendous effect on the intervention's implementation success and effectiveness as a whole [22,23]. Pursuant to this, the primary aim of this study is to optimize the existing app intervention by exploring its usability and acceptability from a user's perspective as well as the possibility of implementing and evaluating it within the school context.

A closer look at the digital mental health landscape indicates that most interventions aim to support specific mental health disorders or symptoms [6,13,18]. However, especially in children where comorbidity rates (ie, transitioning between disorders and showing symptoms from more than one disorder) are high, it has been shown that transdiagnostic interventions are not only highly beneficial for this group but also that traditional psychological interventions can be further improved by adding a transdiagnostic focus [24]. Furthermore, transdiagnostic interventions have also been regarded as highly suitable for the school context, as they reduce the burden on teachers having to deliver multiple, highly fragmented, targeted interventions [25].

Evidence suggests that emotion regulation is an important transdiagnostic mechanism that underlies a wide range of mental health disorders and predicts later levels of psychopathology. Emotion regulation can be described as the extrinsic and intrinsic processes through which individuals monitor, evaluate, and modify emotional reactions to accomplish their goals [26]. There has been an increasing interest in emotion regulation as a treatment and prevention target because of growing evidence demonstrating that emotion regulation difficulties are not only present in most mental health disorders but are also a significant risk factor for future mental health difficulties [27,28]. This is supported by a number of systematic reviews [29] and meta-analyses [30,31], which have indicated that positive changes in emotion regulation are associated with a reduction in anxiety, depression, substance abuse, eating, and borderline personality disorder symptoms.

However, most digital mental health interventions have been developed for specific disorders, thereby leaving a significant gap in technologies addressing transdiagnostic factors such as emotion regulation. Although this could ultimately support a wider range of mental health problems, it is also highly suitable for the school context [32]. Most existing digital emotion regulation interventions have focused on emotion regulation deficits specific to either autism spectrum disorders or attention-deficit/hyperactivity disorder [33,34]. Although it has been shown that these interventions could potentially improve some of the specific deficits [35-37], the interventions were specifically designed for this unique population, thereby making

it difficult to transfer them to a wider population. A very limited amount of research has explored the use of mobile apps to enhance emotion regulation in young people more generally [38] or as a means to prevent difficulties from arising. One exception is the new music app by Hides et al [38], which aims to enhance emotion regulation in adolescents and young adults. Their initial findings based on 169 young people (age=16-25) suggested that the app could potentially enhance emotion regulation in this population; however, further testing is required to determine its effectiveness.

To the best of our knowledge, there is currently no app intervention that targets emotion regulation in late childhood or preadolescence, despite an increasing number of scholars highlighting this period as a critical developmental stage to achieve maximum effect [39] in terms of youth mental health prevention. More specifically, past evidence has indicated that the median onset for most mental health disorders lies at the age of 11 years [40], with the majority reaching a peak during adolescence [41]. Recent developmental models suggest that adolescents are particularly prone to developing social-emotional disorders, which are triggered by a combination of developmental changes and preadolescent risk factors related to emotion regulation [39]. It has been argued that developmental changes hamper important emotion regulation processes, which subsequently lead to increased mental health difficulties later in adolescents [42]. Moreover, in the United Kingdom and many other countries, late childhood is marked by a transition period in which children have to leave primary school to enter secondary school. This period is frequently experienced as highly stressful by children [43], thereby supporting our case to strengthen children's emotion regulation before this significant transition period.

Taken together, we believe that by addressing emotion regulation during childhood, our app intervention presents a promising means not only to prevent mental health difficulties from arising but also to serve as an early intervention tool for children who might already be experiencing such difficulties.

This Study

There is a significant lack of digital mental health interventions for children that target important transdiagnostic factors, such as emotion regulation. The development of an acceptable, usable, and engaging emotion regulation app, which can be implemented successfully in the school context, will be highly beneficial for supporting children's mental health on a larger scale. Therefore, this study explores and evaluates the use of a new emotion regulation app for children with the primary aim of optimizing it further and informing future development stages with the ultimate goal of making it highly suitable for the user group and context [18].

The following research questions are addressed:

1. How acceptable and usable is the app from the children's perspective in the school context?
2. How do children interact and engage with the app at school?
3. What are the perceived barriers to and facilitators of implementing and delivering the intervention in the school context?
4. How can the existing app intervention be further improved?
5. What possibilities, facilitators, and barriers exist in terms of evaluating the app in the school setting?

Methods

Recruitment and Participants

The study was advertised on the Anna Freud National Centre for Children and Families organization's website and in a newsletter that was sent out monthly to a network of schools and related organizations across the United Kingdom. Newsletter recipients had previously signed up to receive newsletters. Initially, 19 schools indicated an interest in participation. Of the 19 schools, 11 were primary schools and were invited to an initial phone call. During the initial phone call, we discussed the research project and intervention as well as the schools' involvement if they agreed to take part in it. Only primary schools in the United Kingdom with access to tablets and wireless internet were eligible. Following the initial phone call, 4 schools were excluded: 3 were not primary schools and one had no phone or tablet policy. Three other schools stepped down before the start of the trial for the following reasons: (1) the research aspect of the intervention would take up too much time, (2) for a large percentage of parents, English was not their first language; hence, they struggled to understand the consent forms or information sheets, and (3) a lack of parental engagement.

Ultimately, 4 primary schools participated in this trial. Only children between the ages of 10 and 12 years with parental consent and child assent were eligible to participate, which resulted in data of 144 children at baseline and 132 children postintervention. Children's ages ranged between 10 and 11 years (mean 10.5, SD 0.49). Of the total sample, 56.3% (81/144) indicated that they were White, 6.9% (10/144) were Black, 18.8% (27/144) were Asian, 15.3% (22/144) were mixed, and 2.8% (4/144) chose *other* as their ethnicity. Of the 144 children taking part, 166 (81.1%) children indicated that English was their first language, 25 (17.4%) children were not, and 2 (1.4%) children preferred not to provide that information. A total of 6 female teachers participated in postintervention interviews; 2 teachers from 1 school were not available because of their illness.

Intervention

Intervention Development and Design Process

The app has been designed as a school-based universal intervention for children at the end of their primary school years (age: 10-12 years). The app is considered as a complex mental health intervention as it involves a set of multiple, interconnected, and interacting components [44,45]. As part of the development process, we created a new design and development framework based on the Medical Research Council (MRC) framework [44,46], the Patient-Clinician-Designer Framework [47], and the co-operative inquiry framework to co-design with children by Druin [48].

At this stage, we only focused on the first three phases of the MRC framework: theory, modeling, and exploratory trials. The theory stage concerns the exploration of relevant theory and a

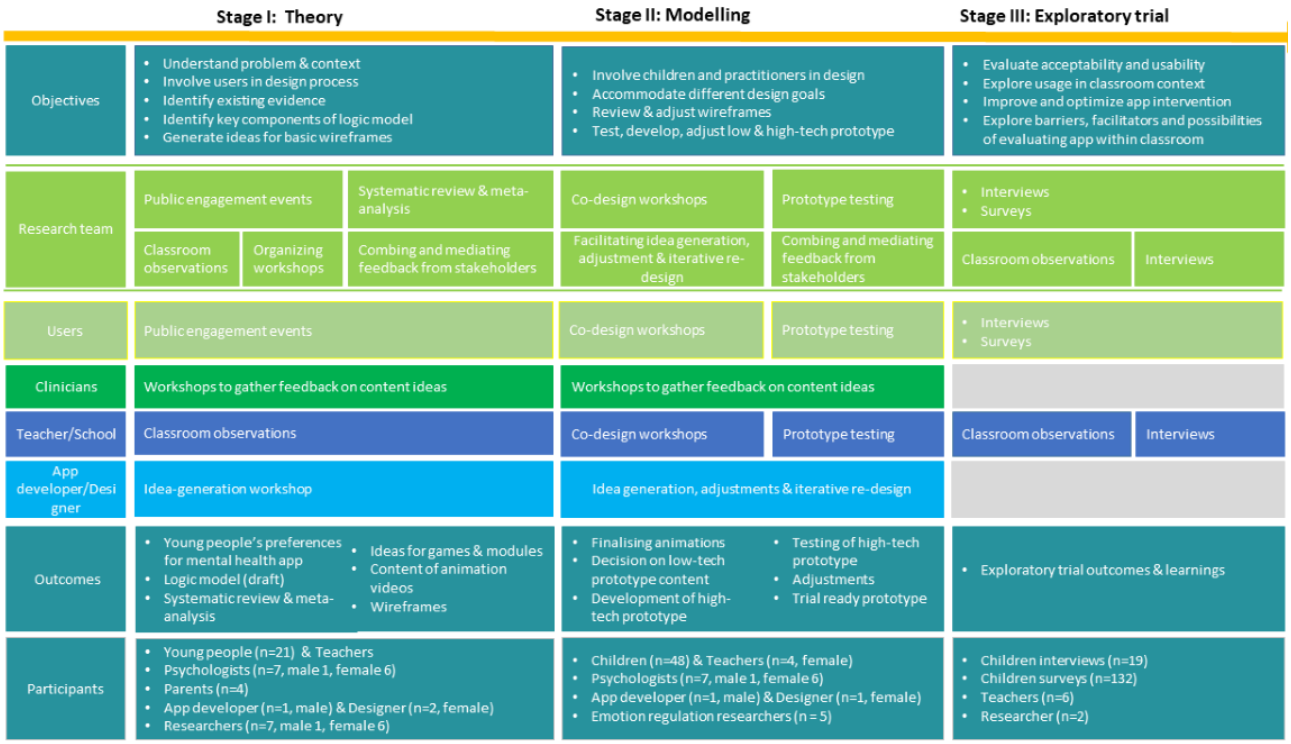
review of existing evidence to ensure that the most reliable intervention components are chosen. The intervention modeling stage suggests that the researchers focus on identifying potential underlying mechanisms that influence the preferred outcome, which are then included in the intervention. Following the design of the initial intervention, the researcher is advised to explore its components further through exploratory trials to identify constant and variable components, including acceptability of intervention, compliance, delivery, recruitment, and retention rate.

The MRC framework provides valuable guidelines for the development and evaluation of complex interventions; however, it provides little information with regard to the actual design of appropriate content [49]. Hence, we drew on two other frameworks rooted in the field of human-computer interaction and design. The Patient-Clinician-Designer Framework provides guidance on the design structure and content creation process of digital interventions for mental illness [47]. It aims to meet the complex requirements when designing user-centered interventions for mental illness by considering different perspectives (eg, in this study, teachers, parents, and children) and design goals.

As our main target group is children, we also drew on the co-operative inquiry framework by Druin [48], which provides specific techniques on involving young users in the design process of technologies. The co-operative inquiry framework highlights the importance of involving children as partners in the whole process, instead of merely letting them test the almost finished prototype or end product. Druin [48,50] also emphasizes on the benefits of conducting fieldwork (ie, *contextual inquiry*), especially when working with children, which allows researchers to detect relevant contextual information, including patterns of activities, ways of communication and other artifacts. In addition, it has been reported that discussing design features in the relevant context (eg, school or home) makes it easier for children to express ideas and provide suggestions.

By combining the three frameworks from different fields, we hope to ensure a truly interdisciplinary approach to developing the app, the lack of which has been frequently criticized in many digital interventions [16,51]. Our development framework consisted of three stages, with each stage using a unique set of objectives, methodologies, and stakeholders (Figure 1). A more detailed description of the development and design process was published as a preprint by the first author (BM) and can be found in a study by Moltrecht et al [52].

Figure 1. The 3-stage interdisciplinary development process, including research activities and stakeholders.



Intervention Description

During the onboarding process, users were presented with a video that explains the purpose of the app and its usage. Next, the user was guided through a process to set up an account and then select a preferred color scheme and profile picture.

Once the user entered the home screen, they could choose from four modules: play (including four games), relax (includes mindfulness and relaxation exercises), watch (includes psychoeducational animations), and tools (includes a list of emotion regulation strategies), which provide users with the opportunity to learn, practice, and develop new emotion regulation skills (Figure 2, Figure 3). The content is presented through audio tracks, images, animated films, and games. A more detailed presentation of each module can be found in Multimedia Appendix 1 as well as the following preprint, which was published in a study by Moltrecht et al [52].

Figure 2. Home screen of the app with 4 main modules and the activated digital agent showing the "tell me something" and "check-in" function.

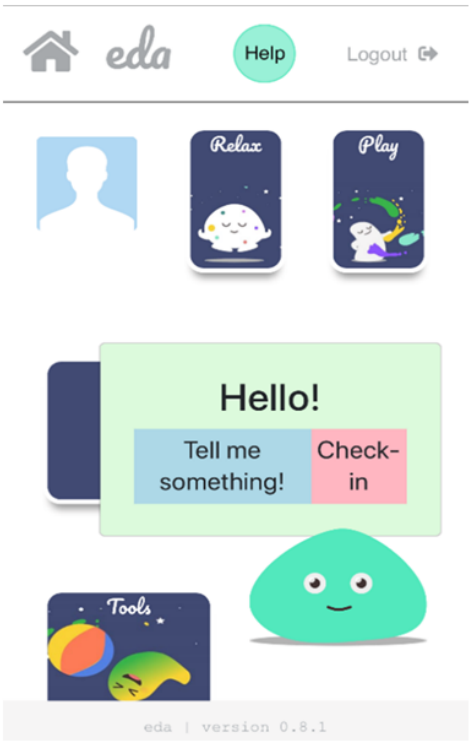
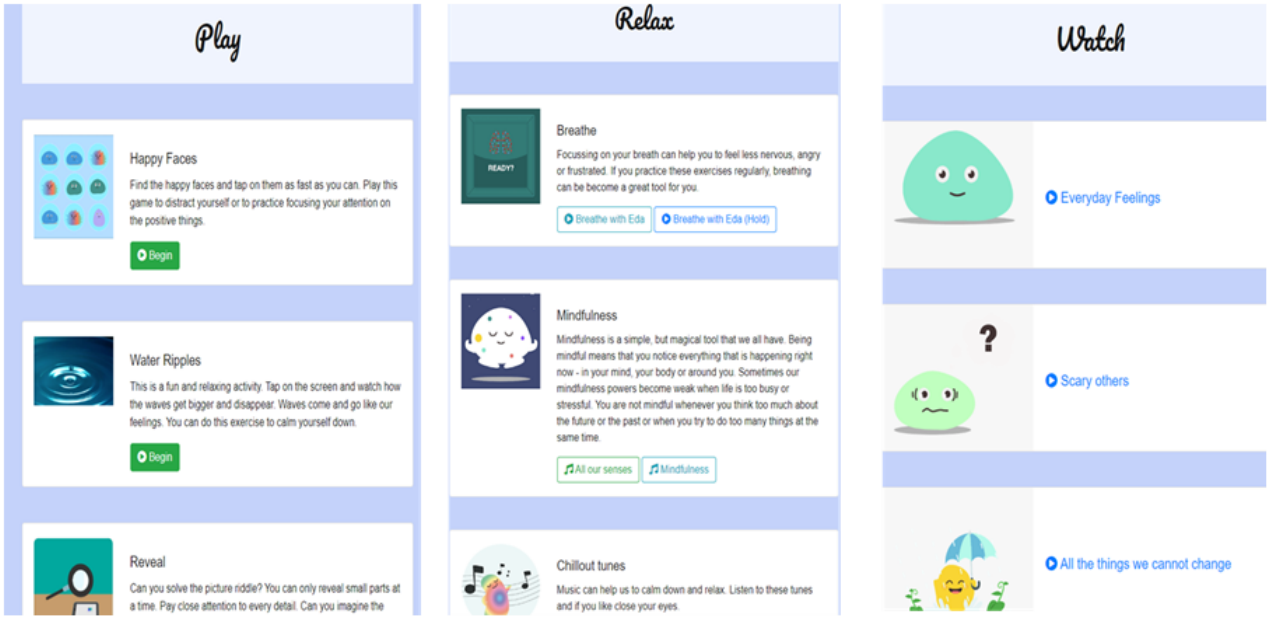
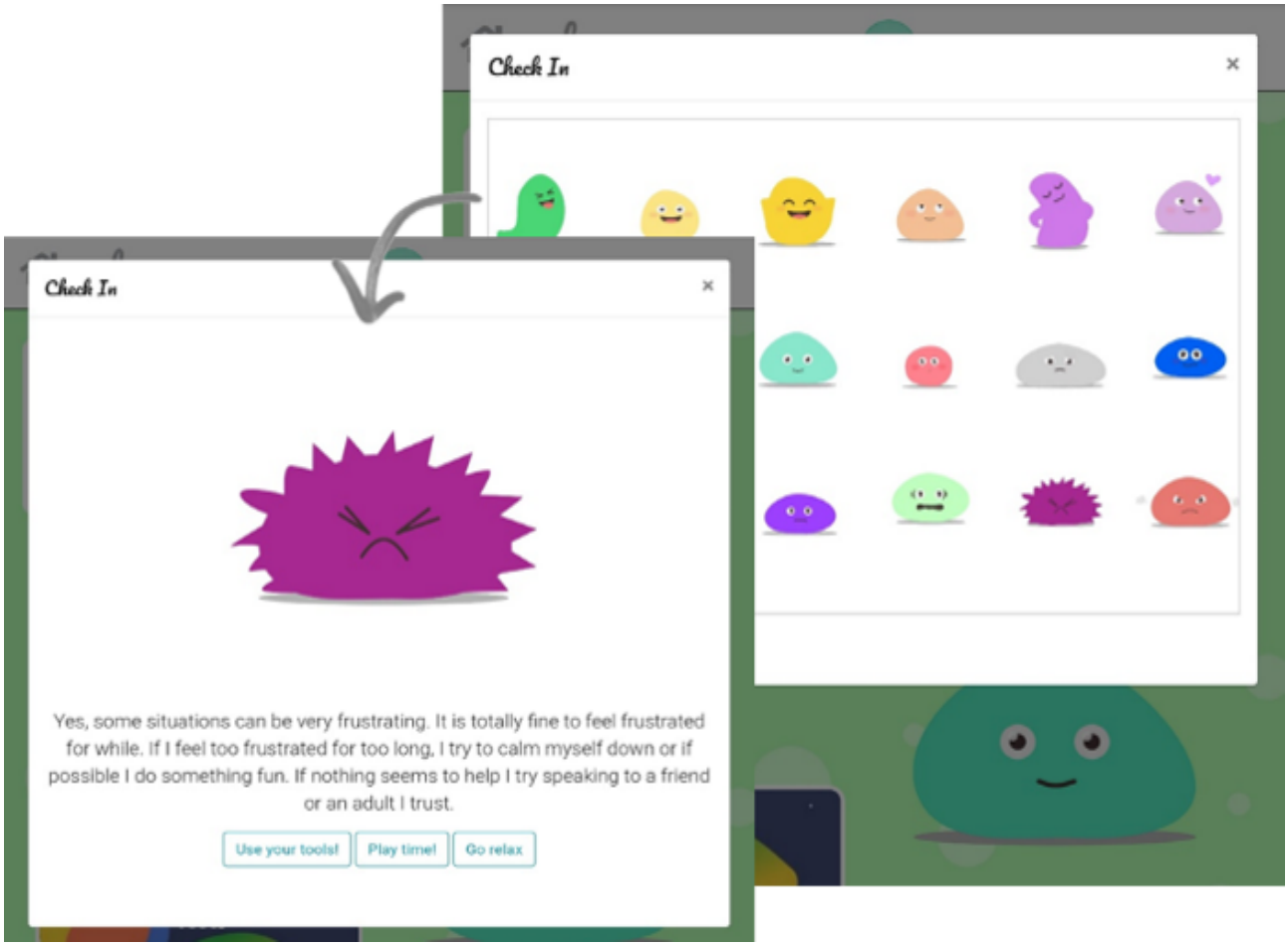


Figure 3. In-app content of the "play", "relax", and "watch" modules.



An animated agent is located at the bottom-right corner of the home screen and opens two more features when the user taps on it. These features are (1) *tell me something*, which activates jokes and funny facts with the aim to increase levels of engagement; and (2) a *check-in* function, where users can select from a range of emotions about how they feel and are subsequently provided with more information about the particular feeling and suggestions for potentially helpful strategies (Figure 4).

Figure 4. Check-in function in the app with "feeling frustrated" being selected.



Teachers and children were instructed to freely explore different ways of using the app intervention. By providing them with the link to the app, it was also possible for children to explore the app outside the school context if they wanted to. This flexible approach was adopted so that children and teachers could use the intervention in their preferred ways and hopefully perceive it as less of a burden. Furthermore, we expected this to increase our understanding about app usage and implementation in future trials.

Technical Specifications

The intervention was developed as a responsive web-based app, which was believed to increase the accessibility of the app, as it allowed users to access it across different mobile devices, as well as desktop computers and smartboards. Although it worked across multiple platforms, it was optimized for tablets, as young children are more likely to have access to tablets at school and at home [53].

The app is delivered through the browser, meaning over-the-wire updates can be pushed out instantly, and the app uses advanced HTML5, CSS3, and JavaScript (ES6) techniques to render a smooth and performant user experience. The underlying development platform used was Meteor.js, a full-stack Node.js application development framework, hosted on a resilient AWS EC2 (Amazon Elastic Compute Cloud) instance with a MongoDB database hosted via MongoDB Atlas. The app only requires an internet connection when users access it for the first time, after which it can be saved to the home screen of the device. This feature was chosen to mitigate the risk that the intervention could not be accessed when schools had reduced or limited Wi-Fi infrastructure. The app does not store any individual user data and adheres to the existing general data protection regulations.

Measures

Demographics

Children reported age, gender, ethnicity, and their primary language spoken.

Internalizing and Externalizing Symptoms

The Short Mood and Feelings Questionnaire consists of 13 items that assess depressive symptoms in children and adolescents [54]. The Short Mood and Feelings Questionnaire has been shown to have good construct and internal validity across clinical (Cronbach α=.85) and community samples [55]. Furthermore, 5 items (“I get very angry,” “I lose my temper,” “I hit out when I am angry,” “I do things to hurt people,” and “I break things on purpose”) of the Me & My Feelings Questionnaire were added to assess externalizing symptoms [56]. The Me & My Feelings Questionnaire was developed as a self-reported mental health measure for the school setting and has been shown to have good psychometric properties across clinical and community samples (Cronbach α=.78-.82) [56,57]. Items on both scales were rated on a 3-point Likert scale, ranging from “Not true” (1), “Sometimes” (2), and “True” (3).

With both scales combined, scores ranged between 18 and 54, with higher scores indicating greater levels of symptoms.

Well-being

The Satisfaction with Life Scale for Children [58] was used to assess participants' personal perceptions of their well-being and satisfaction in life. The Satisfaction with Life Scale for Children consists of five items that are rated on a 5-point Likert scale, ranging from "disagree a lot" (1) to "agree a lot" (5) and has been reported to have good psychometric properties (Cronbach α=.84 [58]). Higher scores indicated greater levels of satisfaction with life and well-being.

Emotion Regulation

The How I feel - Questionnaire is a multidimensional self-report scale to assess emotional arousal and regulation abilities in children. It consists of 30 items rated on a 5-point Likert scale ranging from "very true of me" (5) to "not at all true of me" (1). The items assess the frequency, intensity, and regulation of five different emotions: sadness, fear, anger, happiness, and excitement. There are three subscales: (1) a positive emotion subscale, where higher scores indicate that happiness and excitement are experienced with high frequency and intensity; (2) a negative emotion subscale with high scores indicating that fear, anger, and sadness are experienced with high frequency and intensity; and (3) an emotion regulation subscale, where high scores reflect a strong ability to regulate the frequency and intensity of either positive or negative emotions.

The How I feel - Questionnaire has been reported to be a reliable and valid measure (Cronbach α=.84-.90) for research and for school interventions targeting pupils' emotion regulation [59,60].

Engagement

Engagement data were collected through Google Analytics and paper questionnaires, in which children were asked how often they had been using the app in the past 3 months.

Teachers were provided with a logbook, which they were asked to complete on a weekly basis to describe how often and in what way they used the app in their classrooms. Items in the logbook included (1) what was the average time or preferred way to use the app, (2) how did children engage with the app (eg, tablet or smartphone), (3) how often was the app used due to classroom disruptions, and (4) average time to reinstate children into the class after app use?

Usability and Acceptability of App and Evaluation Measures

To increase our understanding of the usability and acceptability of the app intervention, we conducted brief semistructured interviews with teachers and children after the 3-month intervention phase. A detailed interview schedule can be found in Multimedia Appendix 2. The interview assessed (1) what children and teachers thought about the app; (2) whether it was easy to use; (3) what aspects of the app they found helpful or unhelpful; and (4) how they used the app, what aspects of it, and in what situations.

The research team also explored the usability of the evaluation methods used. The researchers were present when the children filled in the questionnaires to observe if they experienced any difficulties when completing them. Furthermore, the research team kept a logbook of significant events and conversations with teachers or any reported difficulties during the study.

Study Design and Procedure

The University College London Research Ethics Committee approved this study (approval number: 7969/001).

Schools signed a memorandum of understanding, which explained the nature of the project and outlined the timeframes and responsibilities of the research team and the school. Parental consent, child assent forms, and parent and child information sheets were sent to the school and distributed by the class teacher. The research team visited the schools on the first day of the intervention to collect parent consent forms and child assent forms. Parents who indicated on the form that they had any remaining questions, where contacted by the research team to answer any remaining questions and obtain their oral consent over the phone, which was audio-recorded. Questionnaires were distributed to all participating children and the app was introduced to the class. After 6 weeks, the research team contacted the school to discuss the use of the app and any difficulties. Following a 3-month intervention phase, questionnaires were distributed again, and a researcher visited each classroom to observe the use of the app. Following this, semistructured interviews with 19 children and 6 teachers were conducted. Some teachers either had spoken to the children beforehand if they wanted to take part in the interviews or had asked the whole class in the presence of the researchers who would like to tell the researchers more about using the app at school. The research team highlighted to all children and teachers that honest answers were the most helpful and that they should not feel shy to report any negative experiences. All interviews with the children were audio-recorded with encrypted dictaphones and later transcribed. Owing to logistical issues, teacher interviews were not audio-recorded, and answers were written down by a researcher during the interview.

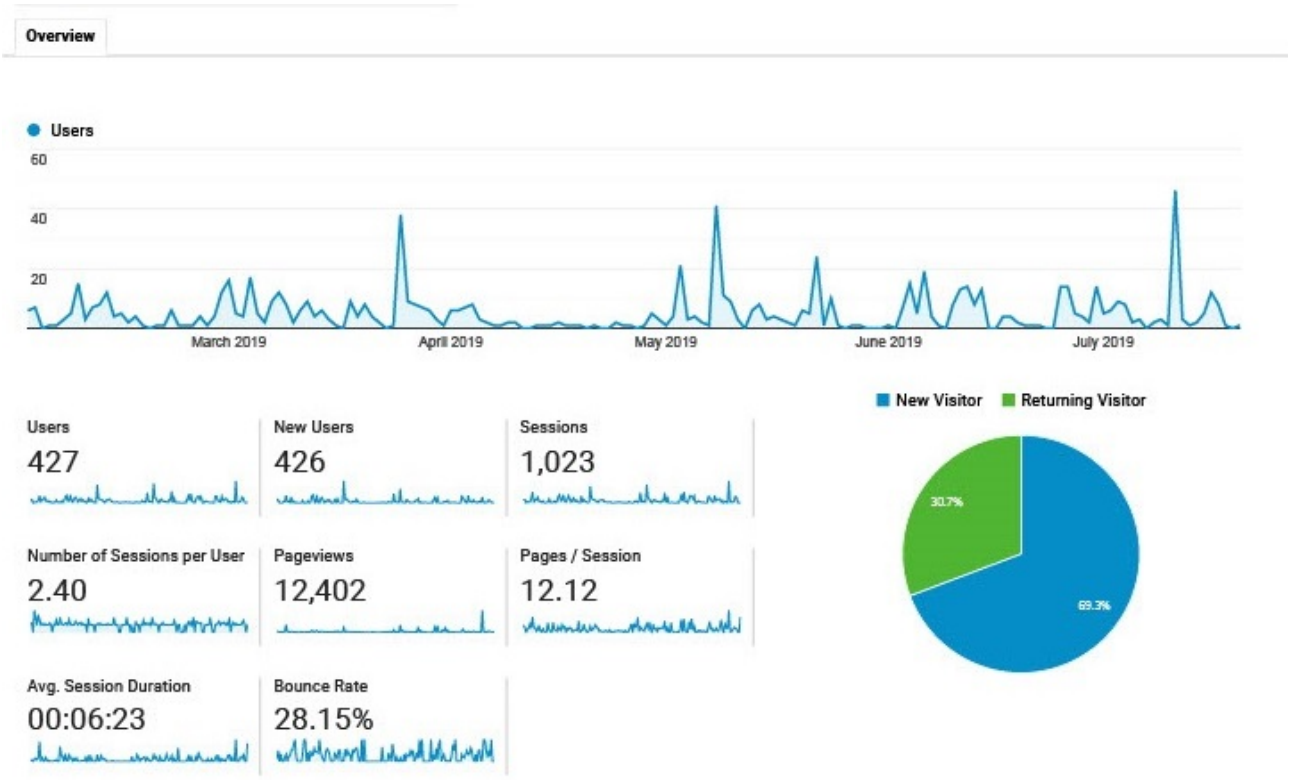
Analytic Strategy

Quantitative Data

Quantitative data from the questionnaires were used to calculate descriptive statistics for the baseline and postintervention assessments using SPSS (IBM Corporation).

Google Analytics data are presented below (Figure 5) to show overall usage and engagement.

Figure 5. Engagement data as derived from Google Analytics.



Qualitative Data

The transcribed interviews and notes taken during the interviews were analyzed using thematic analysis [61].

Thematic analysis is a flexible method that can be used to analyze qualitative data by identifying patterns in the data. In this study, no existing framework was used but patterns were identified with the specific research questions regarding usability, acceptance, user-intervention interaction, and implementation in mind. Braun and Clarke [61] outlined 6 steps as a structured but flexible way to conduct thematic analysis (the article by Braun and Clarke [61] provides a detailed description of the six steps).

Results

Participants, Recruitment, and Retention

In total, 19 schools indicated an interest in taking part, of which we assessed 11 schools for their eligibility. Seven of these schools met our criteria and were eligible for participation. Of the seven eligible schools, 3 schools stepped down before the trial (see the reasons provided in the *Recruitment and Participants* section). In the final study, which included 4 primary schools, a 57% (4/7) retention rate was achieved.

In total, 144 children (female: n=79; male: n=62; not specified: n=3) completed the surveys at baseline and 132 children completed the surveys postintervention, thereby resulting in an attrition rate of 91.6% (132/144).

Six out of eight teachers, all female, participated in the postintervention interviews. None of the teachers completed the weekly usage logbook.

Mental Health and Emotion Regulation Measures

The mean scores and SDs for pre- and postintervention assessments are presented in Table 1. Between 87.5% (126/144) and 91.6% (132/144) of the children completed the measures at baseline and follow-up. Although the mean scores and SDs are comparable with those found in other population and community samples [56,60,62], children in this study reported some difficulties with items on how they feel. Some children were unsure about the difference between items asking them about *strong* feelings and items asking them about *powerful* feelings.

Table 1. Descriptive statistics for mental health and emotion regulation questionnaires (N=144).

Outcome	Participants, n (%)	Score, mean (SD)
SMFQ[a] baseline	144 (100)	25.09 (6)
SMFQ postintervention	132 (91.6)	24.22 (5.99)
SWLS-C[b] baseline	140 (97.2)	4.13 (0.79)
SWLS-C postintervention	126 (87.5)	4.13 (0.83)
HIFQ-PES[c] baseline	144 (100)	3.87 (0.84)
HIFQ-PES postintervention	132 (91.6)	3.72 (0.87)
HIFQ-NES[d] baseline	144 (100)	2.11 (0.79)
HIFQ-NES postintervention	132 (91.6)	1.88 (0.72)
HIFQ-ERS[e] baseline	144 (100)	3.37 (0.83)
HIFQ-ERS postintervention	132 (91.6)	3.35 (0.87)

[a]SMFQ: Short Mood and Feelings Questionnaire. Lower scores indicate fewer internalizing and externalizing difficulties.

[b]SWLS-C: Satisfaction With Life Scale. Greater scores (maximum total score=5) indicate greater satisfaction with life.

[c]HIFQ-PES: How I feel - Questionnaire–Positive Emotion Scale. Greater scores indicate a greater frequency and intensity of happiness and excitement.

[d]HIFQ-NES: How I feel - Questionnaire–Negative Emotion Scale. Greater scores indicate a greater frequency and intensity of sadness, anger, and anxiety.

[e]HIFQ-ERS: How I feel - Questionnaire–Emotion Regulation Scale. Greater scores indicate the greater regulation of the frequency and intensity of emotions.

Engagement

In total, 57.5% (76/132) of all children from all schools indicated that they had used the app at least once, whereas 40.1% (53/132) indicated that they had never used the app in the past 3 months. At postassessment, 7.6% (10/132) of the children reported that they had used the app weekly. Of the 76 children who indicated that they had used the app, 67% (51/76) said that they found it helpful, and 58% (44/76) said that they would recommend the app to a friend.

No data were gathered through the teacher logbooks; therefore, we were unable to analyze any quantitative data on average usage times per classroom over the 12 weeks or in what way the app was engaged with over time.

Data collected via Google Analytics indicated that 426 users had accessed the website, of which 30% (128/426) were returning users and 70.2% (299/426) were new visitors. Furthermore, the average time spent on the app per session was 6 minutes and 22 seconds and the *play* module was most frequently visited, followed by *relax* and *watch*.

Qualitative Outcomes

Children's Reported Usability and Acceptability

In total, 19 children shared their experiences using the app during the interviews. Most children reported positive experiences with the app and provided insights regarding specific strengths and weaknesses.

Feeling Calm and Relaxed

Nearly all children reported that using the app made them feel calm and relaxed. They indicated using the app, especially during stressful times (eg, *test at school, argument with friend or sibling, or having a bad day at school*). Some children reported using the app to fall asleep at night. Children seemed to enjoy two features in particular, the *Water Ripples* game and the *music* function:

> *I think the app's helpful, um, if you're stressed or if you like, it's a good way to relax and you can use it to calm down.*

Helpfulness

Most children reported that they found some of the features (ie, *check-in* function, videos, and tools list) particularly helpful and useful:

> *...the thing I also found was quite helpful was where you could sort of tell how you were feeling, and it sorta gave what you should do.*

They explained that it increased their understanding or knowledge of their feelings and provided suggestions regarding possible solutions or actions to take:

> *...like it makes you understand it, your feelings are something in you and it's ok to have them.*

Design and Technical Issues

With respect to the app's limitations, the children most frequently reported technical or design issues. Most commonly mentioned were problems (eg, "didn't work," "too slow," or "took too long to load") related to the *Reveal* game and video clips, which appeared to happen more frequently on some devices than others (this was assumingly related to certain web-browsers). Furthermore, children mentioned that they would prefer to have more options to choose from for the color scheme and design of the home screen and other personalization features.

Feelings of Anger

A few children reported that they found the app to be less helpful when they were very emotional or experiencing strong feelings of anger:

> *I liked the app. Although it did not help me with my anger, no one could help me with this yet.*

In relation to this, one child explained that when they were angry, they preferred to "do something to kind of get it out" instead of engaging in calming or relaxing exercises.

Children's Interaction With the Intervention

Children reported different preferences for where, when, and how to use the app.

Location

Most children used the app at school where it was introduced to them. Some children reported using the app primarily during times when teachers allowed them to choose an activity for a certain amount of time. Other children said that they had agreed with their teacher to use it in certain situations when they struggled to concentrate or participate in class. Almost half of the children (9/19, 47%) reported that they also used the app at home.

Emotional Prompts

Many children suggested that the app is most suitable during stressful times or when someone struggles with their feelings. They provided examples, including having a fight with someone, not being able to concentrate, or feeling bored:

> *The best way to use the app is if you're stressed out, or um, if you need something to take your mind off something.*

Children reported less frequent use of the app during less stressful times and when they felt generally happy:

> *It's not for someone who is happy, but some people get a bit angry sometimes, I'd recommend it to them.*

Design Prompts

Although most children reported that the app was easy to use ("I just knew how to use it"), it became apparent that some features, such as the help function, had not been accessed or had not been discovered by them. Furthermore, children reported that they had forgotten about the app when they had not used it for a while.

Access Barriers

As the app was web-based, many children were not able to find or download the app through the app store, which they reported was their primary way to access apps. Hence, this was one of the major barriers to accessing the app. One child said, "I couldn't remember what it was called, so I couldn't find it."

Furthermore, in relation to the school setting, children reported difficulties accessing tablets as they were either not permanently available or locked in a drawer, so they had to ask for it. The latter was also perceived as a barrier, as children were too shy to request it and "didn't want to ask the teacher for it."

Teacher-Reported Facilitators to and Barriers of Implementation and Delivery

Access Barriers

Teachers also mentioned that the app was hard to find if the link was not available or in reach, which seemed to inhibit the use of the app from the teacher's perspective.

Technical Issues

Teachers reported experiencing technical issues of video crashing or content taking too long to load, which was perceived as a barrier to using it.

Flexible Use

Every teacher reported a slightly different way of using the app in their classroom, with some preferring a whole-class approach and others directing individuals to the app. The freedom to use the intervention in different ways was perceived as a facilitator, as it made it easier for them to find opportunities to use the app in the classroom.

Compatibility With Teaching Style

Teachers were more likely to use the app if they were compatible with their existing teaching methods and did not require additional work or adjustments. Moreover, teachers liked that they could direct children individually to the app, when needed, and that it "doesn't take away too much time from the teaching" or interrupts the classroom atmosphere.

Furthermore, in classrooms where mindfulness and relaxation exercises were already used in other ways, teachers reported to primarily use the relaxation module by projecting the exercises on the smartboard or playing the music, which seemed to "[help] to calm them [the children] down during work times."

In relation to this, teachers provided further app suggestions to support their teaching (eg, a timer and noise meter with a traffic light system to signal children when they are too noisy) and could therefore further facilitate the implementation of the app.

Teachers who did not see suitable opportunities to integrate the app in their teaching reported that it took them some time to "remember the app" and that there was a tendency to rely on "old habits" or methods in difficult situations ("in the heat of the moment"). They also reported feeling confident that using the app could become a habit.

Recommendations for Further Improvements

Content and Functionality

Both children and teachers reported that they would like more content. Some children reported that they "got a bit bored" by having to play the same game. Others mentioned that there were "only four videos" to watch, which resulted in decreased interaction with the app over time. In line with this, teachers indicated that it was more likely that they continued using the app if there were updates that were more frequent, including new content.

Furthermore, children reported a wish for more features, such as making the embodied agent more responsive, so that more

interactions are possible. One child compared it with "a robot that you can talk to."

Individual Needs

Some teachers mentioned that *activating* features, would be helpful "for children with too much energy," "when they are angry," or after "a long time of sitting." One teacher suggested the use of dance videos as an activating exercise during breaks.

In addition, teachers saw a need for more interventions that specifically support children with learning disabilities or autism spectrum disorder, as they seem to be more likely to experience specific emotion regulation difficulties associated with the disorder.

Discussion

Acceptability and Usability of the App

Through questionnaires and postintervention interviews, the present study collected data on the perceived usability and acceptability of the app. The interviews suggest that children perceived the app as acceptable, usable, and helpful. The interviews provided preliminary evidence that the app helped children to calm down and relax in stressful situations, and potentially increased their understanding and knowledge of emotions. However, this needs to be thoroughly tested in future studies.

Some children reported that they found the app to be less helpful when they experienced anger. This suggests that there is a need for different types of support with respect to different emotional experiences. A similar idea was suggested in past research with infants, whereby certain strategies (ie, distraction) were more effective in regulating anger than fear [63].

In terms of the intervention's usability, most parts of the intervention functioned well; however, certain technical issues were reported that led to reduced engagement levels. These issues must be addressed before future evaluations.

Interaction and Engagement

Children reported that they used the app primarily at school, whereas others accessed it at home. Most indicated that stressful situations were one of the main motives for accessing the app. Children and teachers did not receive specific instructions on how to use the app; therefore, differences occurred between classes, with some teachers directing certain children to the app in a special area in the classroom, whereas others used it primarily with the whole classroom. Children were provided with a link to the app at school and they could access it at home if they followed the same link. Although this made it more difficult to exactly track usage, the open approach helped us understand how and when children used the app depending on the context (eg, listening to music to concentrate in class vs listening to music to fall asleep at home).

Engagement data from Google Analytics suggested that 30% (128/426) to 37.1% (158/426) of the users repeatedly accessed the app over a 3-month period. Although this number would ideally be higher, it is similar to adherence rates reported for other mental health apps [64].

In an attempt to mitigate low levels of engagement, one of the most common limitations in digital health interventions, we involved children and young people throughout the development and design of the app intervention. Interviews with the children indicated that the *Water Ripples* game was perceived as one of the most positive and helpful features in the app, which supports our interdisciplinary development approach.

With respect to children's and teachers' requests for having more content updates to maintain the level of novelty, future research could explore the use of timed updates, whereby sections of the content are released one after another. We recognize that this was only one piece of the puzzle. Issues surrounding user engagement are a recurring topic in the field, posing the question of how much engagement is actually needed for an intervention to be effective. The data from this study do not provide sufficient evidence to determine this, and further research is required.

Nevertheless, based on the children's reports, it can be assumed that some of the features positively influenced user engagement, such as the digital agent, which should be further explored (eg, chatbots and getting clothes or objects as rewards to change its appearance).

Delivery and Implementation

Teachers play a significant role in the intervention's implementation and delivery. Therefore, we tried to gain insights into potential barriers to and facilitators for implementing the app in a school setting.

Although teachers reported positive experiences, they also reported access difficulties and lack of content as the main barriers to implementing the intervention. Our findings suggest that teachers are more likely to use the app if they are compatible with their existing teaching methods. Teachers who saw fewer natural opportunities to integrate it reported more barriers to use the app. This is in line with previous research findings [18] and highlights the importance of considering teachers' perspectives when agreeing on design goals and intervention features. At the same time, teachers' requests and any subsequent design implications will have to be carefully considered in relation to the intervention's overall goal (ie, supporting children's emotion regulation) and any conflicting interests with young users' requests need to be mediated. For instance, although some teachers suggested features such as a noise meter, the research team needs to explore whether such a feature would primarily benefit teachers (eg, as a classroom management strategy) or young users as well.

Another barrier related to the school environment was the small number of tablets available per class, which limited the accessibility of the intervention. In addition, some schools only provided access to tablets upon request, thereby limiting ease of access. However, the possibility of accessing the app through other devices, such as computers or smartboards, enhanced the general uptake of the intervention in the school context.

Possibilities, Facilitators to, and Barriers of Conducting a School-Based App Evaluation

Important conclusions can be drawn regarding the possibility of evaluating the app in a school setting. In terms of school recruitment, we retained 57% (4/7) of the originally recruited schools, which could be further improved. Schools reported that they feared that their resources were limited to facilitate the research. Furthermore, some schools reported that there were significant issues due to unmet translation needs, which suggests that translated information sheets and consent forms should be made available in a future trial. Some schools also mentioned that it would have been helpful if the research team had provided guidance that was more specific on the app usage or if the research team had explored different means to use the app with the teachers beforehand.

In terms of the measures used, a high number of questionnaires were completed at baseline and follow-up (baseline: 126/144, 87.5%; follow-up: 132/144, 91.6%), thereby suggesting acceptable completion rates. However, none of the teachers completed the weekly logbooks, and it was not possible to record any of the interviews with the teachers, which highlights specific barriers in terms of data collection from teachers.

Despite the identified barriers, the findings suggest that it is possible to implement and evaluate the app in a school setting. However, we suggest that a comprehensive feasibility trial is conducted next to ensure (1) an enhanced recruitment and assessment strategy, (2) improved integration of the app intervention in the school curriculum and teaching methods, and (3) easier access to the app by making it available on the app store. In line with this, we also suggest that a set of feasibility criteria is defined beforehand, so that informed decisions can be made as to whether an effectiveness trial is the next appropriate step [65].

Considerations for Future App Features and Research

In addition to the suggestions above, we would like to share further lessons learned, which will hopefully help improve the present and similar school-based app interventions.

The School Setting

With respect to one of the primary design goals and the prioritization of engagement, we opted for a multimedia app that included various audio and video materials. However, this partly presented itself as unsuitable for the school environment, as sounds can be disturbing or require access to headphones. This observation emphasizes that new types of interventions are accompanied by new challenges, which need to be explored further and taken into account.

Specific Emotion Regulation App Features

On the basis of previous research, we included a range of mindfulness and relaxation exercises [66], which children experienced as positive. Although mindfulness and relaxation apps have gained increased popularity, this study shows that for certain negative emotions, *activating* features instead of calming features could be equally important. These features seem to tap into a different set of emotion regulation strategies (eg, behavior activation and physical activity), which should not be neglected [67,68]. The finding that some children needed something else when they were angry is in line with emotion regulation research indicates that effective emotion regulation not only is merely about the frequency by which we apply certain strategies [69,70] but also is about whether we have access to a diverse set of emotion regulation strategies that can be flexibly applied depending on the situation's demands [71,72]. This highlights the complexity of the emotion regulation construct, which is sometimes applied in an overly simplistic manner [73,74]. To account for this complexity, we chose to develop psychoeducational films to introduce children to the more complex concept of emotion regulation. Interviews showed that films could have the potential to increase children's understanding and knowledge of emotion regulation, but more research is needed to prove this.

Another important feature to enhance emotion regulation was the *check-in* function, which was designed to support children by accessing adaptive emotion regulation strategies and increasing their awareness of the diverse range of emotions. Research has shown that both are related to positive mental health outcomes [75-77]. Children in this study found this feature helpful and used it either as a guide to identify their feelings or to explore and learn about different strategies. We believe that this feature could also serve as an ideal assessment tool for collecting data about users' day-to-day feelings.

Concerning the *help* function, we found that many children had not used or seen it until a researcher mentioned it in the interview. After consulting the children, most agreed that it was a good idea to provide guided support for emotionally intense situations. However, they also explained that having it in the app did not address their needs in the situation, due to a lack of permanent or immediate access to a device. We suggest that future research could explore this idea further through wearable devices, which also provide an ideal opportunity to integrate advanced health technology concepts, such as just-in-time adaptive interventions [78].

Strengths and Limitations

A significant strength of this research concerns the development process of the app, for which we included children and young people at every stage and adopted a truly interdisciplinary approach [52]. Another strength relates to the collaborative approach with schools, which has various benefits. It ensured regular access to the user group and helped identify context-specific design goals at an early stage. Teachers also contributed tremendously to their views and expertise from working with the user group. The lack of teachers' roles as intervention deliverers is a limitation that needs to be addressed further. Furthermore, teachers had very limited time available; therefore, we suggest conducting more regular classroom observations to adjust the app for the classroom setting.

Finally, we would like to mention that in some cases, teachers played a significant role in selecting children for the interviews. Although we asked teachers to provide us with a representative sample, it is possible that teachers unconsciously selected children who were more likely to report positive aspects. This assumption is based on previous research showing that children with certain characteristics, such as lower academic achievement

or greater externalizing tendencies, were less likely to be considered for taking part in research [79].

Conclusions

We explored the possibilities of using, implementing, and evaluating a new app intervention to improve children's emotion regulation abilities in the school context. The results suggest that the intervention presents a promising opportunity to enhance emotion regulation abilities by considering the complex nature of the construct itself. The app aims to assist children with their emotion regulation abilities by offering guidance in identifying feelings and selecting adaptive emotion regulation strategies. The app was perceived as acceptable and usable, although some technological issues need to be addressed before any further evaluation. The data provided valuable insights regarding important facilitators and barriers to implementing and evaluating the app in the school setting. Important *lessons learned* are shared and will hopefully be beneficial in the development and evaluation of similar interventions. We hope that this work will motivate the development of further technology-based interventions that target transdiagnostic mechanisms such as emotion regulation in youth, as these are of particular importance considering the high comorbidity rates and less specific symptom profiles in children and young people.

Acknowledgments

BM received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklowdowska-Curie grant (grant 722561). The authors would like to thank all the schools, children, teachers, and parents who participated in this project.

Authors' Contributions

BM is a research fellow in the Evidence-Based Practice Unit at the University College London, United Kingdom, and conducted this research as part of her PhD. She received funding for this study from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant 722561. The European Union's Horizon 2020 Research and Innovation Programme had no role in the study design, collection, analysis, or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication. BM was involved in conceptualizing, data collection, data analysis, writing, editing, and reviewing the manuscript. JD is a professor of the Evidence-Based Practice Unit at the University College London, United Kingdom. JD was involved in conceptualizing, writing, editing, and reviewing. PP is an associate professor at the MRC Unit for Lifelong Health and Ageing and the Center for Longitudinal Studies at University College London. PP was involved in conceptualizing, writing, editing, and reviewing. JEC is an associate professor at the Evidence-Based Practice Unit at the University College London, United Kingdom. JEC was involved in conceptualizing, writing, editing, and reviewing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Animation video showing an overview of app features and usage.
[MP4 File (MP4 Video), 29334 KB - mhealth_v9i7e21837_app1.mp4]

Multimedia Appendix 2
Interview script.
[DOCX File , 24 KB - mhealth_v9i7e21837_app2.docx]

References

1. Patalay P, Gage SH. Changes in millennial adolescent mental health and health-related behaviours over 10 years: a population cohort comparison study. Int J Epidemiol 2019 Oct 01;48(5):1650-1664 [FREE Full text] [doi: 10.1093/ije/dyz006] [Medline: 30815691]
2. Erskine HE, Moffitt TE, Copeland WE, Costello EJ, Ferrari AJ, Patton G, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. Psychol Med 2014 Dec 23;45(7):1551-1563. [doi: 10.1017/s0033291714002888]
3. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct;378(9801):1515-1525. [doi: 10.1016/s0140-6736(11)60827-1]
4. Izard C, Fine S, Schultz D, Mostow A, Ackerman B, Youngstrom E. Emotion knowledge as a predictor of social behavior and academic competence in children at risk. Psychol Sci 2001 Jan;12(1):18-23 [FREE Full text] [doi: 10.1111/1467-9280.00304] [Medline: 11294223]
5. Catalino LI, Fredrickson BL. A Tuesday in the life of a flourisher: the role of positive emotional reactivity in optimal mental health. Emotion 2011 Aug;11(4):938-950 [FREE Full text] [doi: 10.1037/a0024889] [Medline: 21859208]

6. Caan W, Cassidy J, Coverdale G, Ha M, Nicholson W, Rao M. The value of using schools as community assets for health. Public Health 2015 Jan;129(1):3-16. [doi: 10.1016/j.puhe.2014.10.006] [Medline: 25481543]
7. Stephan SH, Weist M, Kataoka S, Adelsheim S, Mills C. Transformation of children's mental health services: the role of school mental health. Psychiatr Serv 2007 Oct;58(10):1330-1338. [doi: 10.1176/ps.2007.58.10.1330] [Medline: 17914011]
8. Weist MD, Evans SW. Expanded school mental health: challenges and opportunities in an emerging field. J Youth Adolescence 2005 Feb;34(1):3-6. [doi: 10.1007/s10964-005-1330-2]
9. Memon A, Taylor K, Mohebati LM, Sundin J, Cooper M, Scanlon T, et al. Perceived barriers to accessing mental health services among black and minority ethnic (BME) communities: a qualitative study in Southeast England. BMJ Open 2016 Nov 16;6(11):e012337 [FREE Full text] [doi: 10.1136/bmjopen-2016-012337] [Medline: 27852712]
10. Weare K, Nind M. Mental health promotion and problem prevention in schools: what does the evidence say? Health Promot Int 2011 Dec 12;26 Suppl 1(suppl 1):29-69. [doi: 10.1093/heapro/dar075] [Medline: 22079935]
11. Bishop D, Bryant KS, Giles SM, Hansen WB, Dusenbury L. Simplifying the delivery of a prevention program with web-based enhancements. J Prim Prev 2006 Jul 9;27(4):433-444. [doi: 10.1007/s10935-006-0042-z] [Medline: 16763766]
12. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr 02;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
13. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
14. Fridrici M, Lohaus A. Stress-prevention in secondary schools: online- versus face-to-face-training. Health Edu 2009 Jun 19;109(4):299-313. [doi: 10.1108/09654280910970884]
15. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
16. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
17. Edbrooke-Childs J, Smith J, Rees J, Edridge C, Calderon A, Saunders F, et al. An app to help young people self-manage when feeling overwhelmed (ReZone): protocol of a cluster randomized controlled trial. JMIR Res Protoc 2017 Nov 03;6(11):e213. [doi: 10.2196/resprot.7019]
18. Edridge C, Deighton J, Wolpert M, Edbrooke-Childs J. The implementation of an mHealth intervention (ReZone) for the self-management of overwhelming feelings among young people. JMIR Form Res 2019 May 02;3(2):e11958 [FREE Full text] [doi: 10.2196/11958] [Medline: 31045499]
19. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
20. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva: World Health Organization; 2016:1-133.
21. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris France p. 2627-2636. [doi: 10.1145/2470654.2481364]
22. Coyle D, McGlade N, Doherty G, O'Reilly G. Exploratory evaluations of a computer game supporting cognitive behavioural therapy for adolescents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver BC Canada p. 2937-2946. [doi: 10.1145/1978942.1979378]
23. Matthews M, Doherty G. In the mood: engaging teenagers in psychotherapy using mobile phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver BC Canada p. 2947-2956. [doi: 10.1145/1978942.1979379]
24. Taylor RD, Oberle E, Durlak JA, Weissberg RP. Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. Child Dev 2017 Jul 07;88(4):1156-1171. [doi: 10.1111/cdev.12864] [Medline: 28685826]
25. Greenberg MT, Weissberg RP, O'Brien MU, Zins JE, Fredericks L, Resnik H, et al. Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. Am Psychol 2003 Jun;58(6-7):466-474. [doi: 10.1037/0003-066X.58.6-7.466]
26. Thompson RA, Goodman M. Development of emotion regulation: more than meets the eye. In: Kring AM, Sloan DM, Gross JJ, Werner KH, Thompson RA, editors. Emotion Regulation and Psychopathology: A Transdiagnostic Approach to Etiology and Treatment. New York, United States: Guilford Press; 2009:1-461.
27. Gross JJ, Muñoz RF. Emotion regulation and mental health. Clinical Psychology: Science and Practice 1995;2(2):151-164. [doi: 10.1111/j.1468-2850.1995.tb00036.x]

28. Keenan K. Emotion dysregulation as a risk factor for child psychopathology. Clinical Psychology: Science and Practice 2000;7(4):418-434. [doi: 10.1093/clipsy.7.4.418]
29. Sloan E, Hall K, Moulding R, Bryce S, Mildred H, Staiger PK. Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review. Clin Psychol Rev 2017 Nov;57:141-163. [doi: 10.1016/j.cpr.2017.09.002] [Medline: 28941927]
30. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. Clin Psychol Rev 2010 Mar;30(2):217-237. [doi: 10.1016/j.cpr.2009.11.004] [Medline: 20015584]
31. Moltrecht B, Deighton J, Patalay P, Edbrooke-Childs J. Effectiveness of current psychological interventions to improve emotion regulation in youth: a meta-analysis. Eur Child Adolesc Psychiatry 2021 Jun 27;30(6):829-848 [FREE Full text] [doi: 10.1007/s00787-020-01498-4] [Medline: 32108914]
32. Forbes MK, Rapee RM, Krueger RF. Opportunities for the prevention of mental disorders by reducing general psychopathology in early childhood. Behav Res Ther 2019 Aug;119:103411 [FREE Full text] [doi: 10.1016/j.brat.2019.103411] [Medline: 31202004]
33. Sonne T, Jensen MM. ChillFish: A respiration game for children with ADHD. In: Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. 2016 Presented at: TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction; February 14-17, 2016; Eindhoven Netherlands p. 271-278. [doi: 10.1145/2839462.2839480]
34. Lobel A, Gotsis M, Reynolds E, Annetta M, Engels R, Granic I. Designing and utilizing biofeedback games for emotion regulation: the case of nevermind. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose California USA p. 1945-1951. [doi: 10.1145/2851581.2892521]
35. Lee CS, Lam SH, Tsang ST, Yuen CM, Ng CK. The effectiveness of technology-based intervention in improving emotion recognition through facial expression in people with autism spectrum disorder: a systematic review. Rev J Autism Dev Disord 2017 Dec 27;5(2):91-104. [doi: 10.1007/s40489-017-0125-1]
36. Grossard C, Grynspan O, Serret S, Jouen A, Bailly K, Cohen D. Serious games to teach social interactions and emotions to individuals with autism spectrum disorders (ASD). Comput Edu 2017 Oct;113:195-211. [doi: 10.1016/j.compedu.2017.05.002]
37. Park JH, Abirached B, Zhang Y. A framework for designing assistive technologies for teaching children with ASDs emotions. In: Extended Abstracts CHI '12 Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin Texas USA p. 2423-2428. [doi: 10.1145/2212776.2223813]
38. Hides L, Dingle G, Quinn C, Stoyanov SR, Zelenko O, Tjondronegoro D, et al. Efficacy and outcomes of a music-based emotion regulation mobile app in distressed young people: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11482 [FREE Full text] [doi: 10.2196/11482] [Medline: 30664457]
39. Rapee RM, Oar EL, Johnco CJ, Forbes MK, Fardouly J, Magson NR, et al. Adolescent development and risk for the onset of social-emotional disorders: a review and conceptual model. Behav Res Ther 2019 Dec;123:103501. [doi: 10.1016/j.brat.2019.103501] [Medline: 31733812]
40. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustun TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007;20(4):359-364. [doi: 10.1097/yco.0b013e32816ebc8c]
41. Powers A, Casey BJ. The adolescent brain and the emergence and peak of psychopathology. J Infant Child Adolesc Psychother 2015 Mar 17;14(1):3-15. [doi: 10.1080/15289168.2015.1004889]
42. Dahl RE. Adolescent brain development: a period of vulnerabilities and opportunities. Keynote address. Ann N Y Acad Sci 2004 Jun;1021:1-22. [doi: 10.1196/annals.1308.001] [Medline: 15251869]
43. Zeedyk MS, Gallacher J, Henderson M, Hope G, Husband B, Lindsay K. Negotiating the transition from primary to secondary school. Sch Psychol Int 2016 Jun 29;24(1):67-79. [doi: 10.1177/0143034303024001010]
44. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
45. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
46. Craig P, Dieppe P, Macintyre S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: an introduction to the new Medical Research Council guidance. In: Killoran A, Kelly MP, editors. Evidence-Based Public Health: Effectiveness and Efficiency. Oxfordshire, United Kingdom: Oxford University Press; 2009:1-592.
47. Marcu G, Bardram J, Gabrielli S. A framework for overcoming challenges in designing persuasive monitoring and feedback systems for mental illness. In: Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare. 2011 Presented at: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare; May 23–26, 2011; Dublin, Republic of Ireland p. 1-8. [doi: 10.4108/icst.pervasivehealth.2011.246097]

48. Druin A. Cooperative inquiry: developing new technologies for children with children. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1999 Presented at: CHI99: Conference on Human Factors in Computing Systems; May 15-20, 1999; Pittsburgh Pennsylvania USA p. 592-599. [doi: 10.1145/302979.303166]
49. Corry M, Clarke M, While AE, Lalor J. Developing complex interventions for nursing: a critical review of key guidelines. J Clin Nurs 2013 Sep 29;22(17-18):2366-2386. [doi: 10.1111/jocn.12173] [Medline: 23551526]
50. Druin A. The role of children in the design of new technology. Behav Info Tech 2002 Jan;21(1):1-25. [doi: 10.1080/01449290110108659]
51. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015 Jul 09;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
52. Moltrecht B, Patalay P, Bear HA, Deighton J, Edbrooke-Childs J. Interdisciplinary development of a transdiagnostic mobile app to enhance children's emotion regulation: sharing insights and lessons learned. PsyArXiv. Preprint posted online on March 2, 2021. [FREE Full text] [doi: 10.31234/osf.io/ks7ag]
53. Children and parents: media use and attitudes report 2019. Ofcom. 2020. URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2019 [accessed 2021-06-12]
54. Angold A, Erkanli A, Silberg J, Eaves L, Costello EJ. Depression scale scores in 8-17-year-olds: effects of age and gender. J Child Psychol Psychiatry 2002 Nov;43(8):1052-1063. [doi: 10.1111/1469-7610.00232] [Medline: 12455926]
55. Sharp C, Goodyer IM, Croudace TJ. The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. J Abnorm Child Psychol 2006 Jun 29;34(3):379-391. [doi: 10.1007/s10802-006-9027-x] [Medline: 16649000]
56. Deighton J, Tymms P, Vostanis P, Belsky J, Fonagy P, Brown A, et al. The development of a school-based measure of child mental health. J Psychoeduc Assess 2013 Jun 20;31(3):247-257 [FREE Full text] [doi: 10.1177/0734282912465570] [Medline: 25076806]
57. Patalay P, Deighton J, Fonagy P, Vostanis P, Wolpert M. Clinical validity of the Me and My School questionnaire: a self-report mental health measure for children and adolescents. Child Adolesc Psychiatry Ment Health 2014;8(1):17 [FREE Full text] [doi: 10.1186/1753-2000-8-17] [Medline: 24963344]
58. Gadermann AM, Schonert-Reichl KA, Zumbo BD. Investigating validity evidence of the satisfaction with life scale adapted for children. Soc Indic Res 2009 Apr 28;96(2):229-247. [doi: 10.1007/s11205-009-9474-1]
59. Walden TA, Harris VS, Catron TF. How I feel: a self-report measure of emotional arousal and regulation for children. Psychol Assess 2003 Sep;15(3):399-412. [doi: 10.1037/1040-3590.15.3.399] [Medline: 14593841]
60. Ciucci E, Baroncelli A, Grazzani I, Ornaghi V, Caprin C. Emotional arousal and regulation: further evidence of the validity of the "How I feel" questionnaire for use with school-age children. J Sch Health 2016 Mar 02;86(3):195-203. [doi: 10.1111/josh.12370] [Medline: 26830506]
61. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
62. Lang J, Schmitz B. German translation of the satisfaction with life scale for children and adolescents. J Psychoeduc Assess 2019 Jun 11;38(3):291-304. [doi: 10.1177/0734282919849361]
63. Buss KA, Goldsmith HH. Fear and anger regulation in infancy: effects on the temporal dynamics of affective expression. Soc Res Child Dev 1998:359-374. [doi: 10.1111/j.1467-8624.1998.tb06195.x]
64. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
65. Edbrooke-Childs J, Edridge C, Averill P, Delane L, Hollis C, Craven MP, et al. A feasibility trial of power up: smartphone app to support patient activation and shared decision making for mental health in young people. JMIR Mhealth Uhealth 2019 Jun 04;7(6):e11677 [FREE Full text] [doi: 10.2196/11677] [Medline: 31165709]
66. Heppner WL, Spears CA, Vidrine JI, Wetter DW. Mindfulness and emotion regulation. In: Ostafin B, Robinson M, Meier B, editors. Handbook of Mindfulness and Self-Regulation. New York: Springer; 2015:107-120.
67. Pascual-Leone A, Gillespie NM, Orr ES, Harrington SJ. Measuring subtypes of emotion regulation: from broad behavioural skills to idiosyncratic meaning-making. Clin Psychol Psychother 2016 May 20;23(3):203-216. [doi: 10.1002/cpp.1947] [Medline: 25704129]
68. Livingstone KM, Srivastava S. Up-regulating positive emotions in everyday life: strategies, individual differences, and associations with positive emotion and well-being. J Res Pers 2012 Oct;46(5):504-516. [doi: 10.1016/j.jrp.2012.05.009]
69. Dixon-Gordon KL, Aldao A, De Los Reyes A. Repertoires of emotion regulation: a person-centered approach to assessing emotion regulation strategies and links to psychopathology. Cogn Emot 2015 Dec 07;29(7):1314-1325. [doi: 10.1080/02699931.2014.983046] [Medline: 25435338]
70. Aldao A, Sheppes G, Gross JJ. Emotion regulation flexibility. Cogn Ther Res 2015 Jan 1;39(3):263-278. [doi: 10.1007/s10608-014-9662-4]
71. Lougheed JP, Hollenstein T. A limited repertoire of emotion regulation strategies is associated with internalizing problems in adolescence. Soc Dev 2012:704-721. [doi: 10.1111/j.1467-9507.2012.00663.x]

72. Aldao A, Nolen-Hoeksema S. The influence of context on the implementation of adaptive emotion regulation strategies. Behav Res Ther 2012 Aug;50(7-8):493-501. [doi: 10.1016/j.brat.2012.04.004] [Medline: 22659159]
73. Yoon J, Li S, Hao Y, Kim C. Towards emotional well-being by design: 17 opportunities for emotion regulation for user-centered healthcare design. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento Italy p. 351-355. [doi: 10.1145/3329189.3329227]
74. Aldao A. The future of emotion regulation research: capturing context. Perspect Psychol Sci 2013 Mar 27;8(2):155-172. [doi: 10.1177/1745691612459518] [Medline: 26172497]
75. Gonçalves SF, Chaplin TM, Turpyn CC, Niehaus CE, Curby TW, Sinha R, et al. Difficulties in emotion regulation predict depressive symptom trajectory from early to middle adolescence. Child Psychiatry Hum Dev 2019 Aug 28;50(4):618-630 [FREE Full text] [doi: 10.1007/s10578-019-00867-8] [Medline: 30689145]
76. Quoidbach J, Gruber J, Mikolajczak M, Kogan A, Kotsou I, Norton MI. Emodiversity and the emotional ecosystem. J Exp Psychol Gen 2014 Dec;143(6):2057-2066. [doi: 10.1037/a0038025] [Medline: 25285428]
77. Ciarrochi J, Heaven PC, Supavadeeprasit S. The link between emotion identification skills and socio-emotional functioning in early adolescence: a 1-year longitudinal study. J Adolesc 2008 Oct;31(5):565-582. [doi: 10.1016/j.adolescence.2007.10.004] [Medline: 18083221]
78. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
79. Noll RB, Zeller MH, Vannatta K, Bukowski WM, Davies WH. Potential bias in classroom research: comparison of children with permission and those who do not receive permission to participate. J Clin Child Psychol 1997 Mar;26(1):36-42. [doi: 10.1207/s15374424jccp2601_4] [Medline: 9118174]

Abbreviations

AWS EC2: Amazon Elastic Compute Cloud
MRC: Medical Research Council

Edited by L Buis; submitted 16.09.20; peer-reviewed by S Byrne, S Hugh-Jones; comments to author 09.11.20; revised version received 03.03.21; accepted 28.04.21; published 14.07.21.

Please cite as:
Moltrecht B, Patalay P, Deighton J, Edbrooke-Childs J
A School-Based Mobile App Intervention for Enhancing Emotion Regulation in Children: Exploratory Trial
JMIR Mhealth Uhealth 2021;9(7):e21837
URL: https://mhealth.jmir.org/2021/7/e21837
doi:10.2196/21837
PMID:34259642

Original Paper

Measuring the Healthiness of Ready-to-Eat Child-Targeted Cereals: Evaluation of the FoodSwitch Platform in Sweden

Antoine Mottas[1], MSc; Veli-Matti Lappi[1], MSc, MD; Johan Sundström[2,3], MD, PhD; Bruce Neal[3,4], MBChB, PhD; Cliona Ni Mhurchu[3,5], BSc (Hons), PhD; Marie Löf[1,6], PhD; Karin Rådholm[3,6], MD, PhD

[1]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[2]Department of Medical Sciences, Uppsala University, Uppsala, Sweden

[3]The George Institute for Global Health, University of New South Wales, Sydney, Australia

[4]Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

[5]National Institute for Health Innovation, School of Population Health, University of Auckland, Auckland, New Zealand

[6]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Karin Rådholm, MD, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Linköping, SE-58183
Sweden
Phone: 46 13 281756
Email: karin.radholm@liu.se

Abstract

Background: Childhood obesity is a major public health issue. The increase in the consumption of foods with poor nutritional value, such as processed foods, contributes to this. Breakfast cereals are often advertised as a healthy way to start the day, but the healthiness of these products varies greatly.

Objective: Our main objective was to gather information about the nutritional characteristics of ready-to-eat breakfast cereals in Sweden and to investigate the healthiness of products targeted at children compared to other cereals by use of the FoodSwitch platform. A secondary objective was to evaluate the alignment between the Keyhole symbol and the Health Star Rating.

Methods: The FoodSwitch app is a mobile health (mHealth) tool used to present nutrition data and healthier alternative products to consumers. Ready-to-eat breakfast cereals from the largest Swedish grocery retailers were collected using the FoodSwitch platform. Products were defined as targeting children if they presented features addressing children on the package.

Results: Overall, information on 261 ready-to-eat cereals was examined. Of this total, 8% (n=21) were targeted at children. Child-targeted cereals were higher in sugar (22.3 g/100 g vs 12.8 g/100 g, P<.001) and lower in fiber (6.2 g/100 g vs 9.8 g/100 g, P<.001) and protein (8.1 g/100 g vs 10.5 g/100 g, P<.001). Total fat (3 g/100 g vs 10.5 g/100 g, P<.001) and saturated fat (0.8 g/100 g vs 2.6 g/100 g, P<.001) were also lower. No difference was found in salt content (P=.61). Fewer child-targeted breakfast cereals displayed an on-pack Keyhole label (n=1, 5% vs n=53, 22%; P=.06), and the mean Health Star Rating value was 3.5 for child-targeted cereals compared to others (mean 3.8, P=.07). A correlation was found between the Keyhole symbol and the Health Star Rating.

Conclusions: Ready-to-eat breakfast cereals targeted at children were less healthy in terms of sugar and fiber content compared to products not targeted at children. There is a need to improve the nutritional quality of child-targeted cereals.

(JMIR Mhealth Uhealth 2021;9(7):e17780) doi:10.2196/17780

KEYWORDS

breakfast cereals; child-targeted cereals; front-of-pack labels; Keyhole symbol; Health Star Rating; FoodSwitch; diet; food intake

Introduction

Childhood obesity is a major public health issue. Over 40 million children under 5 years of age are overweight worldwide with a majority living in low- or middle-income countries [1]. The increasing rates of obesity in children are linked to noncommunicable diseases such as type 2 diabetes and cardiovascular diseases in young adults [2]. In addition,

individuals with obesity at an early age are more prone to obesity later in life [3]. Unhealthy diet is the leading risk factor for obesity in children and adults. Processed foods, which are energy dense and of poor nutritional value, are a major cause of obesity [4]. In the last few decades, the consumption of processed and ultraprocessed food has increased and correlates with the elevated prevalence of obesity [5].

Ready-to-eat breakfast cereals include a large range of products from unprocessed to ultraprocessed. This large diversity in products and the omnipresence of nutritional claims on the packages of breakfast cereals make it hard for the consumer to identify actual healthy options. In addition, there are data suggesting that highly advertised products targeted at children may be less healthy [6]. To help consumers, front-of-pack nutrition labels have been created. There are many types of labeling, often specific to a country or area, but all aim to promote healthier food choices and, in the long term, to urge manufacturers to reformulate their products [7]. The Swedish National Food Administration introduced the Keyhole symbol that endorses products with healthier fat composition, less sugars and salt, and more dietary fiber, whole grains, fruits, and vegetables than other similar foods [8]. The Health Star Rating (HSR), developed by the Australian government, grades foods from 0.5 to 5 stars [9]. Grades are attributed based on energy content, saturated fat, sugar, salt, protein, fiber, as well as fruit, vegetable, nut, and legume content, in the product. In other countries different front-of-pack labels are used. For instance, nutrient-specific labels that assess the percentage of each nutrients (eg, Traffic Lights labeling and nutrient-specific warnings that indicate an excessive amount of critical nutrients). The weakness of most systems is that they remain voluntary and rely on industry interest. In addition, the coexistence of numerous front-of-pack labels in the marketplace, as well as the great disparity between their criteria, can be confusing for consumers [10].

The use of mobile apps could be an innovative alternative way for consumers to access more information on the nutrients and healthiness of the foods they buy. Today, 88% of people in Sweden over the age of 12 years own a smartphone, which offers great potential for mobile health (mHealth) apps in all age groups [11]. The FoodSwitch app, developed by the George Institute for Global Health in Australia, was designed to offer a tool to promote healthier food choices and is currently being used in Australia, New Zealand, China, the United Kingdom, India, the United States, Kuwait, South Africa, Fiji, and Hong Kong. By scanning bar codes of packaged foods, the consumer obtains at-a-glance information on the product as well as suggestions of similar healthier products [12]. The FoodSwitch solution includes the FoodSwitch app, a database with packaged products organized in a categorized system that is applicable to all countries and enables comparisons, and a data collection application (DataCollector). The FoodSwitch database has been previously used in several studies to measure the healthiness of packaged food, to evaluate variations in specific nutrients, and to compare front-of-pack labeling systems [13-15].

In this study, we aim to use the FoodSwitch platform in the Swedish market to compare the healthiness and nutritional values of ready-to-eat breakfast cereals targeted at children to non–child-targeted, ready-to-eat cereals. Furthermore, we aim to assess the alignment of the Nordic Keyhole symbol and the HSR system for these products.

Methods

Data Collection

Ready-to-eat cereals from 4 supermarkets (ICA, Coop, Hemköp, and Lidl), representing 90% of the market share in Sweden, were analyzed [16]. The largest supermarkets in the Stockholm area were targeted to ensure the presence of the entire range of products. Prior to data collection, which took place in February and March 2019, consent forms were sent to the store managers. One store denied us permission without giving more details on why they chose to do so, and thus this retailer was not included. On site, the cereal department and all shelves containing breakfast cereals, including gluten-free sections, were identified. Oatmeal for porridge was regarded as a grain and not used as a breakfast cereal per se. Muesli, however, was included. There is a thin line between cereal bars and candy, and cereal bars are categorized differently from breakfast cereal on the FoodSwitch platform. Thus, they were not deemed breakfast cereal and were excluded.

The DataCollector app of the FoodSwitch platform (Android, version 2.7) from the George Institute for Global Health was used to collect information from packages. First, the bar codes were scanned using a smartphone camera. Then, numerous pictures of the front of the pack, the Nutrition Information Panel (NIP), and the ingredients were taken to collect all relevant information.

All products were manually entered in the system by 1 researcher (AM). The NIP, the gluten status and the Keyhole symbol, if present, were recorded. All child-targeted information on the package was identified. The criteria were as follows: the presence of cartoons, games, toys, children's movie references, or text addressed to children on the package. A second researcher (V-ML) reviewed all the information entered and confirmed the presence of marketing targeted at children.

Finally, products were categorized and an HSR was generated. HSR scores of 0.5 to 5 stars in 10 half-star increments were assigned to all scanned products, where a higher number of stars represents healthier products. In the HSR system, each packaged food item is categorized into 1 of 6 categories depending on food type. All breakfast cereals were assigned to category 2. The HSR score was calculated via baseline and modifying points using the following formula: HSR score = baseline points – modifying points. Baseline points depended on energy content, saturated fat, sugar, and salt, and modifying points were based on protein, fiber, fruit, and vegetable content. The final assignment of the HSR score depended on which category the product was assigned to [17].

Statistical Analysis

Categorical variables were summarized as the number of products and corresponding percentages, and continuous variables were summarized as mean (SD) and median (IQR). A Kolmogorov-Smirnov test was used to determine whether the data were normally distributed or not. A Mann-Whitney *U*

test and an independent-samples *t* test were used to compare the nutritional values and HSR of child-targeted cereals to the nontargeted ones for nonnormal and normal distribution, respectively. In addition, the HSR scores were divided into two groups, healthier (HSR≥3.5) and less healthy (HSR<3.5) [18]. After dichotomization of HSR scores, the Fisher exact test was performed to assess correlation to the presence of the Keyhole symbol. The significance level was set to $P<.05$. The statistical analysis was performed using SPSS Statistics 25 (IBM Corp).

Ethical Considerations

This was a study of packaged food supplies in supermarkets in Sweden and did not involve study participants or animal testing. Therefore, no ethical permission was sought.

Results

By use of the data collection application of the FoodSwitch platform, we collected information on a total of 261 ready-to-eat breakfast cereals, of which 21 (8%) were targeted at children and 240 (92%) were not. Table 1 summarizes the nutritional content from the NIP of all products as well as their HSR score and the number of products displaying the Keyhole symbol. Child-targeted cereals contained more sugars, with a mean carbohydrate content of 78.3 g compared to 62.0 g per 100 g for the non–child-targeted group ($P<.001$). This applied to sugar as well, where the mean in child-targeted products was nearly twice as high than in the non–child-targeted ones (22.3 g vs 12.8 g per 100 g, $P<.001$). Total fat, saturated fat, fiber, and protein were all lower in the child-targeted cereals compared to the cereals not targeted at children ($P<.001$ for all). There was no difference in salt content ($P=.61$) between the groups. The main categories of cereals not targeted at children were muesli and granola, while cocoa-based and sweetened cereals accounted for the majority of child-targeted cereals (Table S1 in Multimedia Appendix 1).

Figure 1 shows that on-pack Keyhole labeling was aligned with products considered as "healthy" (≥3.5) according to the HSR ($P<.001$). No products with an HSR score <4 were labeled with the Keyhole symbol. However, 60% (n=92) of products with 4 stars or more were not Keyhole labeled. According to the HSR, there was no significant difference between cereals targeted at children and those that were not (mean 3.5 for the child-targeted group vs 3.8 for the non–child-targeted group, $P=.07$).

Table 1. Nutritional overview of ready-to-eat breakfast cereals targeted at children (n=21) compared to those not targeted at children (n=240) and their estimated healthiness. The mean fiber amount in child-targeted cereals was based on 20 products since 1 product did not display its fiber content.

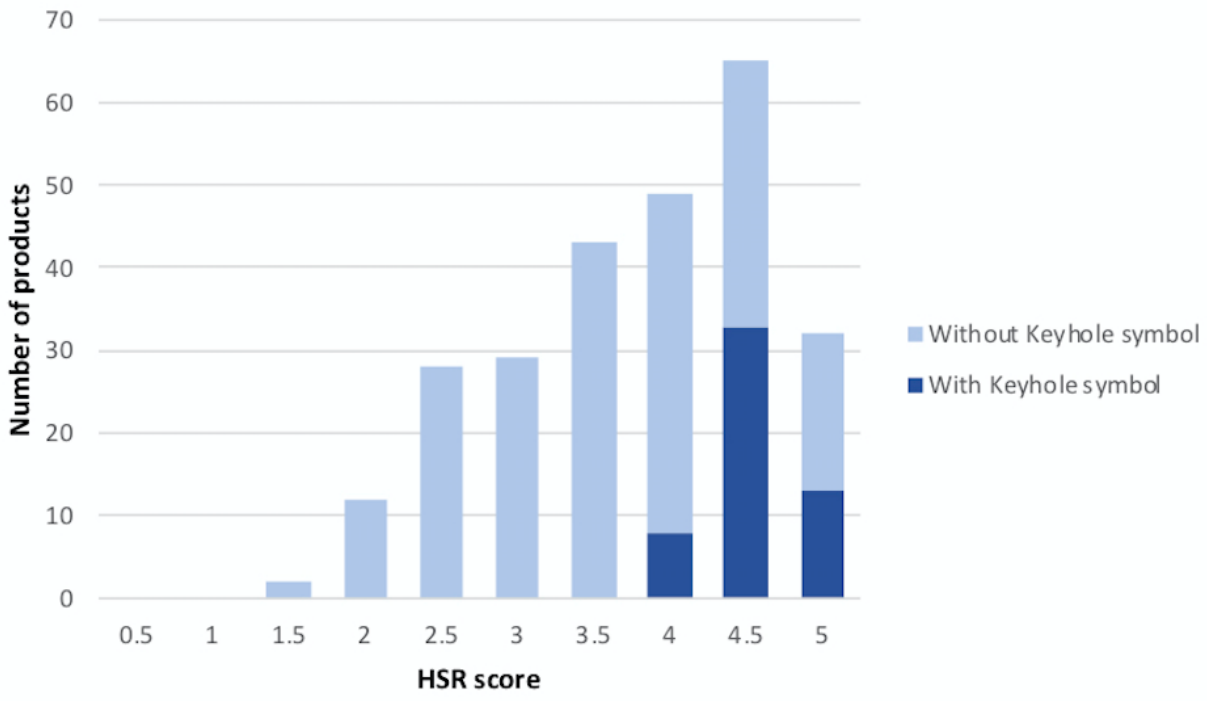
Nutrition[a]	Child-targeted cereals	Non–child-targeted cereals	*P* value[b]
Energy (kJ/100 g)			.18
Mean (SD)	1628 (68)	1699 (202)	
Median (IQR)	1618 (49)	1647 (225)	
Total fat (g/100 g)			<.001
Mean (SD)	3.0 (2.8)	10.5 (9.6)	
Median (IQR)	2.3 (1.5)	7.6 (12.8)	
Saturated fat (g/100 g)			<.001
Mean (SD)	0.8 (0.8)	2.6 (3.0)	
Median (IQR)	0.6 (0.6)	1.8 (3.0)	
Carbohydrates (g/100 g)			<.001
Mean (SD)	78.3 (5.3)	62.0 (13.5)	
Median (IQR)	78.0 (6.1)	62.0 (13.4)	
Sugars (g/100 g)			<.001
Mean (SD)	22.3 (7.5)	12.8 (7.8)	
Median (IQR)	23.5 (13.0)	11.0 (12.1)	
Protein (g/100 g)			.002
Mean (SD)	8.1 (1.7)	10.5 (2.6)	
Median (IQR)	8.2 (2.2	10.0 (3.2))	
Fiber (g/100 g)			.01
Mean (SD)	6.2 (2.5)	9.8 (5.0)	
Median (IQR)	6.8 (3.6)	9.0 (4.8)	
Salt (g/100 g)			.61
Mean (SD)	0.5 (0.3)	0.5 (0.4)	
Median (IQR)	0.6 (0.6)	0.4 (0.6)	
HSR[c] score			.07
Mean (SD)	3.5 (0.5)	3.8 (0.9)	
Median (IQR)	3.8 (1.0)	4.0 (1.5)	
Keyhole symbol, n (%)	1 (5)	53 (22)	.06

[a]Data are n (%) for the categorical variable and mean (SD) and median (IQR) for continuous variables.

[b]Comparison of child-targeted versus non–child-targeted breakfast cereals was analyzed with a Mann-Whitney *U* test.

[c]HSR: Health Score Rating.

Figure 1. The calculated Health Star Rating (HSR) score of breakfast cereals in supermarkets in Sweden (light blue) and the HSR scores of packaged products displaying an on-pack Keyhole symbol (dark blue).



Discussion

Principal Findings

This study compared the healthiness of ready-to-eat breakfast cereals targeted at children to cereal products not specifically targeted at children across supermarkets in Sweden using the FoodSwitch platform. Despite lower levels of saturated fat, the child-targeted cereals were overall less healthy, according to the nutritional content. Child-targeted products had a greater amount of sugar, with a mean nearly 2 times higher than the cereals not branded toward children. Furthermore, they were lower in fiber and protein. Our results are in line with 3 similar studies conducted in Canada, Australia, New Zealand, Guatemala, and the United States [19-23]. However, this is the first study focusing on the healthiness of packaged cereal products in Europe. Interestingly, the number of products targeted at children is relatively small. Compared to the markets analyzed in other countries, we obtained the smallest ratio between child-targeted cereals and non–child-targeted ones [19-23]. A few cereals rated high in healthiness according to the HSR system although these had a high sugar content. This aligns with previous findings showing that the HSR rating algorithm does not sufficiently penalize products with a high added sugar content [24].

Societal Relevance

In a context where childhood obesity is at the heart of public health concerns, our findings are highly relevant. The direct link between added sugar intake and obesity is well established [25,26]. Recently, the American Heart Association reported associations between added sugars and cardiovascular disease risk factors among children at levels well below the actual average intake [27]. Therefore, sugar levels in foods intended for children should be as low as possible or at least be similar to products not branded toward children. Dietary fibers are known to improve satiety, and fiber consumption is associated with lower body weight [28]. Even though it has been shown that consumers of breakfast cereals are more likely to reach the recommended daily amount of fiber, there should be no difference in fiber between food items branded toward children and cereals not targeted at children [29]. The presence of more total fat and saturated fat in non–child-targeted products was probably due to cereal type. When the products were categorized, we noticed that granola and muesli accounted for a majority. These types of cereals often contain nuts and seeds, which are high in fat. Salt content was similarly low in all products. In fact, many sodium-reduction initiatives have been introduced as part of the European Union salt reduction framework [30]. One study showed a significant drop in salt levels in breakfast cereals in the United Kingdom [31]. We can assume that the same happened in Sweden since the brands sold in both countries are very similar; however, this would be possible to investigate in the future by using this study as the reference values for 2019.

Marketing toward children often depicts cartoon characters to influence their food preferences [32]. In addition, children exposed to child-targeted cereals via TV commercials tend to have higher intake of these advertised cereals [33]. Problematically, products advertised to children are often of poor nutritional quality, as demonstrated in our study [34]. Unfortunately, packaging also affects parents, and parents being role models influence the eating habits of their children [35]. In fact, children depend on their parents' food choices. Finally, it has been shown that NIPs and front-of-pack labels can

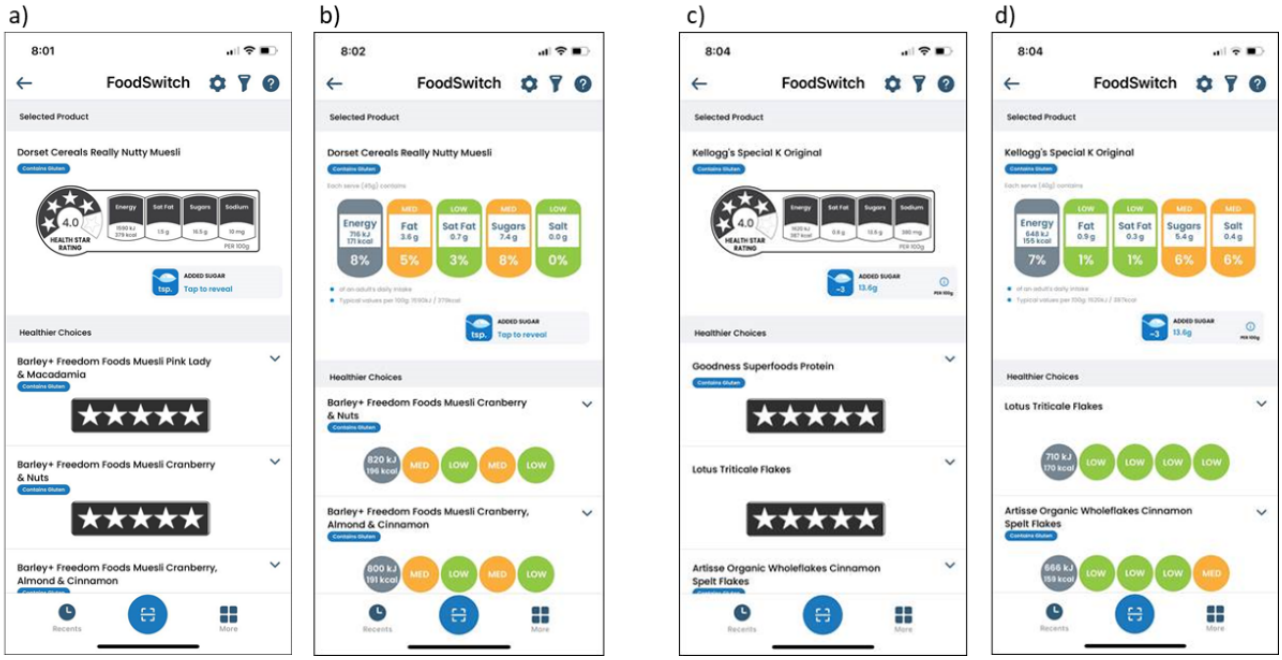
sometimes be difficult to interpret [36]. Therefore, there is a need to provide easy-to-understand solutions to better guide consumers and implement healthy dietary habits from the earliest age.

Potential for the FoodSwitch Platform in Sweden

A substantial part of the foods sold and eaten in high-income countries is preprepared packaged foods, and processed foods are largely responsible for an excessive intake of saturated fat, energy, added sugars, and sodium in the diet [37]. Thus, the food industry plays an increasingly important role in public health. The aim of the FoodSwitch platform is to change consumers' behaviors and legislators' actions. Front-of-pack labels are well established in influencing consumers toward healthier food choices. The Keyhole symbol used in Sweden was implemented 30 years ago. This front-of-pack logo is well known by Swedes as well as Scandinavians [38]. Although it is hard to assess its effect, it seems to positively affect the consumers' behavior [39]. Unfortunately, not all manufacturers use the logo even though it is free of charge. The criteria required to make a product eligible for the Keyhole symbol are numerous and vary according to product type. Thus, producers might not be aware when a product meets the standard. Moreover, displaying the symbol induces extra costs for new packaging. Manufacturers often prioritize the sales of their foods. Therefore, another reason for the lack of compliance to the Keyhole symbol might be that producers are not convinced of any added value by the logo on their products. Throughout our investigation, we realized that many products considered as "healthy" by the HSR system (>3.5) did not display the Keyhole symbol. It is very likely that some of them were suitable to be branded with the Keyhole symbol.

Because of this, introducing the FoodSwitch app in Sweden could be of interest to Swedish consumers. First, it would allow a virtual front-of-pack labeling of all products (Figure 2). In fact, with information from the database, the app could generate the HSR score and the Traffic Lights label for any packaged product and suggest a healthier alternative. However, these two front-of-pack labeling systems used in the FoodSwitch app to direct consumers toward a healthier food choice are not used in Sweden. Fortunately, both the HSR and Traffic Lights label are self-explanatory and should not impede compliance of Swedish users to the mobile app's recommendations. In addition, it would be possible to add a filter on the app that would generate the Keyhole symbol. Second, the implementation of FoodSwitch in Sweden could pressure manufacturers to reformulate their products. If consumers choose healthier packaged foods, producers would have to adapt to evolving client demand. Finally, the FoodSwitch database would be a very useful tool for conducting research on many aspects of the packaged food market. Through crowdsourcing, the information from packaged foods are always updated, which allows a tracking of the nutritional composition of the food supply over time.

Figure 2. Screenshots of the Australian version of the FoodSwitch app displaying the estimated healthiness of products using Health Star Ratings (panels A and C) and Traffic Lights ratings (panels B and D) for the scanned products Dorset Cereals Really Nutty Muesli (panels A and B) and Kellogg's Special K Original (panels C and D), and suggestions for healthier options.



Strengths, Limitations, and Future Directions

We consider the conclusions drawn from our primary investigation to be reliable for the Swedish market because we visited the largest grocery retailers, representing 90% of market shares [16]. Another strength is the reliability of our data. During data collection, the data were double-checked by the collector prior to data entry. Once entered, the data were checked by a second researcher.

Limitations include possible transcribing errors and, in some cases, minor discrepancies that can occur between the NIP information shown on the package and the real value [40]. In terms of the method of assessing the healthiness of products,

in Europe, displaying fiber content on the package is not mandatory [41]; however, it is often included on the NIP and only one product in our sample did not display fiber content. Finally, another limitation is that our results are limited to the Swedish market.

A further research objective could be to expand this assessment of nutrition to all child-targeted products. Additionally, data collection of breakfast cereals can be redone to provide a longitudinal perspective.

Conclusion

In conclusion, we showed that the nutritional quality of ready-to-eat cereals targeted at children was overall not significantly unhealthier than ready-to-eat breakfast cereals not targeted at children. However, cereals targeted at children were high in sugar and low in fiber. Thus, we conclude that there is a need to improve the dietary quality of child-targeted breakfast cereals in Sweden.

Authors' Contributions

AM, CNM, BN, and KR contributed to the concept and rationale of the study. The data were collected and analyzed by AM. V-ML conducted the data review. AM and KR drafted the manuscript. All authors contributed to the interpretation of the results and the discussion, and reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ready-to-eat breakfast cereals in Swedish supermarkets in 2019, categorized according to the FoodSwitch platform categories.
[DOCX File , 20 KB - mhealth_v9i7e17780_app1.docx]

References

1. Noncommunicable diseases: Childhood overweight and obesity. World Health Organization. 2020 Oct 10. URL: http://www.who.int/dietphysicalactivity/childhood/en/ [accessed 2019-10-22]
2. Lobstein T, Jackson-Leach R. Estimated burden of paediatric obesity and co-morbidities in Europe. Part 2. Numbers of children with indicators of obesity-related disease. Int J Pediatr Obes 2006 Jan;1(1):33-41. [doi: 10.1080/17477160600586689] [Medline: 17902213]
3. Singh A, Mulder C, Twisk J, van Mechelen W, Chinapaw M. Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes Rev 2008 Sep;9(5):474-478. [doi: 10.1111/j.1467-789x.2008.00475.x]
4. Juul F, Hemmingsson E. Trends in consumption of ultra-processed foods and obesity in Sweden between 1960 and 2010. Public Health Nutr 2015 Mar 25;18(17):3096-3107. [doi: 10.1017/s1368980015000506]
5. Costa CS, Del-Ponte B, Assunção MCF, Santos IS. Consumption of ultra-processed foods and body fat during childhood and adolescence: a systematic review. Public Health Nutr 2017 Jul 05;21(1):148-159. [doi: 10.1017/s1368980017001331]
6. Potvin Kent M, Rudnicki E, Usher C. Less healthy breakfast cereals are promoted more frequently in large supermarket chains in Canada. BMC Public Health 2017 Nov 13;17(1):877 [FREE Full text] [doi: 10.1186/s12889-017-4886-3] [Medline: 29132346]
7. van der Bend D, Lissner L. Differences and Similarities between Front-of-Pack Nutrition Labels in Europe: A Comparison of Functional and Visual Aspects. Nutrients 2019 Mar 14;11(3):626 [FREE Full text] [doi: 10.3390/nu11030626] [Medline: 30875797]
8. Regulations amending the National Food Agency's regulations (SLVFS 2005:9) on the use of a particular symbol. Swedish Food Agency. 2015 Jan 30. URL: http://extwprlegs1.fao.org/docs/pdf/swe199505.pdf [accessed 2019-11-05]
9. Health Star Rating System. Department of Health. Canberra: Australian Government URL: http://healthstarrating.gov.au/internet/healthstarratin [accessed 2019-04-11]
10. Kelly B, Jewell J. What is the evidence on the policy specifications, development processes and effectiveness of existing front-of-pack food labelling policies in the WHO European Region? In: Health Evidence Network Synthesis Report, No. 61. Copenhagen: WHO Regional Office for Europe; 2018.
11. Svenskarna och internet 2018. Internetstiftelsen. 2018 Oct 3. URL: https://2018.svenskarnaochinternet.se/?referer=www.svenskarnaochinternet.se [accessed 2019-05-23]
12. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: A Mobile Phone App to Enable Consumers to Make Healthier Food Choices and Crowdsourcing of National Food Composition Data. JMIR Mhealth Uhealth 2014 Aug 21;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]
13. Dunford E, Huang L, Peters S, Crino M, Neal B, Ni Mhurchu C. Evaluation of Alignment between the Health Claims Nutrient Profiling Scoring Criterion (NPSC) and the Health Star Rating (HSR) Nutrient Profiling Models. Nutrients 2018 Aug 10;10(8):1065. [doi: 10.3390/nu10081065]

14. Crino M, Sacks G, Dunford E, Trieu K, Webster J, Vandevijvere S, et al. Measuring the Healthiness of the Packaged Food Supply in Australia. Nutrients 2018 May 31;10(6):702 [FREE Full text] [doi: 10.3390/nu10060702] [Medline: 29857517]
15. Coyle D, Ndanuko R, Singh S, Huang P, Wu J. Variations in Sugar Content of Flavored Milks and Yogurts: A Cross-Sectional Study across 3 Countries. Curr Dev Nutr 2019 Jun;3(6):nzz060 [FREE Full text] [doi: 10.1093/cdn/nzz060] [Medline: 31187086]
16. Sweden: market share of selected grocery retailers 2017. Statista. 2018. URL: https://www.statista.com/statistics/565650/market-share-of-selected-grocery-retailers-in-sweden/ [accessed 2019-05-01]
17. Guide for Industry to the Health Star Rating Calculator. Commonwealth of Australia. 2019. URL: http://www.healthstarrating.gov.au/internet/healthstarrating/publishing.nsf/Content/E380CCCA07E1E42FCA257DA500196044/$File/Health-Star-Rating-system-Calculator-and-Style-Guide.pdf [accessed 2020-12-14]
18. Dunford E, Cobcroft M, Thomas M, Wu JH. Technical Report: Alignment of NSW Healthy Food Provision Policy with the Health Star Rating System. Sydney, NSW, Australia: NSW Ministry of Health; 2015.
19. Tong T, Rangan A, Gemming L. Evaluating the Nutritional Content of Children's Breakfast Cereals in Australia. Children (Basel) 2018 Jun 21;5(7):84 [FREE Full text] [doi: 10.3390/children5070084] [Medline: 29933610]
20. Devi A, Eyles H, Rayner M, Ni Mhurchu C, Swinburn B, Lonsdale-Cooper E, et al. Nutritional quality, labelling and promotion of breakfast cereals on the New Zealand market. Appetite 2014 Oct;81:253-260. [doi: 10.1016/j.appet.2014.06.019] [Medline: 24953195]
21. Soo J, Letona P, Chacon V, Barnoya J, Roberto CA. Nutritional quality and child-oriented marketing of breakfast cereals in Guatemala. Int J Obes (Lond) 2016 Jan 21;40(1):39-44. [doi: 10.1038/ijo.2015.161] [Medline: 26293234]
22. Potvin Kent M, Cameron C, Philippe S. The healthfulness and prominence of sugar in child-targeted breakfast cereals in Canada. Health Promot Chronic Dis Prev Can 2017 Sep;37(9):266-273 [FREE Full text] [doi: 10.24095/hpcdp.37.9.02] [Medline: 28902476]
23. Schwartz MB, Vartanian LR, Wharton CM, Brownell KD. Examining the nutritional quality of breakfast cereals marketed to children. J Am Diet Assoc 2008 Apr;108(4):702-705. [doi: 10.1016/j.jada.2008.01.003] [Medline: 18375229]
24. Dishonest health star ratings on Milo and Kellogg's cereals. CHOICE. 2016 Feb 15. URL: https://www.choice.com.au/food-and-drink/nutrition/food-labelling/articles/nestle-milo-kelloggs-cereals-misuse-health-star-ratings [accessed 2021-06-07]
25. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ 2012 Jan 15;346:e7492. [doi: 10.1136/bmj.e7492] [Medline: 23321486]
26. de Ruyter JC, Olthof MR, Seidell JC, Katan MB. A Trial of Sugar-free or Sugar-Sweetened Beverages and Body Weight in Children. N Engl J Med 2012 Oct 11;367(15):1397-1406. [doi: 10.1056/nejmoa1203034]
27. Vos MB, Kaar JL, Welsh JA, Van Horn LV, Feig DI, Anderson CA, et al. Added Sugars and Cardiovascular Disease Risk in Children: A Scientific Statement From the American Heart Association. Circulation 2017 May 09;135(19):e1017-e1034. [doi: 10.1161/cir.0000000000000439]
28. Slavin JL. Position of the American Dietetic Association: health implications of dietary fiber. J Am Diet Assoc 2008 Oct;108(10):1716-1731. [doi: 10.1016/j.jada.2008.08.007] [Medline: 18953766]
29. Williams P. The benefits of breakfast cereal consumption: a systematic review of the evidence base. Adv Nutr 2014 Sep 15;5(5):636S-673S [FREE Full text] [doi: 10.3945/an.114.006247] [Medline: 25225349]
30. European Commission, Health and Consumers. Survey on Members States' Implementation of the EU Salt Reduction Framework. Luxembourg: Publications Office of the European Union; Mar 07, 2014.
31. Pombo-Rodrigues S, Hashem KM, He FJ, MacGregor GA. Salt and sugars content of breakfast cereals in the UK from 1992 to 2015. Public Health Nutr 2017 Feb 07;20(8):1500-1512. [doi: 10.1017/s1368980016003463]
32. Lapierre MA, Vaala SE, Linebarger DL. Influence of licensed spokescharacters and health cues on children's ratings of cereal taste. Arch Pediatr Adolesc Med 2011 Mar 07;165(3):229-234. [doi: 10.1001/archpediatrics.2010.300] [Medline: 21383272]
33. Longacre MR, Drake KM, Titus LJ, Harris J, Cleveland LP, Langeloh G, et al. Child-targeted TV advertising and preschoolers' consumption of high-sugar breakfast cereals. Appetite 2017 Jan 01;108:295-302 [FREE Full text] [doi: 10.1016/j.appet.2016.10.014] [Medline: 27746213]
34. Levin AM, Levin IP. Packaging of healthy and unhealthy food products for children and parents: The relative influence of licensed characters and brand names. J Consum Behav 2010 Sep 28;9(5):393-402. [doi: 10.1002/cb.326]
35. Yee AZH, Lwin MO, Ho SS. The influence of parental practices on child promotive and preventive food consumption behaviors: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Apr 11;14(1):47 [FREE Full text] [doi: 10.1186/s12966-017-0501-3] [Medline: 28399881]
36. Kelly B, Hughes C, Chapman K, Louie JC, Dixon H, Crawford J, et al. Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market. Health Promotion International 2009 Mar 31;24(2):120-129. [doi: 10.1093/heapro/dap012]
37. van Raaij J, Hendriksen M, Verhagen H. Potential for improvement of population diet through reformulation of commonly eaten foods. Public Health Nutr 2009 Mar;12(3):325-330. [doi: 10.1017/S1368980008003376] [Medline: 18671891]

38. Mørk T, Grunert KG, Fenger M, Juhl HJ, Tsalis G. An analysis of the effects of a campaign supporting use of a health symbol on food sales and shopping behaviour of consumers. BMC Public Health 2017 Mar 09;17(1):239 [FREE Full text] [doi: 10.1186/s12889-017-4149-3] [Medline: 28274208]
39. Sjölin KL. Nordic Keyhole: Experience and challenges in Sweden, Norway, Denmark and Iceland [Presentation on the Internet]. Swedish National Food Agency. 2013 May 16. URL: https://www.who.int/nutrition/events/2013_FAO_WHO_workshop_frontofpack_nutritionlabelling_presentation_Sjolin.pdf?ua=1 [accessed 2019-05-02]
40. Jumpertz R, Venti CA, Le DS, Michaels J, Parrington S, Krakoff J, et al. Food label accuracy of common snack foods. Obesity (Silver Spring) 2013 Jan 12;21(1):164-169 [FREE Full text] [doi: 10.1002/oby.20185] [Medline: 23505182]
41. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011. Official Journal of the European Union. 2011 Nov 22. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0018:0063:EN:PDF [accessed 2019-05-02]

Abbreviations

HSR: Health Star Rating
mHealth: mobile health
NIP: Nutrition Information Panel

Edited by G Eysenbach, R Kukafka; submitted 13.01.20; peer-reviewed by Y Li, J Alvarez Pitti; comments to author 17.09.20; revised version received 19.10.20; accepted 07.05.21; published 22.07.21.

Please cite as:
Mottas A, Lappi VM, Sundström J, Neal B, Mhurchu CN, Löf M, Rådholm K
Measuring the Healthiness of Ready-to-Eat Child-Targeted Cereals: Evaluation of the FoodSwitch Platform in Sweden
JMIR Mhealth Uhealth 2021;9(7):e17780
URL: https://mhealth.jmir.org/2021/7/e17780
doi:10.2196/17780
PMID:34292165

Original Paper

Efficacy and Safety of Text Messages Targeting Adherence to Cardiovascular Medications in Secondary Prevention: TXT2HEART Colombia Randomized Controlled Trial

Anderson Bermon[1,2], MSc, MD; Ana Fernanda Uribe[3], PhD; Paula Fernanda Pérez-Rivero[3], MSc; David Prieto-Merino[4,5], PhD; Jose Federico Saaibi[6], MD; Federico Arturo Silva[7], MD; Diana Ivonne Canon[8], MD, FACC; Karol Melissa Castillo-Gonzalez[1], RN; Diana Isabel Cáceres-Rivera[9], PhD; Elizabeth Guio[10], MSc; Karen Janneth Meneses-Castillo[1], MD; Alberto Castillo-Meza[1], MD; Louise Atkins[11], PhD; Robert Horne[12], PhD; Elizabeth Murray[13], PhD; Norma Cecilia Serrano[14], MSc; Caroline Free[4], PhD; Juan Pablo Casas[15,16], PhD; Pablo Perel[17], PhD

[1]Research Center, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[2]Epidemiology an Biostatistics, Escuela de Graduados, Universidad CES, Medellín, Colombia

[3]Faculty of Psychology, Universidad Pontificia Bolivariana - Seccional Bucaramanga, Floridablanca, Colombia

[4]Epidemiology and Population Health Faculty, London School of Hygiene & Tropical Medicine, London, United Kingdom

[5]Applied Statistical Methods in Medical Research Group, Universidad Católica San Antonio de Murcia, Murcia, Spain

[6]Departament of Cardiovascular Surgery, Division of Vascular and Endovascular Surgery, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[7]Neurovascular Science Group, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[8]Departament of Cardiology, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[9]Nursing Faculty, Universidad Cooperativa de Colombia, Bucaramanga, Colombia

[10]Metabolism and Genoma Laboratory, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[11]Research Department of Epidemiology and Public Health, University College London, London, United Kingdom

[12]University College London School of Pharmacy, London, Colombia

[13]Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[14]Direction of Research, Fundación Cardiovascular de Colombia, Floridablanca, Colombia

[15]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), Boston, MA, United States

[16]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[17]Centre for Global Chronic Conditions, London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Anderson Bermon, MSc, MD
Research Center
Fundación Cardiovascular de Colombia
Calle 155A #23-58
Floridablanca, 681001
Colombia
Phone: 57 3002053041
Email: andebermon@gmail.com

Abstract

Background: Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of mortality worldwide, with a prevalence of approximately 100 million patients. There is evidence that antiplatelet agents and antihypertensive medications could reduce the risk of new vascular events in this population; however, treatment adherence is very low. An SMS text messaging intervention was recently developed based on behavior change techniques to increase adherence to pharmacological treatment among patients with a history of ASCVD.

Objective: This study aims to evaluate the efficacy and safety of an SMS text messaging intervention to improve adherence to cardiovascular medications in patients with ASCVD.

Methods: A randomized controlled clinical trial for patients with a prior diagnosis of cardiovascular events, such as acute myocardial infarction, unstable angina, cerebrovascular disease, or peripheral artery disease, in one center in Colombia was conducted. Patients randomized to the intervention arm were assigned to receive SMS text messages daily for the first 4 weeks, 5 SMS text messages on week 5, 3 SMS text messages each in weeks 6 and 7, and 1 SMS text message weekly from week 8 until week 52. In contrast, patients in the control arm received a monthly SMS text message reminding them of the next study appointment and the importance of the study, requesting information about changes in their phone number, and thanking them for participating in the study. The primary endpoint was the change in low-density lipoprotein cholesterol (LDL-C) levels, whereas the secondary endpoints were the changes in thromboxane B2 levels, heart rate, systolic and diastolic blood pressure, medication adherence, cardiac and noncardiac mortality, and hospitalization. Linear regression analyses and bivariate tests were performed.

Results: Of the 930 randomized patients, 805 (86.5%) completed follow-up and were analyzed for the primary endpoint. There was no evidence that the intervention changed the primary outcome (LDL-C levels; *P*=.41) or any of the secondary outcomes evaluated (all *P*>.05). There was also no evidence that the intervention was associated with adverse events.

Conclusions: In this study, there was no evidence that a behavior modification intervention delivered by SMS text messaging improved LDL-C levels, blood pressure levels, or adherence at 12 months. More research is needed to evaluate whether different SMS text messaging strategies, including personalized messages and different timings, are effective; future studies should include mixed methods to better understand why, for whom, and in which context (eg, health system or social environment) SMS text messaging interventions work (or not) to improve adherence in patients with ASCVD.

Trial Registration: ClinicalTrials.gov NCT03098186; https://clinicaltrials.gov/ct2/show/NCT03098186

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-028017

(JMIR Mhealth Uhealth 2021;9(7):e25548) doi:10.2196/25548

KEYWORDS

randomized controlled trial; Colombia; text messaging; cardiovascular disease; secondary prevention

Introduction

Background

Cardiovascular diseases are the leading cause of mortality worldwide. In 2017, approximately 17.5 million people died from cardiovascular diseases. Atherosclerotic cardiovascular disease (ASCVD) was responsible for 7.3 million deaths in 2007, which increased to 8.93 million in 2017. During the same period, mortality associated with cerebrovascular disease increased from 5.29 to 6.17 million events. Moreover, 82% of deaths in people ≤70 years occurred in low- and middle-income countries [1].

In 2015, more than 100 million people worldwide were diagnosed with ASCVD [2]. This population has been estimated to have a four- to five-fold increased risk of a new cardiovascular event in comparison with individuals without ASCVD history [3].

Robust evidence indicates that the use of antiplatelet agents, β-blocker agents, angiotensin-converting enzyme inhibitors (ACEIs), and statins reduces the incidence of fatal and nonfatal cardiovascular events in this population and is cost-effective. These medications are recommended by all international guidelines for the management of ASCVD [4,5].

However, long-term adherence to medication regimens continues to be suboptimal, and many patients stop medication for various reasons other than adverse side effects [6,7]. Only less than half of the patients with known ASCVD in high-income countries are receiving this group of cardiovascular medications, and the situation is much worse in low- to middle-income countries (LMICs), where only 1 in 20 patients with ASCVD received all four types of cardiovascular drugs in 2011 [8].

The widespread use of mobile devices allows the implementation of strategies such as text messaging to increase medication adherence. It has shown some promising results among patients with diabetes [9], HIV [10], and tuberculosis [11] and may therefore help improve adherence for patients with ASCVD [12,13]. In addition, access to the use of mobile telephones globally has increased in recent years. For example, in Colombia, telephone coverage increased from 84% in 2009 to 98.1% in 2019 [14].

A 2017 Cochrane review [15] evaluated the effects of SMS text messaging on medication adherence in patients with ASCVD. The review included 7 trials (n=1310) and reported the beneficial effect of SMS text messaging on adherence to medications in 6 of these trials. However, the quality of the evidence was very low. The Cochrane review identified the following limitations: (1) trials had small sample sizes (n=34-521); (2) most trials had a short follow-up period (<6 months); (3) the primary outcomes reported were of limited clinical relevance; (4) most studies recruited only patients with acute coronary syndrome and excluded an important group of patients with other arterial occlusive events (eg, stroke, peripheral vascular disease, and programmed coronary revascularization) who should be amenable for this type of intervention; (5) few studies were performed in LMICs; and (6) most trials did not describe the processes for SMS text messaging content generation, and the few trials that did report these processes did not target the key knowledge and attitudinal factors that are known to influence adherence to medication; instead, the interventions were simple *reminders* [15]. In summary, although there are some promising small studies, there is a need to provide high-quality evidence to assess the effect of SMS text messaging using behavior change techniques to increase long-term medication adherence in patients with ASCVD in LMICs.

Aims

This study aims to fill this gap and provide evidence on whether theory-based and context-specific SMS text messages increase medication adherence for the secondary prevention of ASCVD in Colombia. We developed an intervention (ie, SMS text message) following the recommendations of Abroms et al [16]: a review of the literature, conduct of qualitative studies, and use of formal theories and behavior change techniques (Transtheoretical Model of Behavior Change). Details of our intervention development have been described previously [17,18].

The main aim of this study is to evaluate the efficacy and safety of an SMS text messaging intervention delivered by mobile phones to improve adherence to cardiovascular medications in patients with ASCVD. The intervention efficacy was assessed by measuring blood serum low-density lipoprotein cholesterol (LDL-C) levels as an indicator of adherence to statins, systolic blood pressure (SBP) as an indicator of adherence to blood-lowering therapies (ACEI or angiotensin II receptor blockers [ARBs]), and heart rate (HR) as an indication of adherence to β-blockers. The secondary objectives are to assess the impact of SMS text messaging on self-reported adherence to medications, hospitalizations, and the composite end point of incident major adverse cardiovascular events at 12 months.

Methods

The full methodology of TXT2HEART Colombia has been previously published [17] and is summarized here. We report the following CONSORT (Consolidated Standards of Reporting Trials) recommendations [19].

Study Design and Participants

In this two-arm parallel, single-blind individually randomized controlled trial, adult patients aged ≥18 years with a history of at least one of the following arterial occlusive events were included: acute coronary syndrome (unstable angina or acute myocardial infarction with or without ST elevation), stable angina, ischemic cerebrovascular disease, peripheral arterial disease, or coronary revascularization (coronary artery bypass surgery or percutaneous transluminal coronary angioplasty). Patients had to own a mobile phone and were able to read the SMS text message. They were excluded if they had a known contraindication to take all appropriate cardiovascular secondary prevention medications. All patients were recruited from a single center, the Fundación Cardiovascular de Colombia, a tertiary hospital serving as a reference center for cardiovascular diseases in Northeastern Colombia. The hospital has a clinical studies office and has been certified in good clinical practice by national and international authorities. All electronic health records were scanned using SQL queries, identifying patients with at least one month and without a maximum limit of time elapsed since the last hospitalization for ASCVD. The records were then manually inspected by 2 experienced medical doctors. Qualifying patients were contacted by phone, and if they met the inclusion criteria and were currently admitted or attended the outpatient clinic with a diagnosis of ASCVD, they were invited to participate. The process for evaluating potentially eligible individuals is described in Multimedia Appendix 1. Written informed consent was obtained from all subjects before the study.

Intervention

The intervention consisted of behavior modification techniques based on the Transtheoretical Model [16] to be delivered via SMS text messaging. In our previous study [18], a protocol was carried out to determine the content, quantity, and frequency of SMS text messages through focus groups, validation of experts, user feedback, and pretest. The messages included information on the health implications of adherence to health habits (or lack thereof) and indications and recommendations on how to take their medication and promote healthy medication habits. They provided or encouraged social support activities for correct compliance with the prescribed treatment. The result of that study was 86 SMS text messages (including 12 SMS text messages of control and one welcome to the study), which were the messages used as an intervention in this clinical trial and the methodology of its delivery.

SMS text messages were sent through an automated text messaging platform (Telerivet), which was fed directly with data registered in Commcare, the platform for patient registration. Volunteers were informed about the unidirectional nature of the text messages and warned that no replies were expected. If the patients replied the word “PARE” or “Detener” (stop in Spanish), then no more messages were delivered. Text messaging was started a day after patient randomization. Messages were delivered every day for the first 4 weeks, and then five messages were delivered in week 5. From week 6 onward, three messages were delivered per week; from week 8 until week 52, one message was delivered per week. Messages were delivered on random weekdays from 8 AM to 6 PM to prevent patients from predicting delivery times, in accordance with a previous validation with study subjects. If the patient withdrew from the study or died, we stopped sending the SMS text messages. No tailoring considerations or modifications were made during the trial.

Control Group

Patients in the control group only received SMS text messages regarding the next study appointment, requesting information about changes in their phone number, acknowledging them for participating in the study, and reminding them of the importance of the study. Messages were sent every month. These messages were also sent to the intervention group and were generated during SMS text message validation in the general population.

Examples of TXT2HEART Colombia SMS text messages are included in Multimedia Appendix 2.

Outcomes

The primary outcome was a change in plasma LDL-C levels at 12 months. Blood samples were obtained at the start and end of the study appointment. An improvement in LDL-C levels was considered a surrogate indicator of adherence to statin treatment. The secondary outcomes were SBP as an indicator of adherence to blood-lowering therapies (ACEI or ARBs), HR as an indicator of adherence to β-blockers, 11-dehydrothromboxaneB2 urine levels adjusted for creatinine

as an indicator of adherence to antiplatelet therapy, self-reported adherence to cardiovascular medications used in secondary prevention as measured using the Medication Adherence Report Scale-5 (MARS-5) questionnaire, and rates of cardiovascular death or hospitalization due to cardiovascular disease and noncardiovascular death or hospitalization due to noncardiovascular disease. We also included road traffic crashes (the only potential known hazard of SMS text messaging) and death due to all causes as secondary outcomes.

The psychometric properties of the MARS-5 have been previously reported [20]. The MARS-5 demonstrated acceptable reliability (internal and test-retest) and validity (criterion-related and construct validity). Internal reliability (Cronbach α) ranged from .67 to .89 across all patient groups; the test-retest reliability (Pearson *r*) was 0.97 for hypertension. Criterion-related validity was established with more adherent patients with hypertension showing better blood pressure control (χ^2_1=4.2; *P*=.04). Construct validity with beliefs about medicines was demonstrated; higher adherence was associated with stronger beliefs in treatment necessity and lower concerns about the medication.

All study participants were seen twice upon admission to the study for baseline assessment and randomization and at the end of the follow-up period for a 12-month office visit. A follow-up telephone call was made 3 months after randomization asking about new hospital admissions, all-cause death, or cardiovascular death, and adverse events were also recorded. SBP, resting HR, and urinary levels of thromboxane B2 were recorded at the first and final visits. Self-informed cardiovascular medication prescription compliance was assessed using the MARS-5 at both visits. The scale was applied by trained personnel, considering automatic compliance if a total score of 25 was achieved. Subjective medication intake compliance was assessed on days 7 and 30. Data obtained about recurrent ASCVD were requested on the phone interview or by physical examination; oral reports by patients or relatives were allowed; and cardiovascular or any cause mortality was recorded. Information obtained on the phone was confirmed in all cases by reviewing medical case notes, registries, or death certificates. Written evidence for any event was requested via electronic mail, WhatsApp messaging, or case note copies. If death or any major event or hospital admission occurred during the follow-up period, a hard copy of the death certificate from the patients' relatives or case notes was requested on the 12-month follow-up visit. At the final follow-up, all biomarkers were processed simultaneously to avoid interference due to the reactive processing, and the simple handling protocol depicted in the protocol was followed [17]. At the initial appointment, we recorded at least three different phone numbers and a complete house address for each subject. A study identification card was provided with a written record of the date of the last visit to the trial, name of the principal investigator, and clinic contact phone numbers. To ensure no loss of follow-up, any home-number modifications were actively searched and recorded. The appointment follow-up interval was kept to avoid any interference with the study results.

Sample Size

The original sample size of the study was 1600 participants, based on a 97% power to detect a 10% difference between arms in adherence. In the published protocol [17] of this trial, a table with power calculations under different assumptions was included. However, due to limited study funding, the final sample recruited was 930 participants, 805 (86.5%) of whom reported LDL-C at the final visit. With 400 patients per arm (based on an expected mean reduction of 80 mg/dl on plasma levels of LDL-C in an adherence population and an expected mean reduction of 16 mg/dl on a nonadherent population to atorvastatin 20 mg) and assuming a 5% type I error, a power >92% was estimated to detect a 10% difference in protocol compliance between both intervention arms. A table with power calculations for this sample size under different assumptions is included in Multimedia Appendix 3.

Randomization

Block randomization was used, with block sizes of 5 patients each in a 1:1 allocation ratio, and assignment was done automatically using a remote computer-based randomization. Once the patient met the inclusion and exclusion criteria and signed the consent form, the data capture platform Commcare (Dimagi) applied a logarithm of randomization assigning the arm for the patient. This information was not shown to the interviewer to maintain blindness, although he did confirm its effective completion on the digital form. The data capture platform accessed the services of the SMS text messaging platform (Telerivet), categorizing the SMS text messaging group to send according to the assigned group.

Blinding

Owing to the nature of the intervention, the participants could not be blinded. However, all investigation personnel inputting data were blinded to the individual's group assignment, and all patients were asked not to reveal their allocation details to the study personnel. The study had an engineer who was the only person who could access the messaging and database platforms. He could access the data to sort patient queries or help solve reception or technical issues. He was specifically trained on the importance of maintaining blinding. Investigators handling and analyzing the data were blinded to the intervention assigned.

Statistical Analysis

The distributions of the baseline characteristics were compared between the intervention and control groups for all randomized patients, those who completed the follow-up and those who did not complete the follow-up, performing an intention to treat analysis. Analysis of the continuous outcomes (ie, LDL-C, thromboxane, HR, SBP, diastolic blood pressure [DBP], and quantitative measures of adherence) was performed using linear regression models. In each model, the dependent variable was the difference between 12-month follow-up and the baseline of the outcome, and the main explanatory variable was the intervention group. Furthermore, the Patient Health Questionnaire-9 (PHQ-9) scale (to measure depression) was collected at baseline, as it was considered a potential confounding factor for adherence to medication. All linear regression models were adjusted by the baseline value of the

outcome, centered on the mean. The effect of the intervention was the difference between the arms in the expected change in an individual and the average outcome value at baseline. Binary outcomes (ie, hospitalization and mortality) were analyzed by comparing the proportion of occurrence in both arms. No adverse events were reported; therefore, we did not conduct any analyses regarding this outcome. All *P* values were from two-sided tests, and the data were analyzed using Stata version 14.0 (StataCorp). No interim analyses were performed.

Ethical Considerations

The Ethics Committee of the Fundación Cardiovascular de Colombia evaluated and approved the trial (reference 375-2015). The study was conducted in compliance with the protocol, regulatory requirements, Good Clinical Practice, the Declaration of Helsinki, and the clinical investigations guidelines of the Fundación Cardiovascular de Colombia.

Data Availability

The data sets generated during this study are not publicly available because availability was not included in the study plan approved by the ethics committee but are available from the corresponding author on reasonable request.

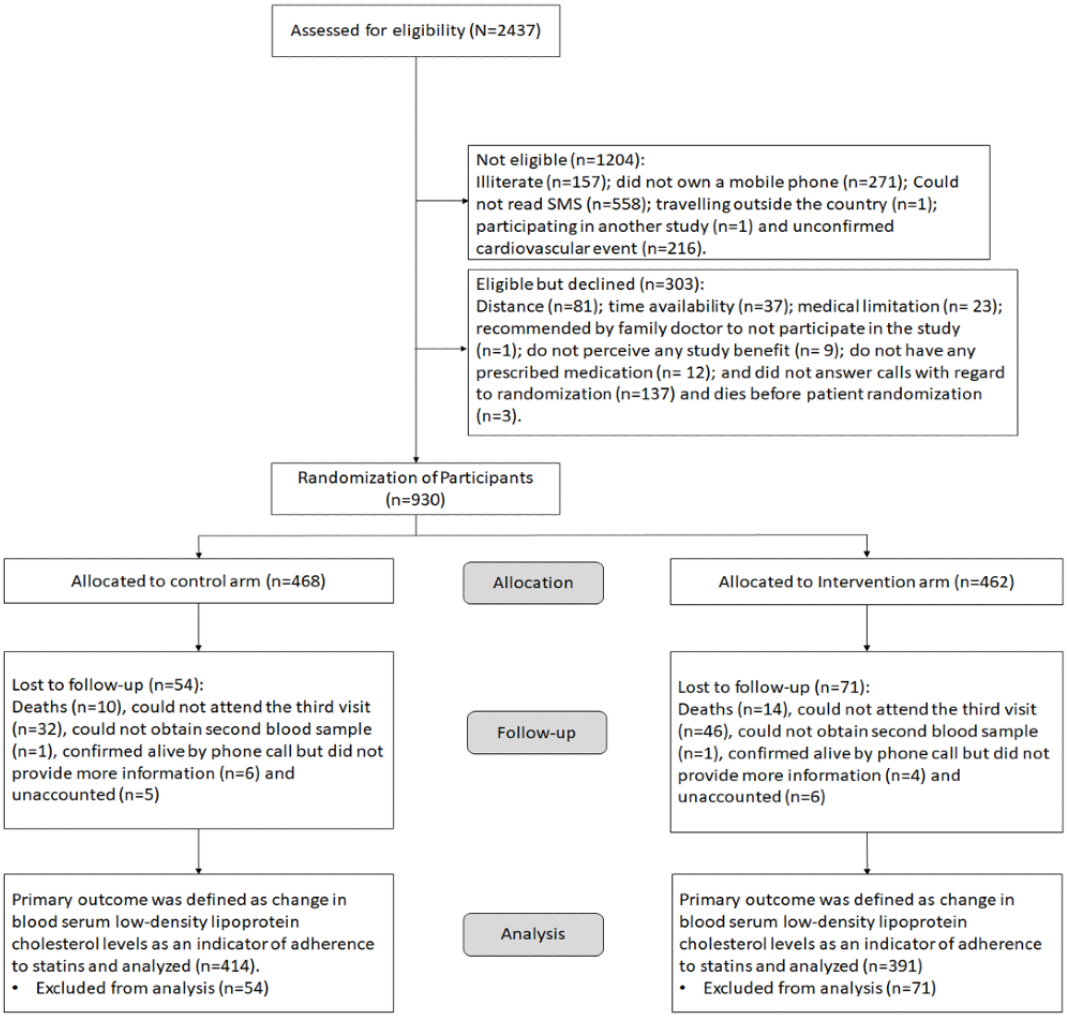
Results

Participant Flow

From April 18, 2017, to August 21, 2018, 930 patients were randomized. A total of 49.7% (462/930) of patients were assigned to the intervention arm and 50.3% (468/930) were included in the control arm (Figure 1), of which 1.7% (16/930) of patients replied “stop” to the messages (6 in intervention group and 10 control), all of whom were followed up until the end of the study. In total, 13.4% (125/930) losses to follow-up occurred, 71 in the intervention group and 54 in the control group.

A total of 86.8% (805/930) of participants completed the trial follow-up at 12 months for the primary outcome (intervention group: n=391; control: n=414; Figure 1). Retention did not differ between two arms (71/462, 15.4% in the intervention group vs 54/468, 11.5% in the control group; *P*=.09). The main predictors of retention were male sex (OR 1.61, 95% CI 1.05-2.46; *P*=.03) and high total PHQ-9 score (OR 0.37, 95% CI 0.15-0.92; *P*=.03). The effect of these predictors did not differ between the groups (interaction test values: *P*=.23 and *P*=.80, respectively). The characteristics of the participants who completed the follow-up and those who did not are reported in Multimedia Appendix 4. For secondary outcomes, HR, SBP, and DBP of 850 patients were evaluated. For TxBA2, only 801 patients were evaluated because 4 patients were unable to deliver the urine sample. In addition, 807 patients were assessed using the MARS-5. New cardiovascular events were evaluated in 910 patients and mortality was evaluated in 919 patients through telephone interviews.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.



Baseline Characteristics

The baseline characteristics of the participants (N=930), which were similar between the two groups, are summarized in Table 1. Overall, most participants (729/930, 78.4%) were men, and the mean age was 63.5 years (SD 9.8); 88.9% (827/930) were using statins, 94.52% (879/930) were on antiplatelet aggregation therapy, and β-blocker use was reported in 83.2% (774/930) of participants, whereas ACEI or ARBs use was observed in 68.1% (633/930) of participants. The average MARS-5 score at baseline was 22.8, and 41.72% (388/930) of participants considered adherent (MARS-5 score=25). Clinical and laboratory characteristics such as PHQ-9, BMI, LDL-C, SBP, DBP, HR, and thromboxane B2 were similar between the two groups (Table 1). The baseline characteristics of participants with primary outcome completers and primary outcome noncompleters are shown in Multimedia Appendix 4, and the MARS-5 scores at baseline are shown in Figure 2.

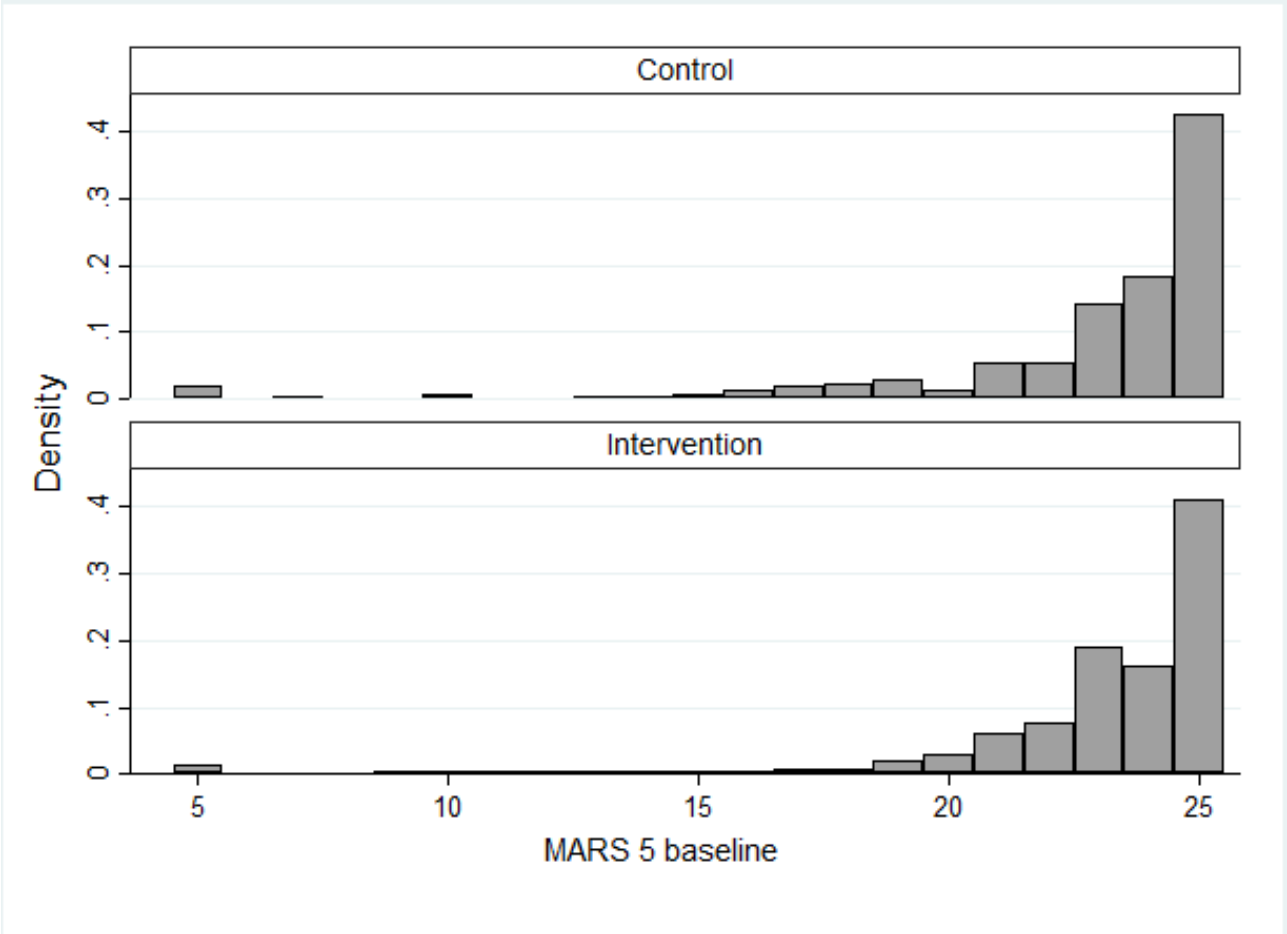
Table 1. Characteristics of study participants at baseline (N=930).

Characteristics	Control group (n=468)	Intervention group (n=462)	Participants (N=930)
Age (years), mean (SD)	63.1 (10)	64.0 (9.7)	63.5 (9.8)
Gender, n (%)			
Female	92 (19.7)	109 (23.6)	201 (21.6)
Male	376 (80.3)	353 (76.4)	729 (78.4)
Time since the last event, n (%)			
Less than 3 months	62 (13.3)	43 (9.3)	105 (11.3)
3 to 12 months	84 (17.9)	97 (21)	181 (19.5)
1 to 3 years	144 (30.8)	122 (26.4)	266 (28.6)
More than 3 years	178 (38)	200 (43.3)	378 (40.7)
Type event, n (%)			
Acute coronary syndrome	336 (71.8)	327 (70.8)	663 (71.3)
Stable angina	33 (7.1)	23 (5)	56 (6)
Ischemic cerebrovascular disease	21 (4.5)	21 (4.6)	42 (4.5)
Peripheral arterial disease	15 (3.2)	21 (4.6)	36 (3.9)
Coronary revascularization	63 (13.5)	70 (15.2)	133 (14.3)
Prescribed medications, n (%)			
Statins	414 (88.5)	413 (89.4)	827 (88.9)
ACEI[a] or ARBs[b]	327 (69.9)	306 (66.2)	633 (68.1)
β-blocker	382 (81.6)	392 (84.9)	774 (83.2)
Platelet aggregation inhibitors	442 (94.4)	437 (94.6)	879 (94.5)
MARS-5[c] score, mean (SD)	22.8 (3.78)	23 (3.31)	22.9 (3.6)
Adherent (MARS-5 score=25 points), n (%)	199 (42.5)	189 (40.9)	388 (41.7)
Self-reported adherence, mean (SD)			
Last 7 days (0-10 scale)	9.1 (2.04)	9.1 (2.19)	9.1 (2.1)
30 days (0-10 scale)	9.1 (1.95)	9.1 (2.05)	9.1 (2)
PHQ-9[d], n (%)			
Minimal depression (<5)	343 (73.3)	323 (69.9)	666 (71.6)
Moderate depression (5-14)	112 (24)	127 (27.5)	239 (25.7)
Moderately severe depression or severe (>14)	12 (2.6)	13 (2.8)	25 (2.7)
Smoking, n (%)			
Smoker	16 (3.4)	12 (2.6)	28 (3)
Never smoked	168 (35.9)	185 (40)	353 (38)
Past smoker	284 (60.7)	265 (57.4)	549 (59)
BMI (m/kg^2), mean (SD)	27.9 (4.2)	27.3 (4.2)	27.6 (4.2)
LDL[e] (mg/dl), mean (SD)	88.2 (37.4)	88.5 (38.0)	88.4 (37.7)
SBP[f] (mmHg), mean (SD)	128.2 (20.8)	129.9 (20.9)	129.0 (20.9)
DBP[g] (mmHg), mean (SD)	71.5 (11.7)	71.9 (11.3)	71.7 (11.5)
Heart rate (bpm), mean (SD)	68.9 (11.7)	68.8 (10.6)	68.8 (11.1)
Thromboxane B2[h] (ng/ml), mean (SD)	64.1 (147.2)	64.2 (167.7)	64.2 (157.6)

[a]ACEI: angiotensin-converting enzyme inhibitor.

[b]ARB: angiotensin II receptor blocker.
[c]MARS-5: Medication Adherence Report Scale-5.
[d]PHQ-9: Patient Health Questionnaire-9.
[e]LDL: low-density lipoprotein.
[f]SBP: systolic blood pressure.
[g]DBP: diastolic blood pressure.
[h]Creatinine adjusted (mg/dl).

Figure 2. Distribution of the MARS-5 scores at baseline. Graphs by Intervention. MARS-5: Medication Adherence Report Scale-5.



Outcomes

We did not find evidence (P=.41) that the intervention was more effective than the control according to changing plasma LDL-C levels (adjusted by baseline value) at 12 months. We also did not find significant differences between the two groups in terms of secondary outcomes analyzed including thromboxane B2 levels, HR, SBP, DBP, adherence measured by MARS-5, or clinical events (hospitalization or death) at one year of follow-up (Table 2).

There were no adverse events related to this study. In total, three falls that required hospitalization were reported; in all of them, medical evaluation and patient interviews were performed, eliminating any relationship between mobile phone use and fall events. No traffic accidents occurred.

Table 2. Summary of primary and secondary outcomes[a].

Outcome	Baseline		Difference[b]		Difference adjusted by baseline			
	Control	Intervention	Control (n=468)	Intervention (n=462)	Coefficient (95% CI)	Odds ratio (95% CI)	Risk ratio (95% CI)	*P* value
Primary outcome, mean (SD)								
LDL[c] (mg/dl)	88.0 (36.9)	88.0 (37.5)	5.1 (31.8)	7.0 (33.8)	1.85 (−2.5 to 6.2)	N/A[d]	N/A	.42
Secondary outcome								
Thromboxane B2 (ng/ml)[e], mean (SD)	61.2 (133.2)	58.8 (138.1)	−19.6 (131.0)	−18.6 (94.0)	−0.28 (−10.54 to 10.0)	N/A	N/A	.96
Heart rate (bpm), mean (SD)	68.6 (11.6)	68.5 (10.5)	−0.1 (13.9)	0.6 (10.5)	0.54 (−1.0 to 2.1)	N/A	N/A	.48
SBP[f] (mmHg), mean (SD)	128.0 (21.2)	129.3 (20.6)	1.3 (19.4)	0.8 (20.7)	0.14 (−2.3 to 2.6)	N/A	N/A	.91
DBP[g] (mmHg), mean (SD)	71.5 (11.9)	71.7 (11.3)	0.7 (11.7)	−0.1 (10.7)	−0.70 (−2.0 to 0.6)	N/A	N/A	.30
MARS-5[h] (score), mean (SD)	22.8 (3. 8)	23.1 (3.1)	0.2 (3.7)	−0.02 (3.4)	−0.01 (−0.4 to 0.4)	N/A	N/A	.96
Self-reported adherence (7 days), mean (SD)	9.1 (2.1)	9.2 (2.0)	0.1 (2.0)	0.2 (2.0)	0.05 (−0.2 to 0.3)	N/A	N/A	.69
Self-reported adherence (30 days), mean (SD)	9.1 (2.0)	9.2 (1.9)	0.1 (1.9)	0.1 (2.0)	0.02 (−0.2 to 0.2)	N/A	N/A	.83
Change in adherence[i], mean (SD)	N/A	N/A	1.2 (0.2)	1.1 (0.20)	N/A	0.94 (0.6 to 1.5)	N/A	.81
Hospitalization for cardiovascular events[j], n (%)	N/A	N/A	32 (6.8)	27 (5.8)	N/A	N/A	0.85 (0.5 to 1.4)	.54
Hospitalization for any cause[k], n (%)	N/A	N/A	49 (10.5)	50 (10.8)	N/A	N/A	1.03 (0.71 to 1.5)	.92
Death from cardiovascular events[l], n (%)	N/A	N/A	3 (0.6)	2 (0.4)	N/A	N/A	0.68 (0.1 to 4.0)	.99
Death from any cause[j], n (%)	N/A	N/A	8 (1.7)	14 (3)	N/A	N/A	1.80 (0.8 to 4.3)	.20

[a]Sample sizes may vary slightly because some individuals have missing values.

[b]Calculated as the difference between final and baseline measurements.

[c]LDL: low-density lipoprotein.

[d]N/A: not applicable.

[e]Creatinine adjusted (mg/dl)

[f]SBP: systolic blood pressure.

[g]DBP: diastolic blood pressure.

[h]MARS-5: Medication Adherence Report Scale-5.

[i]Within each group is the number of patients that lose adherence over those that became adherent.

[j]Within each group is the proportion of patients who experienced the event.

Discussion

Principal Findings

In this study, a behavioral modification intervention delivered by SMS text messages did not decrease LDL-C level, the primary outcome, and did not find evidence of an impact on any of the other biological markers assessed, including blood pressure, HR, or thromboxane; clinical events; or medication adherence, as measured by the MARS-5 or self-reporting.

This study has many strengths. First, a thorough and detailed formative research was followed to develop a tailored behavioral modification intervention, which has been previously published [17] alongside the study protocol. Different outcomes have been collected, including self-reporting of adherence, validated adherence tools (ie, the MARS-5 has demonstrated acceptable reliability [internal and test-retest] and validity [criterion-related and construct validity]) [20], and proxy biological markers, and they were triangulated to evaluate adherence. A rigorous plan was put in place to ensure that the intervention was delivered appropriately, study investigators collecting data were blinded to the patient allocation arm, and most patients were followed up for the primary and secondary outcomes at 12 months. Finally, even if the sample size was smaller than originally planned, to the best of our knowledge, this study is the largest to date to assess the effect of a behavioral modification intervention delivered by SMS text messages to increase adherence in people with ASCVD.

Limitations

This study also presented some limitations; although sending text messages could be confirmed, neither the correct reading of the message nor the stage of the transtheoretical behavioral modification reached by each participant was evaluated [18]. In addition, the inability to blind patients would have increased the likelihood of underreporting nonadherence. This is a common problem with self-reports when patients may exaggerate their adherence if they believe that reports of nonadherence will disappoint their health provider (self-presentational bias) [21]. The MARS-5 addresses this problem by taking steps to diminish self-presentational bias. Introductory statements normalize nonadherence, conveying a *no-blame* approach [8]. As another step to minimize self-presentational bias, patients were told that their responses to the study questionnaires would not be seen by the health care professionals providing care. In addition, we conducted a case analysis on all lost cases, assuming that all related losses were random, but due to the low rate of loss to follow-up and that its main predictors (sex and PHQ-9) were similar in both arms of the study, we do not consider that these losses had a major impact on the results of the trial.

Another limitation was that SMS text messages were not customized according to the patient categories. However, as mentioned earlier, this formative research was detailed, and a formal theory was applied to develop the content of the messages. Unfortunately, due to the lack of funding, we could not conduct a comprehensive mixed methods process evaluation to shed light on some of the mechanisms and contexts that could explain the effect on specific subgroups of patients. Finally, measuring adherence is always challenging; a validated scale (ie, MARS-5) was used and self-reporting measures were used, but pill count was not taken into consideration; however, the direct measurement of adherence was complemented with indirect measurements, such as those for LDL-C, blood pressure, HR, and thromboxane.

A previous Cochrane review [15] published in 2017 identified studies reporting positive effects; however, all studies had small sample sizes and high risk of bias and therefore provided a low level of evidence. The TEXT ME study was a randomized controlled trial that included 710 patients with coronary artery disease who did not report medication adherence but did report that four text messages sent per week led to reductions in LDL-C levels and blood pressure at 6 months [22]; however, a more recent multi-center randomized controlled trial in China including 822 patients with coronary artery disease reported that SMS text messages did not reduce blood pressure or LDL-C levels at 6 months [23].

The lack of evidence of a beneficial effect reported by the intervention could be due to different reasons; the study might have lacked the power to detect an effect as the sample size was lower than intended; however, with 930 patients, it was well powered to detect a reasonable, modest, clinical benefit. The lack of benefit could be potentially explained by the fact that the baseline levels of medication adherence were already quite high, with a mean MARS-5 of 9.1 out of 10 at 7 and 30 days; therefore, there was little room for improvement. Another possible explanation could be related to intervention content. Although the intervention went beyond simple prompts and reminders (a thorough process was followed with the objective of changing beliefs and motivations), it was not tailored enough to change beliefs and motivations in the study context, or perhaps in this highly adherent population, these issues were not the main drivers of nonadherence. The findings could also be related to issues related to the delivery of the SMS text messages (ie, timing, frequency, and length of intervention), due to the effects of fatigue, overload, or loss of interest in SMS text messages. SMS text messaging interventions have been shown to be effective in modifying lifestyle issues (eg, tobacco cessation) where the goals of the patient and the intervention are closely aligned; however, they might be less effective on adherence [24]. Another possible explanation is that the results were measured at 12 months, and it has been shown that adherence interventions are more effective in the short term (3 to 6 months) [25]. Planning only one intermediate measure at 3 months to assess the presence of rehospitalizations or death may limit the comparison of this study with others, but the protocol prioritized the pragmatic conditions of the trial, avoiding face-to-face contacts (eg, adherence questions), because this would represent an additional study activity to the real scenario of the patients, leading to a potential Hawthorne effect that can alter the results of the effectiveness of SMS text messages. Another explanation for the difference found in the studies included in the Cochrane review could be related to the fact that only a third of the patients recruited in our study had an index event within the last year.

Finally, having data from a single site is a limitation in generalizing our results to different scenarios; however, it is

important to note that the institution where the study was conducted is a reference center in Northeastern Colombia, and its area of influence included a population of 5 million.

Conclusions

This study did not find evidence that a behavior modification intervention delivered by SMS text messaging improves medication adherence, LDL-C, or blood pressure levels at 12 months. Although its potential use to increase medication adherence in patients with ASCVD remains as a potentially attractive and scalable solution to a very important problem, further research is needed. Future research should include interventions, including SMS text messages blended with other components, tailored to the specific issues and beliefs of individual participants. Implementation studies using mixed methods and innovative approaches that evaluate how different intervention characteristics (behavioral modification component), SMS text message delivery strategies (eg, timing of initiation, frequency, duration, and personalization), and context (eg, type of patients, health system, and social environment) influence the effect of this intervention should be conducted.

Acknowledgments

This work was supported by the Ministerio de Ciencia Tecnología e Innovación (code: 656672553352; grants 899-2015 and 753 de 2016); Fundación Cardiovascular de Colombia, Floridablanca; UK Medical Research Council Funded Reference (reference number: MR/N021304/1); and the Universidad Pontificia Bolivariana, Bucaramanga.

Authors' Contributions

JPC, PP, CF, NCS, AB, AFU, EM, RH, and LA carried out the conceptualization of the study. JPC, PP, NCS, AB, and KMCG organized the logistics of the clinical trial. AB, JPC, PP, CF, AFU, PP, RH, JFS, FAS, DICR, KMCG, and PFPR evaluated the questionnaires to be applied. JFS, FAS, DIC, ACM, KJMC, PFPR, and KMCG perform validation of the clinical criteria of the study. DPM and AB performed the statistical analysis plan and performed all the statistical analyzes; DPM and PP validated the statistical results. DPM, PP, and AB made the formal presentation of the results. PP, JPC, NCS, AFU, and AB acquired and managed the funds. NCS and EG validated the laboratory results. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluating potential eligible individuals.
[PNG File , 252 KB - mhealth_v9i7e25548_app1.png]

Multimedia Appendix 2
Example of the SMS text messaging intervention.
[DOCX File , 36 KB - mhealth_v9i7e25548_app2.docx]

Multimedia Appendix 3
Sample size calculations.
[DOCX File , 15 KB - mhealth_v9i7e25548_app3.docx]

Multimedia Appendix 4
Baseline characteristics.
[DOCX File , 44 KB - mhealth_v9i7e25548_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (v.1.6.1).
[PDF File (Adobe PDF File), 1208 KB - mhealth_v9i7e25548_app5.pdf]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. J Am Coll Cardiol 2017 Jul 4;70(1):1-25 [FREE Full text] [doi: 10.1016/j.jacc.2017.04.052] [Medline: 28527533]

3. Kerr AJ, Broad J, Wells S, Riddell T, Jackson R. Should the first priority in cardiovascular risk management be those with prior cardiovascular disease? Heart 2009 Feb;95(2):125-129. [doi: 10.1136/hrt.2007.140905] [Medline: 18381374]
4. Yusuf S. Two decades of progress in preventing vascular disease. Lancet 2002 Jul 6;360(9326):2-3. [doi: 10.1016/S0140-6736(02)09358-3] [Medline: 12114031]
5. Model List of Essential Medicines, 21st List. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/325771/WHO-MVP-EMP-IAU-2019.06-eng.pdf?ua=1 [accessed 2020-01-12]
6. Wei MY, Ito MK, Cohen JD, Brinton EA, Jacobson TA. Predictors of statin adherence, switching, and discontinuation in the USAGE survey: understanding the use of statins in America and gaps in patient education. J Clin Lipidol 2013;7(5):472-483. [doi: 10.1016/j.jacl.2013.03.001] [Medline: 24079289]
7. Lauffenburger JC, Robinson JG, Oramasionwu C, Fang G. Racial/Ethnic and gender gaps in the use of and adherence to evidence-based preventive therapies among elderly Medicare Part D beneficiaries after acute myocardial infarction. Circulation 2014 Feb 18;129(7):754-763 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.002658] [Medline: 24326988]
8. Yusuf S, Islam S, Chow CK, Rangarajan S, Dagenais G, Diaz R, Prospective Urban Rural Epidemiology (PURE) Study Investigators. Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE Study): a prospective epidemiological survey. Lancet 2011 Oct 1;378(9798):1231-1243. [doi: 10.1016/S0140-6736(11)61215-4] [Medline: 21872920]
9. Haider R, Sudini L, Chow CK, Cheung NW. Mobile phone text messaging in improving glycaemic control for patients with type 2 diabetes mellitus: A systematic review and meta-analysis. Diabetes Res Clin Pract 2019 Apr;150:27-37. [doi: 10.1016/j.diabres.2019.02.022] [Medline: 30822496]
10. 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 Syst Rev 2012 Mar 14(3):CD009756 [FREE Full text] [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
11. Nglazi MD, Bekker L, Wood R, Hussey GD, Wiysonge CS. Mobile phone text messaging for promoting adherence to anti-tuberculosis treatment: a systematic review. BMC Infect Dis 2013 Dec 2;13:566 [FREE Full text] [doi: 10.1186/1471-2334-13-566] [Medline: 24295439]
12. Palmer MJ, Henschke N, Villanueva G, Maayan N, Bergman H, Glenton C, et al. Targeted client communication via mobile devices for improving sexual and reproductive health. Cochrane Database Syst Rev 2020 Jul 14;8:CD013680. [doi: 10.1002/14651858.CD013680] [Medline: 32779730]
13. Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, Zheng G, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015 Sep;12(9):e1001876 [FREE Full text] [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
14. Estrategia Integral para mejorar las condiciones de prestación de servicios fijos y móviles en Colombia. MinTIC. Bogotá; 2020 Jun. URL: https://www.mintic.gov.co/portal/715/articles-145908_recurso_1.pdf [accessed 2021-02-02]
15. Adler AJ, Martin N, Mariani J, Tajer CD, Owolabi OO, Free C, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4:CD011851 [FREE Full text] [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
16. Abroms LC, Whittaker R, Free C, van Alstyne J, 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]
17. Bermon A, Uribe-Rodríguez AF, Pérez-Rivero PF, Prieto-Merino D, Cáceres Rivera DI, Guio E, et al. Evaluation of the efficacy and safety of text messages targeting adherence to cardiovascular medications in secondary prevention: the txt2heart Colombia randomised controlled trial protocol. BMJ Open 2019 Dec 8;9(12):e028017. [doi: 10.1136/bmjopen-2018-028017] [Medline: 31818831]
18. Uribe-Rodriguez A, Perez-Rivero P, Free C, Perel P, Murray E, Serrano N. Designing a Text Messaging Program to Increase Adherence to Medication for the Secondary Prevention of Cardiovascular Disease. medRxiv. 2019. URL: https://www.medrxiv.org/content/10.1101/19002683v1.full-text [accessed 2021-02-20]
19. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
20. Chan AH, Horne R, Hankins M, Chisari C. The Medication Adherence Report Scale: a measurement tool for eliciting patients' reports of nonadherence. Br J Clin Pharmacol 2020 Jul;86(7):1281-1288 [FREE Full text] [doi: 10.1111/bcp.14193] [Medline: 31823381]
21. Stirratt MJ, Dunbar-Jacob J, Crane HM, Simoni JM, Czajkowski S, Hilliard ME, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. Transl Behav Med 2015 Dec 9;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
22. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]

23. Zheng X, Spatz ES, Bai X, Huo X, Ding Q, Horak P, et al. Effect of text messaging on risk factor management in patients with coronary heart disease: the chat randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019 Apr;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
24. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55. [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
25. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2008 Apr 16(2):CD000011. [doi: 10.1002/14651858.CD000011.pub3] [Medline: 18425859]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin II receptor blocker
ASCVD: atherosclerotic cardiovascular disease
CONSORT: Consolidated Standards of Reporting Trials
DBP: diastolic blood pressure
HR: heart rate
LDL-C: low-density lipoprotein cholesterol
LMIC: low- to middle-income country
MARS-5: Medication Adherence Report Scale-5
PHQ-9: Patient Health Questionnaire-9
SBP: systolic blood pressure

Edited by L Buis; submitted 06.11.20; peer-reviewed by J Redfern, C Rios-Bedoya, M Choi, C Bedard; comments to author 29.12.20; revised version received 22.02.21; accepted 21.05.21; published 28.07.21.

Please cite as:
Bermon A, Uribe AF, Pérez-Rivero PF, Prieto-Merino D, Saaibi JF, Silva FA, Canon DI, Castillo-Gonzalez KM, Cáceres-Rivera DI, Guio E, Meneses-Castillo KJ, Castillo-Meza A, Atkins L, Horne R, Murray E, Serrano NC, Free C, Casas JP, Perel P
Efficacy and Safety of Text Messages Targeting Adherence to Cardiovascular Medications in Secondary Prevention: TXT2HEART Colombia Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(7):e25548
URL: https://mhealth.jmir.org/2021/7/e25548
doi:10.2196/25548
PMID:34319247

Original Paper

Perspectives and Preferences of Adult Smartphone Users Regarding Nutrition and Diet Apps: Web-Based Survey Study

Maria F Vasiloglou[1], MSc; Stergios Christodoulidis[1,2], PhD; Emilie Reber[3], PhD; Thomai Stathopoulou[1], MSc; Ya Lu[1], PhD; Zeno Stanga[3], PhD; Stavroula Mougiakakou[1,4], PhD

[1]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

[2]MICS Laboratory, Biomathematics Research Group, CentraleSupélec, Gif-sur-Yvette, Paris, France

[3]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[4]Department of Emergency Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Corresponding Author:
Stavroula Mougiakakou, PhD
ARTORG Center for Biomedical Engineering Research
University of Bern
Murtenstrasse 50, 3008
Bern
Switzerland
Phone: 41 316327592
Email: stavroula.mougiakakou@artorg.unibe.ch

Abstract

Background: Digital technologies have evolved dramatically in recent years, finding applications in a variety of aspects of everyday life. Smartphones and mobile apps are being used for a steadily increasing number of tasks, including health monitoring. A large number of nutrition and diet apps are available, and some of them are very popular in terms of user downloads, highlighting a trend toward diet monitoring and assessment.

Objective: We sought to explore the perspectives of end users on the features, current use, and acceptance of nutrition and diet mHealth apps with a survey. We expect that this study can provide user insights to assist researchers and developers in achieving innovative dietary assessments.

Methods: A multidisciplinary team designed and compiled the survey. Before its release, it was pilot-tested by 18 end users. A 19-question survey was finally developed and was translated into six languages: English, German, French, Spanish, Italian, and Greek. The participants were mainly recruited via social media platforms and mailing lists of universities, university hospitals, and patient associations.

Results: A total of 2382 respondents (1891 female, 79.4%; 474 male, 19.9%; and 17 neither, 0.7%) with a mean age of 27.2 years (SD 8.5) completed the survey. Approximately half of the participants (1227/2382, 51.5%) had used a nutrition and diet app. The primary criteria for selecting such an app were ease of use (1570/2382, 65.9%), free cost (1413/2382, 59.3%), and ability to produce automatic readings of caloric content (1231/2382, 51.7%) and macronutrient content (1117/2382, 46.9%) (ie, food type and portion size are estimated by the system without any contribution from the user). An app was less likely to be selected if it incorrectly estimated portion size, calories, or nutrient content (798/2382, 33.5%). Other important limitations included the use of a database that does not include local foods (655/2382, 27.5%) or that may omit major foods (977/2382, 41%).

Conclusions: This comprehensive study in a mostly European population assessed the preferences and perspectives of potential nutrition and diet app users. Understanding user needs will benefit researchers who develop tools for innovative dietary assessment as well as those who assist research on behavioral changes related to nutrition.

(JMIR Mhealth Uhealth 2021;9(7):e27885) doi:10.2196/27885

KEYWORDS

dietary assessment; end-users; mHealth; mobile apps; smartphone; survey; apps; nutrition; diet; mobile health; users; behavior; behavior change

Introduction

To optimally quantify food intake, it is essential to assess nutritional risk, understand dietary patterns, and identify nutrition-related health problems. Accuracy and efficiency in tracking dietary intake requires tools that are validated and easy to use, as conventional methods of assessing diets are prone to errors [1].

Digital technologies, particularly the use of mobile apps, have evolved dramatically in recent years. An estimated 3.7 billion app downloads were performed in 2017. At that time, there were 325,000 mobile health (mHealth) apps available in app stores related to health and fitness [2]. Data collected by smartphones help users improve their self-management and assist behavior changes. Moreover, mobile apps enhance communication between customers and health care professionals, reduce health costs, and improve the dissemination of public health information [3].

The large number of nutrition and diet apps and their numbers of downloads indicate that there is great interest in diet monitoring and assessment [4]. Nutrition and diet apps are among the most widely mHealth used services to promote a healthy lifestyle. Image-based apps that use artificial intelligence (AI) and computer vision are able to recognize the type of food, segment the different parts of the meal, and accurately calculate a meal's energy content, macronutrients, and partial micronutrients [5,6]. A wide variety of factors influence the uptake and impact of mHealth services, which are related both to the individual person (eg, health-related goals) and the technology (eg, usability and accuracy) [6]. To be able to efficiently use those apps, users need to be able to assess them. However, comparing and evaluating these apps is difficult [6].

Nutrition and diet app engagement relies on several elements, such as the time and effort required to manually track food intake [7]. Although some attempts have been made to reduce the effort involved in taking food records (eg, digital scales) [8] or with the list of food items in nutrient databases [9], these features have not yet been integrated into commercially available nutrition and diet apps.

There are few literature reports on the characteristics and preferences of potential nutrition and diet app users. Social norms appear to play a significant role in adopting nutrition and diet apps, as more technically skilled individuals are generally more likely to engage with digital media [10]. Lee et al [11] reported that five factors played a major role in predicting intention to continue using nutrition and diet apps, namely, recordability, networkability, credibility, comprehensibility, and trendiness. A survey of college-aged individuals in the United States investigated their perspectives on health and fitness apps, and it was reported that the respondents valued the low cost and simplicity of the apps as well as the enjoyment of using them [12]. A review on determining the components that facilitate user engagement with digital health interventions to encourage behavior change and weight management showed how crucial it is to incorporate user perspectives from a very early stage of app development to promote app engagement [13].

There are very few broad studies that investigate users' opinions on nutrition and diet apps. To increase acceptance and adoption of those apps, it is necessary to gain insight into user perspectives on nutrition and diet apps while encouraging engagement for continued use. In this study, we aimed to explore the perspectives of users on the features, current use, and acceptance of nutrition and diet apps. To the best of our knowledge, this is the most comprehensive nutrition and diet app user survey, both in sample size and questionnaire detail.

Methods

Survey

We performed a web-based quantitative survey to collect data on nutrition and diet apps and how users perceive them. We adopted the process of drafting, reviewing, and finalizing the questionnaire used in another survey, and the precise procedure can be found in an earlier paper [14]. The following is a brief description of the basic steps taken.

Based on a thorough literature review of surveys on mHealth apps used for dietary monitoring and assessment, we developed 22 preliminary questions. These questions covered basic demographic information, current use of nutrition and diet apps, criteria for nutrition and diet app selection, and barriers to using these apps. Furthermore, respondents were asked to state their opinion on the importance of specific features as well as their preferences for logging meals and how the results were presented. Next, an interdisciplinary team of experts in AI, computer scientists, dietitians, physicians, pharmacists, and psychologists reviewed the questions and made suggestions for revision.

The survey was submitted for a pilot test to determine whether it was simple, clear, concise, and user friendly. The group of 24 users invited to this pilot survey comprised 11 colleagues from the ARTORG Center for Biomedical Research (nonmembers of the research group) and 13 members of the general public. Of these, 75% (18/24) agreed to take part in the survey. All the participants held a BSc degree, their age range was 22-41 years, and they were familiar with app use. They were asked to provide feedback on the structure, content, readability, flow of questions, and duration of the survey. The interdisciplinary team revised the survey based on the feedback received.

The final survey consisted of 19 questions in English. The survey was also translated into German, French, Italian, Spanish, and Greek by certified translators. It was structured in a multiple-choice format, and data were collected and managed using Research Electronic Data Capture (REDCap); participants took 5-10 minutes to complete the survey.

Details regarding the structure of the survey are described in the following sections. Additionally, a comprehensive Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [15] is provided in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Eligible respondents were adults who were able to understand one of the following languages: English, German, French,

Italian, Spanish, or Greek. Signed informed consent was required prior to participation. Exclusion criteria included any inability to understand and comply with written and verbal instructions, or inability to give consent.

Recruitment

The respondents were recruited via (1) social media platforms (ie, Twitter, Facebook), (2) mailing lists of the collaborative universities across Europe, (3) patient associations and foundations, and (4) outpatients of collaborating university hospitals.

Statistical Analysis

Descriptive statistics were used to present quantitative data (only from completed surveys). A chi-square test of independence was performed to examine the differences between groups. Multiple binary logistic regressions were performed to explain relationships between independent variables (eg, age, sex, education, BMI) and categorical dependent variable (ie, users who used a nutrition and diet app). The results were expressed as odds ratios (ORs) with 95% confidence intervals representing the odds that a participant will use a nutrition and diet app based on their aforementioned variables compared to the odds that they will not use it. Statistical significance was indicated with P=.04. RStudio, version 1.0.153 (RStudio PBC) was used for the statistical analysis.

Ethical Approval

The study was reviewed and declared exempt from ethics review by the Cantonal Ethics Committee, Bern, Switzerland (KEK 2019-00102).

Results

A total of 3587 people accessed the survey link, and 2399 completed the survey (66.9%). After 17 respondents from were eliminated from the sample for not providing informed consent, data from 2382 respondents (2382/3587, 66.4%) were included in our analyses. The vast majority of the respondents were from Europe (2333/2382, 97.9%). The demographic characteristics of the respondents are provided in Table 1.

Table 1. Demographic characteristics of the survey respondents (N=2382).

Characteristic	Value
Sex, n (%)	
Female	1891 (79.4)
Male	474 (19.9)
Neither/prefer not to disclose	17 (0.7)
Age (years)	
Average (SD)	27.2 (8.5)
Range, n (%)	
18-29	1770 (74.3)
30-39	413 (17.3)
40-49	111 (4.7)
50-59	64 (2.7)
≥60	24 (1.0)
BMI (kg/m^2), n (%)	
18.5-24.9 (normal weight)	1759 (73.8)
25-29.9 (overweight)	392 (16.5)
<18.5 (underweight)	111 (4.7)
30-34.9 (class I obesity)	82 (3.4)
35-39.9 (class II obesity)	30 (1.3)
>40 (class III obesity)	8 (0.3)
Smoker, n (%)	
No	2014 (84.5)
Yes	349 (14.7)
No answer	19 (0.8)
Highest educational level, n (%)	
Bachelor's degree	958 (40.2)
High school/apprenticeship	744 (31.2)
Master's degree	512 (21.5)
PhD	114 (4.8)
Other	41 (1.7)
Primary/intermediate School	8 (0.3)
No schooling completed	5 (0.2)
Diseases or conditions,[a] n (%)	
No illnesses or health problems	1495 (62.8)
Overweight/obesity	194 (8.1)
Lactose intolerance	138 (5.8)
Inflammatory bowel syndrome	137 (5.8)
Anemia	105 (4.4)
Food allergy	105 (4.4)
Eating disorders	93 (3.9)
Acid reflux	92 (3.9)
High total cholesterol	62 (2.6)

Characteristic	Value
Other diseases/conditions[b]	322 (13.5)
No answer	53 (2.2)
Smartphone user, n (%)	
No	154 (6.5)
Yes	2228 (93.5)
Operating system, n (%)	
Android	1381 (61.9)
iOS (Apple iPhone)	828 (37.2)
Windows	4 (0.2)
I do not know/I do not want to answer	11 (0.5)
Other	4 (0.2)

[a]Multiple answers could be selected.

[b]Any disease/condition with a prevalence of <2.5% was summed in "Other diseases/conditions."

Tracking of Food Intake

It was reported that 29.1% (693/2382) of the respondents tracked their food intake and their tracking methods. Asked about which methods they used, 40.4% (280/693) declared that they used apps, 36.4% (252/693) chose paper and pencil methods, and 22.1% (153/693) preferred other methods. Moreover, 8.2% (57/695) tracked their food by photos/videos, and 4% (28/700) did not express their opinion.

Use of Nutrition and Diet Apps

In total, 51.5% (1227/2382) of the respondents had used a nutrition and diet app. Among these, 20.1% (247/1227) used the app daily, and 8.3% (102/1227) mentioned using the app only for specific foods or beverages. We also asked how the users identified those apps. Most of the respondents reported finding them via the App Store/Google Play store (467/1227, 38.1%), via social media (192/1227, 15.6%), or on a friend's recommendation (156/1227, 12.7%). Only 2% (25/1227) found the apps through their dietitian, 0.7% (8/1227) found them through their medical doctor, and 0.5% (6/1227) expressed no opinion. Finally, 5% (59/1227) mentioned other sources, such as recommendations on the web or from their fitness trainers. The 2 most popular apps were MyFitnessPal (239/1227, 19.5%) and Yazio (184/1227, 15%). Both apps are primarily used as calorie counters.

Reasons for Not Having Used a Nutrition and Diet App

When respondents were asked about the reasons for never having used a nutrition and diet app, the majority (667/1155, 57.8%) indicated that they were not interested, and approximately one-third (340/1155, 29.4%) considered nutrition and diet apps to be too time consuming. Other frequently mentioned reasons were lack of awareness of the apps' existence (194/1155, 16.8%) and privacy as well as security concerns (184/1155, 15.9%). A few respondents preferred paper and pencil methods (44/1155, 3.8%). Free text responses provided additional reasons for nonuse, which included the strong focus of the currently available apps on weight loss and the fear of developing eating disorders. Of those who had not used a nutrition and diet app (1155/2382), 40.3% (465/1155) were optimistic about trying one in the future, 49.6% (573/1155) were negative, and 10.1% (117/1155) did not express their opinion.

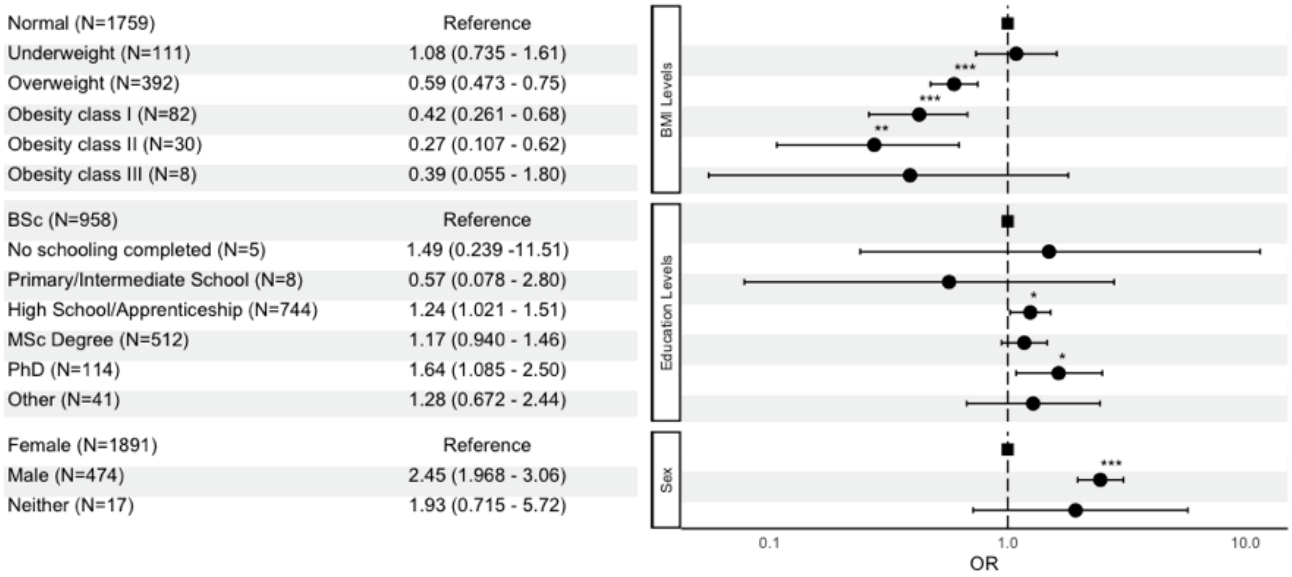
Associations Between Different Variables and Nutrition and Diet App Use

Nutrition and diet app users and nonusers differed significantly by gender (P<.001), BMI (P<.001), and educational level (P=.003). Logistic regression, when adjusted for gender, BMI, and educational level, indicated that the odds that an individual used a nutrition and diet app were 2.5 times greater for male participants compared to female participants (OR 2.45, 95% CI 1.96-3.06; P<.001).

Based on the participants' self-reported weight and height, we described the distribution of nutrition and diet app use according to the World Health Organization classification of BMI. Thus, 54.1% (60/111) of participants in the underweight category, 54.2% (954/1759) of those in the normal weight category, 44.6% (175/392) of those in the overweight category, and 31.7% (38/120) of those in the obese category reported that they had used nutrition and diet apps. According to the logistic regression model mentioned above, overweight reduces the likelihood of using nutrition and diet apps compared to normal weight (OR 0.59, 95% CI 0.47-0.74; P<.001). Individuals with class I or II obesity are less likely to use an app than individuals with normal weight (obese class I: OR 0.42, 95% CI 0.26-0.67, P<.001; obese class II: OR 0.27, 95% CI 0.1-0.62, P=.03).

With regard to the level of education, the logistic regression model showed that having completed only high school or having obtained a PhD increased the probability of using a nutrition and diet app (high school: OR 1.24, 95% CI 1.02-1.51, P=.03; PhD: OR 1.64, 95% CI 1.08-2.49, P=.02). Figure 1 shows the results of the logistic regression for the variables in relation to the use of nutrition and diet apps.

Figure 1. Forest plot showing the influence of BMI, educational levels, and sex on the use of nutrition and diet apps. *P<.05; **P<.01; ***P<.001.



We also investigated the respondents who were living with a disease or health condition and had used a nutrition and diet app. We focused on the most prominent diseases and conditions in our sample (obesity/overweight, lactose intolerance, irritable bowel syndrome [IBS], anemia, food allergy, eating disorders, acid reflux, high cholesterol levels). Only 32.5% (63/194) of people who declared they were living with obesity or overweight were using nutrition and diet apps. The percentage of people who used nutrition and diet apps and were affected by lactose intolerance was 42.8 (59/138), with 31.4% (43/137) for people with IBS, 32.4% (34/105) for people with anemia, 42.9% (45/105) for people with food allergies, 13% (12/93) for people with eating disorders, 38% (35/92) for people with acid reflux, and 47% (29/62) for people with high cholesterol levels. According to the logistic regression model between app users and each disease, individuals with overweight/obesity, IBS, anemia, or eating disorders were less likely to use a nutrition and diet app than individuals who did not have the respective conditions. The ORs for the most prominent diseases are provided in detail in Table 2.

Table 2. Logistic regression of nutrition-related diseases and the use of nutrition and diet apps.

Disease	Odds ratio (95% CI)
Overweight/obesity* (n=194)	0.49 (0.34-0.70)
Lactose intolerance (n=138)	0.90 (0.61-1.32)
Irritable bowel syndrome** (n=137)	0.58 (0.38-0.86)
Anemia** (n=105)	0.49 (0.31-0.77)
Food allergy (n=105)	0.88 (0.56-1.35)
Eating disorders* (n=93)	0.16 (0.08-0.29)
Acid reflux (n=92)	0.82 (0.51-1.30)

*P<.001.
**P<.01.

Criteria for Selecting and Reasons for Not Selecting a Nutrition and Diet App Among the Whole Sample

A detailed overview of the criteria for selecting a nutrition and diet app is provided in Table 3. The most prominent criteria were that the app was easy to use, was free of charge, supported automatic calorie/nutrient estimation, and integrated automatic food recording (eg, bar code readers, meal images).

Asked to indicate what they considered to be barriers to selecting a nutrition and diet app, participants reported that they would not choose such an app if major foods were missing, if it gave incorrect estimations of calories or nutrients, if local foods were not supported, or if the estimation of portion size was not accurate. They also emphasized the vital role of personalization (language, measurement units, etc) in the apps. A detailed list of the barriers and their selection frequencies are provided in Table 3.

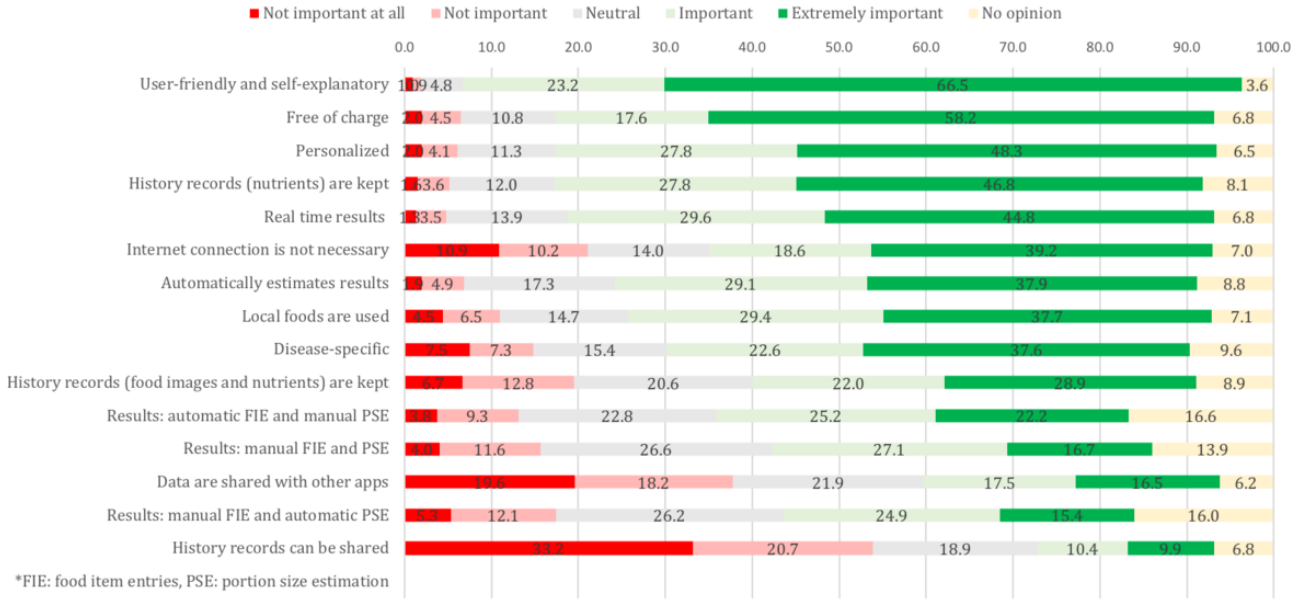
Table 3. Criteria for selecting and reasons for not selecting nutrition and diet apps (N=2382).

Criteria and barriers	Value, n (%)
Criteria for selecting a nutrition and diet app	
Easy to use/convenient	1570 (65.9)
Costless/free of charge	1413 (59.3)
Automatic calorie estimation is supported	1231 (51.7)
Automatic nutrient estimation is supported	1117 (46.9)
Automatic food recording is supported	1115 (46.8)
Validated and certified	1036 (43.5)
Self-explanatory (no need for training)	953 (40)
History records are supported	755 (31.7)
Recipe and menu import functions are supported	729 (30.6)
Provides an option for adding nutritionally related events (eg, reflux)	443 (18.6)
A link for data transmission (eg, other apps) is supported	217 (9.1)
No opinion	195 (8.2)
Other	95 (4)
Barriers to selecting a nutrition and diet app	
Major foods are missing	977 (41.0)
Incorrect calorie and nutrient estimation	798 (33.5)
Local foods are not supported	655 (27.5)
Unconvincing portion size estimation	648 (27.2)
Not personalized	503 (25.1)
Not validated and not certified	567 (23.8)
No opinion	522 (21.9)
Only manual entry of food type is allowed	464 (19.5)
Incorrect automatic food item recognition	448 (18.8)
Recipes are not considered for nutrient estimations	376 (15.8)
Only manual entry of portion size is allowed	345 (14.5)
User is required to be tech savvy	310 (13.0)
History records are not supported	295 (12.4)
Sharing of history records is not supported	88 (3.7)
Other	83 (3.5)

Importance of Specific Features in Nutrition and Diet App Selection

Asked to rate the importance of specific features for a nutrition and diet app on a scale from 1 (not important at all) to 5 (extremely important), users considered user-friendliness, self-explanatory nature of the app, provision of real-time results, and the app being free of charge as important characteristics (Figure 2). Moreover, the automatic estimation and recording of results was judged as important. On the other hand, respondents ranked the ability to share history records and the fact that the data could be shared with other apps as least important.

Figure 2. Importance of features of nutrition and diet apps (%).



Comparing the Criteria for and Barriers to Nutrition and Diet App Use of Users and Nonusers

We also investigated possible differences between the users and nonusers of nutrition and diet apps. In fact, both groups mentioned that their top 3 reasons for selecting a nutrition and diet app were free cost, ease of use, and integration features that automatically estimated energy and nutrient content. The main difference identified was that nutrition and diet app users stated that validation of the apps was among the top 5 important criteria for choosing an app, while nonusers mentioned that automatic food recording would be of great importance if it were supported by nutrition and diet apps.

In terms of barriers, both groups reported that they would not choose an app with incorrect energy and nutrient estimates that did not provide correct portion sizes. Regarding differences, app users viewed not including the most important foods in the app's nutritional database as a major barrier. In contrast, nonusers stated that lack of personalization and lack of inclusion of local foods in the nutrient database would be their greatest barriers.

Preferences Regarding On-screen Display of Nutrient Content and Time Required to Capture Data

When respondents were asked about their preferences regarding on-screen display of the calorie estimations of a meal, 43.3% (1031/2382) indicated that they would prefer to receive the exact value. Regarding the on-screen display of macronutrients, 26.6% (632/2382) would prefer they be presented as a traffic light system (green for low, yellow for moderate, red for high), and 26.5% (631/2382) favored the display of either an accurate value or a combination of a traffic light system and accurate values.

The respondents were also asked about their preferences for the display of the amount of food or drink; they were mostly in favor of values (eg, grams, milliliters) (1284/2382, 53.9%), while a considerable number of respondents (784/2382, 32.9%) preferred common household measures (eg, cups, spoons). Fewer respondents (191/2382, 8%) preferred an abstract portion size (small, medium, large), and 5.2% (124/2382) expressed no opinion.

When asked about the relationship between the response time and the accuracy of the results, respondents were given the following choices: (1) taking 2 photos and obtaining results within 5 seconds that were as accurate as results from a dietitian, (2) taking just 1 photo and obtaining results in real time that were less accurate than those from a dietitian, and (3) recording a video and obtaining results after 5 seconds that were more accurate than those from a dietitian. Most of the respondents preferred the second option (1305/2382, 54.8%), followed by the third (721/2382, 30.3%) and the first (354/2382, 14.9%).

Discussion

Principal Findings

We investigated users' perspectives and preferences regarding nutrition and diet apps. Approximately half of our respondents (1227/2382, 51.5%) reported having used a nutrition and diet app. These results are consistent with data obtained among 1191 respondents in a study investigating weight-management apps in Saudi Arabia, in which 43.1% (513/1190) of the respondents had used weight management apps [16]. A similar cross-sectional survey by Krebs et al [17] among US mobile phone users (n=1604) found that 58.2% (934/1604) of the respondents had downloaded a health-related mobile app.

In our study, the most frequently chosen criteria for selecting a nutrition and diet app were ease of use (1570/2382, 65.9%); lack of cost (1413/2382, 59.3%); automatic energy estimation (1231/2382, 51.7%); automatic nutrient estimation (1117/2382, 46.9%); and automatic food recording (1115/2382, 46.8%). These results reflect those of several studies that investigated the usability of apps as a valuable factor and that concluded that complex and difficult-to-use apps would not be preferred [16-21]. In a study conducted by König et al [7] in 2018 regarding the adoption of nutrition and fitness apps, it was stated that the decision to use an app may be influenced by whether

the data collection is active or passive. For instance, fitness apps that gather data automatically have better use rates than apps with manual data entry [7]. In terms of automatic (ie, AI) features in nutrition apps, however, despite the implied support from the AI technology to yield easy-to-use apps, [4] an analysis of popular nutrition apps (n=13) found no application of AI technologies (eg, image recognition or natural language processing) in those apps.

Issues of App Accuracy and Food Databases as Barriers

One-third of the participants (794/2382) in our study mentioned that incorrect nutrient and energy output would be a barrier to selecting a specific nutrition and diet app. This is also in accordance with observations from another study, which showed that the accuracy and trustworthiness of these apps are important. Respondents expressed concerns stemming either from the inherent characteristics of the apps, such as whether they were developed by experts, or from uncertainty coming from the user (eg, human errors of forgetting) [18]. In fact, in a study that aimed to analyze common user errors when using an image-based app, 12.8% of the acquired images had to be discarded due to mistakes in the capturing procedure [22].

Our study also revealed that respondents would not select apps that had issues related to their food and nutrient databases, such as absence of major or local foods and incorrect estimations of calories or nutrients. Respondents in a qualitative study for weight management (n=24) considered that larger databases were more convenient and easier to use than smaller databases, even though they had difficulty identifying the right foods among the numerous food options [13]. Another database-related barrier reported by some respondents was that of missing foods (ie, ethnic/traditional foods) and misreported content (ie, provision of content of calories but not of macronutrients) in the databases [21]. The same concern was reported by users in another study, in which some respondents expressed doubts about the accuracy of nutrient databases and others expressed concerns about incorrect micronutrient calculations [21]. Moreover, in a qualitative study exploring the experiences of mHealth app use by young adults (n=19), it was mentioned that respondents were doubtful about the reliability of the portion size estimation, which may lead to lower confidence in monitoring intake [23].

User Preferences Concerning Nutrient and Energy Output of Nutrition and Diet Apps

Participants in our study preferred the display of energy and macronutrient content over other options, the former as an accurate value and the latter as an accurate value or traffic light system. In a study of nutrition and diet apps in China, it was found that the most frequently provided output was energy (38/44, 86%); however, none of the detected apps provided any information on macronutrients or salt [24], [23]. Mixed results were found in a study of 24 healthy volunteers, in which color coding was found to be effective by some respondents, while others considered that this approach might promote negative feelings. Some female respondents (n=8) mentioned that they were concerned because the apps might make them compulsive about using them. It was implied that these apps could lead to the development of eating disorders, as they mostly target calories and body weight [21].

In our study, some respondents indicated that current apps focus largely on weight loss rather than on behavioral change, implying a heightened risk that app users could develop an eating disorder. The dynamic nature of weight loss was discussed in [24]. The authors considered that apps can contribute to the development or exacerbation of eating disorders but can also be used to treat these disorders.

Reasons for Not Having Used Nutrition and Diet Apps

The majority of the 48.5% (1155/2382) self-declared nonusers of nutrition and diet apps in our survey said they did not use them due to lack of interest. Approximately one-third of nonusers believed that nutrition and diet apps are time-consuming, and smaller proportions reported lack of knowledge of their existence as well as lack of trust as reasons for nonuse. These findings are in line with a survey conducted among 1604 US citizens, in which 42% of respondents (674/1604) reported never having used a health app. Among these nonusers, the main reasons for nonuse were lack of interest, high app cost, and lack of need [17].

Furthermore, in a qualitative study on health apps conducted among 20 adolescents, the main barriers to using the apps included unfamiliarity with app functionalities and ignorance of their existence. In general, many adolescents did not consider health management as a high priority among their interests [19]. Similarly, another qualitative study [20] examined user perspectives (44/77 participants, 57%, owned health apps) in terms of the content elements of health apps that encourage or hinder their use. Those who did not use apps mentioned impeding factors that included unfamiliarity with health apps (8/33, 25%). Additionally, in corroboration of our findings, some respondents indicated that they had no need for health apps, either because they were using other tools such as websites or because they did not consider themselves in need of such an app because they had already adopted healthy habits. In the same study, a lack of app literacy was noted in that participants did not know which apps were considered good or did not know how to use them [20].

In our study, we found that people with specific nutrition-related diseases, such as obesity, IBS, anemia, or eating disorders, were less likely to use a nutrition and diet app than individuals without the respective conditions. Although we had expected the opposite outcome, we can only speculate that this finding can be attributed to unfamiliarity with the available apps or possible difficulties in finding an app that is dedicated to their condition or is trustworthy. Further research is needed to confirm those findings and explore the aforementioned contradiction.

Where Do Users Hear About or Find Nutrition and Diet Apps?

Slightly over half of the participants in our study reported having used nutrition and diet apps. The majority found them by searching app stores, reading about them on social media, or receiving a recommendation from a friend. Similar findings from another study [17] revealed that respondents learned about the apps from app stores (327/934, 35%) or from friends or

family (287/934, 30.7%), but only 20.37% (210/1031) learned about the apps from a physician's recommendation [17]. Our results match those observed by Peng et al [20], who found that respondents were recommended a health app via a family member or friend; this suggests a substantial social influence. Analogously, the same outcome was found in a web-based survey, in which the participants selected apps based on recommendation from friends or family (154/513, 30%) or from social media influencers (93/513, 18.1%) [16].

Comparison of Criteria and Barriers for App Users in the General Public and Health Care Professionals Who Recommend Nutrition and Diet Apps

A survey of health care professionals (n=1001) [14] showed that they would choose an app if it were easy to use (872/1001, 87.1%) and free of charge (727/1001, 72.6%). However, they also prioritized validation of the app (682/1001, 68.1%) and only then considered the importance of an automatic system for recording food (566/1001, 56.5%), followed by automatic nutritional estimation (525/1001, 52.4%). In terms of barriers, health care professionals agree with users from the general public that food database inaccuracy, missing food items, and lack of personalization are critical issues. However, the fourth criterion for health care professionals was technical knowledge (433/1001, 43.3%), whereas few (310/2382, 13%) end users mentioned that technical knowledge would hinder their use of such apps.

Perception of Obesity

Approximately one-fifth of individuals in our study (476/2382) were classified as either overweight or obese based on BMI calculations of self-reported weight and height. However, when the individuals were asked if they had any health or medical conditions, only 8.1% (194/2382) declared that they were obese or overweight, thus showing a possible weight misperception. Data from the National Health and Nutrition Examination Survey of 4784 individuals living with overweight or obesity showed that 71% (3397/4784) of the participants misperceived their weight. This misperception was associated with lower likelihood of interest in weight loss and less physical activity [25]. Moreover, another study found that people can report their weight and height with reasonable accuracy, but most people with obesity do not consider themselves as obese. It was also noted that adults with obesity who were unable to correctly classify themselves as such may neglect health messages related to obesity and lack motivation to lose weight [26].

Strengths and Limitations

This is the largest study to date that documents the perspectives of European citizens in relation to nutrition and diet apps. These results add to the rapidly expanding field of apps in dietary monitoring and assessment by enhancing the understanding of the needs of users; thus, it creates a clearer idea of their preferences. Another strength is that the survey was made available in 6 different languages, namely, English, Spanish, German, French, Italian, and Greek, which has not been implemented in any other study on this topic. Furthermore, we aimed to minimize bias in terms of the investigators' professions. For this reason, the study was designed by a multidisciplinary team that represented diverse opinions from different scientific fields. Moreover, the distribution of the survey was not restricted to specific user groups but aimed to reach a general population sample. In addition, the distribution of the survey was not limited to social media but was also extended to mailing lists of collaborating universities, patient associations and foundations, and outpatients of collaborating university hospitals.

As a potential limitation, users with greater interest in nutrition apps or those with prior experience using them may have introduced self-selection bias, as they were more likely to participate in the survey. Our sample consisted to a large extent of Swiss and Greek persons; therefore, we cannot generalize our findings to the whole European population. Another drawback is that the survey was limited to people who had internet access. Additionally, in our study, 4.7% (112/2382) of the respondents were underweight, 73.8% (1758/2382) were in the normal weight range, 16.5% (393/2382) were overweight, and 5% (119/2382) had obesity. However, the findings of the current study do not support the European data. More specifically, the average BMI in the European Union population in 2014 was as follows: 2.3% underweight, 46.1% normal weight, 35.7% overweight, and 15.9% obese [27]. Given the convenience sampling that was adopted, the generalizability of these findings may not represent the distribution of nutrition and diet app users.

Future Work

The insights gained from this study may provide a foundation for further studies that will include a broader sample size consisting of larger percentages of participants from different European countries. Larger studies are needed to be able to draw reliable conclusions on the overall opinions of smartphone users on nutrition and diet apps. Interpreting user perspectives on nutrition and diet apps provides important clues for app development and improvement. However, intervention studies are needed to test the usability of the apps as well as whether this theory-based information could lead to persistent nutrition and diet app use.

Future apps that focus on eating disorder recovery should explore the different types of feedback in terms of visualization. This is an important aspect in that colors may considerably impact users' emotional responses, and red and green patterns seem to promote negative behaviors [28]. Weight loss and calorie counting should be treated with caution, because patients may engage in compulsive logging; accurate macronutrient and energy tracking may then encourage unhealthy diets or disordered eating behavior.

Finally, in our study, only 1% (25/2382) of respondents reported that an app was recommended to them by their dietitian. Because nutrition and diet app technology is pervasive, continuous professional development is crucial for health care professionals, especially dietitians, who are responsible for assessing people's nutritional status. Therefore, these professionals should be well informed and keep up to date to be able to suggest reliable apps to their clients and patients. Thereby, the apps would act as invaluable tools for better self-management and dietary

monitoring, while the users would be aware of which apps they should avoid.

Conclusions

This comprehensive study in a mostly European population assessed the preferences and perspectives of potential nutrition and diet app users. The findings suggest that users would select nutrition and diet apps that are easy to use, are free of charge, and automatically estimate the energy and nutrient content of foods with automatic food recording capabilities. Significant barriers to selecting a nutrition and diet app include inaccurate food databases that omit key foods, inaccurate energy and nutrient estimations, and lack of validation and personalization. Understanding user needs will benefit both researchers who develop tools for innovative dietary assessment and those who assist research on behavioral changes related to nutrition. Researchers from different fields, such as nutrition, medicine, computer science, and AI, who are involved in nutrition and diet app development need the insight of the user perspective to design and develop apps that meet users' requirements and needs.

Acknowledgments

We would like to thank all the participants for dedicating their time and sharing their perspectives on apps and the pilot-testers for their valuable input. We also thank Yvonne Bogenstätter for her substantial contribution and the Department of Emergency Medicine, Bern University Hospital, and the University of Bern for their assistance. Special thanks to the University of Bern, Berner Fachhochschule, the Department of Nutrition and Dietetics of the University of Thessaly, Greece, the European Federation of Association of Dietitians, the Swiss Association for Clinical Nutrition, the Hellenic Dietetic Association, the Union of Dietitians Nutritionists Greece, and the Greek Center for Education and Treatment of Eating Disorders for their willingness and kindness to share the word.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOC File , 55 KB - mhealth_v9i7e27885_app1.doc]

References

1. Reber E, Gomes F, Vasiloglou MF, Schuetz P, Stanga Z. Nutritional risk screening and assessment. J Clin Med 2019 Jul 20;8(7):1065-1075 [FREE Full text] [doi: 10.3390/jcm8071065] [Medline: 31330781]
2. Research2Guidance. mHealth App Economics 2017/2018; Current Status and Future Trends in Mobile Health. Research2Guidance. URL: https://research2guidance.com/wp-content/uploads/2017/11/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf [accessed 2020-11-30]
3. Buis L. Implementation: the next giant hurdle to clinical transformation with digital health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
4. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR mHealth uHealth 2016 Aug 01;4(3):85-90 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
5. Vasiloglou MF, Mougiakakou S, Aubry E, Bokelmann A, Fricker R, Gomes F, et al. A comparative study on carbohydrate estimation: GoCARB vs. dietitians. Nutrients 2018 Jun 07;10(6):741 [FREE Full text] [doi: 10.3390/nu10060741] [Medline: 29880772]
6. Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, et al. goFOOD: an artificial intelligence system for dietary assessment. Sensors 2020 Jul 31;20(15):4283 [FREE Full text] [doi: 10.3390/s20154283] [Medline: 32752007]
7. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR mHealth uHealth 2018 Mar 13;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
8. Kumar N, Lopez C, Caldeira C, Pethe S, Si B, Kobsa A. CalNag: effortless multiuser calorie tracking. 2016 Presented at: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops); March 14-18, 2016; Sydney p. 1-4. [doi: 10.1109/PERCOMW.2016.7457051]
9. Andrew A, Borriello G, Fogarty J. Simplifying mobile phone food diaries. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; May 5-8, 2013; Venice, Italy p. 260-263. [doi: 10.4108/icst.pervasivehealth.2013.252101]
10. Okumus B, Bilgihan A, Ozturk AB. Factors affecting the acceptance of smartphone diet applications. J Hosp Mark Manage 2015 Nov 03;25(6):726-747 [FREE Full text] [doi: 10.1080/19368623.2016.1082454]
11. 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]

12. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 Model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
13. Sharpe EE, Karasouli E, Meyer C. Examining factors of engagement with digital interventions for weight management: rapid review. JMIR Res Protoc 2017 Oct 23;6(10):e205 [FREE Full text] [doi: 10.2196/resprot.6059] [Medline: 29061557]
14. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. What Healthcare Professionals Think of "Nutrition & Diet" Apps: An International Survey. Nutrients 2020 Jul 24;12(8) [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
15. 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]
16. Aljuraiban GS. Use of weight-management mobile phone apps in Saudi Arabia: a web-based survey. JMIR mHealth uHealth 2019 Feb 22;7(2):e12692 [FREE Full text] [doi: 10.2196/12692] [Medline: 30794205]
17. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
18. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
19. Chan A, Kow R, Cheng J. Adolescents' perceptions on smartphone applications (Apps) for health management. JournalMTM 2017 Aug;6(2):47-55 [FREE Full text] [doi: 10.7309/jmtm.6.2.6]
20. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
21. Lieffers J, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb 15;118(2):229-239.e3 [FREE Full text] [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
22. Vasiloglou MF, van der Horst K, Stathopoulou T, Jaeggi MP, Tedde GS, Lu Y, et al. The human factor in automated image-based nutrition apps: analysis of common mistakes using the goFOOD lite app. JMIR mHealth uHealth 2021 Jan 13;9(1):e24467 [FREE Full text] [doi: 10.2196/24467] [Medline: 33439139]
23. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Mar;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
24. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China App Store: assessment of functionality and quality. JMIR mHealth uHealth 2019;7(7):e13261 [FREE Full text] [doi: 10.2196/13261] [Medline: 31364606]
25. Duncan DT, Wolin KY, Scharoun-Lee M, Ding EL, Warner ET, Bennett GG. Does perception equal reality? Weight misperception in relation to weight-related attitudes and behaviors among overweight and obese US adults. Int J Behav Nutr Phys Act 2011 Mar 22;8:20 [FREE Full text] [doi: 10.1186/1479-5868-8-20] [Medline: 21426567]
26. Truesdale KP, Stevens J. Do the obese know they are obese? N C Med J 2008;69(3):188-194 [FREE Full text] [Medline: 18751350]
27. European Health Interview Survey: almost 1 adult in 6 in the EU is considered obese. eurostat. 2016 Oct 20. URL: https://ec.europa.eu/eurostat/documents/2995521/7700898/3-20102016-BP-EN.pdf/c26b037b-d5f3-4c05-89c1-00bf0b98d646 [accessed 2021-01-25]
28. Eikey EV, Reddy MC. "It's definitely been a journey": a qualitative study on how women with eating disorders use weight loss apps. 2017 May Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; 6-11 May 2017; Denver Colorado, USA URL: https://doi.org/10.1145/3025453.3025591 [doi: 10.1145/3025453.3025591]

Abbreviations

AI: artificial intelligence
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IBS: irritable bowel syndrome
mHealth: mobile health
OR: odds ratio
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 11.02.21; peer-reviewed by S Kriventsov, P Zhang; comments to author 15.03.21; revised version received 18.03.21; accepted 10.05.21; published 30.07.21.

Please cite as:
Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, Mougiakakou S
Perspectives and Preferences of Adult Smartphone Users Regarding Nutrition and Diet Apps: Web-Based Survey Study
JMIR Mhealth Uhealth 2021;9(7):e27885
URL: https://mhealth.jmir.org/2021/7/e27885
doi:10.2196/27885
PMID:34328425

Original Paper

Examining the Impact of an mHealth Behavior Change Intervention With a Brief In-Person Component for Cancer Survivors With Overweight or Obesity: Randomized Controlled Trial

Jane C Walsh[1*], PhD; Janice Richmond[2], DNSc; Jenny Mc Sharry[1], PhD; AnnMarie Groarke[1], PhD; Liam Glynn[3], MD; Mary Grace Kelly[2], MSc; Owen Harney[1], PhD; Jenny M Groarke[4*], PhD

[1]School of Psychology, National University of Ireland, Galway, Galway, Ireland

[2]Letterkenny University Hospital, Donegal, Ireland

[3]Health Research Institute and Graduate Entry Medical School, University of Limerick, Limerick, Ireland

[4]Centre for Improving Health-Related Quality of Life, School of Psychology, Queen's University Belfast, Belfast, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Jenny M Groarke, PhD
Centre for Improving Health-Related Quality of Life
School of Psychology
Queen's University Belfast
David Keir Building
Belfast, BT7 1NN
United Kingdom
Phone: 44 02890974886
Email: j.groarke@qub.ac.uk

Abstract

Background: Cancer survivorship in Ireland is increasing in both frequency and longevity. However, a significant proportion of cancer survivors do not reach the recommended physical activity levels and have overweight. This has implications for both physical and psychological health, including an increased risk of subsequent and secondary cancers. Mobile health (mHealth) interventions demonstrate potential for positive health behavior change, but there is little evidence for the efficacy of mobile technology in improving health outcomes in cancer survivors with overweight or obesity.

Objective: This study aims to investigate whether a personalized mHealth behavior change intervention improves physical and psychological health outcomes in cancer survivors with overweight or obesity.

Methods: A sample of 123 cancer survivors (BMI≥25 kg/m^2) was randomly assigned to the standard care control (n=61) or intervention (n=62) condition. Group allocation was unblinded. The intervention group attended a 4-hour tailored lifestyle education and information session with physiotherapists, a dietician, and a clinical psychologist to support self-management of health behavior. Over the following 12 weeks, participants engaged in personalized goal setting to incrementally increase physical activity (with feedback and review of goals through SMS text messaging contact with the research team). Direct measures of physical activity were collected using a Fitbit accelerometer. Data on anthropometric, functional exercise capacity, dietary behavior, and psychological measures were collected at face-to-face assessments in a single hospital site at baseline (T0), 12 weeks (T1; intervention end), and 24 weeks (T2; follow-up).

Results: The rate of attrition was 21% (13/61) for the control condition and 14% (9/62) for the intervention condition. Using intent-to-treat analysis, significant reductions in BMI ($F_{2,242}$=4.149; P=.02; ηp^2=0.033) and waist circumference ($F_{2,242}$=3.342; P=.04; ηp^2=0.027) were observed in the intervention group. Over the 24-week study, BMI was reduced by 0.52 in the intervention condition, relative to a nonsignificant reduction of 0.11 in the control arm. Waist circumference was reduced by 3.02 cm in the intervention condition relative to 1.82 cm in the control condition. Physical activity level was significantly higher in the intervention group on 8 of the 12 weeks of the intervention phase and on 5 of the 12 weeks of the follow-up period, accounting for up to 2500 additional steps per day (mean 2032, SD 270).

Conclusions: The results demonstrate that for cancer survivors with a BMI≥25 kg/m^2, lifestyle education and personalized goal setting using mobile technology can yield significant changes in clinically relevant health indicators. Further research is needed

to elucidate the mechanisms of behavior change and explore the capacity for mHealth interventions to improve broader health and well-being outcomes in the growing population of cancer survivors.

Trial Registration: ISRCTN Registry ISRCTN18676721; https://www.isrctn.com/ISRCTN18676721

International Registered Report Identifier (IRRID): RR2-10.2196/13214

(JMIR Mhealth Uhealth 2021;9(7):e24915) doi:10.2196/24915

KEYWORDS

cancer survivors; overweight; obesity; health behavior; goals; accelerometry; text messaging; technology; Ireland; self-management; mobile phone

Introduction

Background

There is an average of 35,000 new cases of cancer diagnosed each year in Ireland, representing a doubling of cases in the past 25 years [1]. At the same time, cancer survivorship in Ireland is also increasing, with survival at 5 years from diagnosis having increased from 42% in 1994 to 62% in 2019, with cancer survivors now making up 4% of the Irish population [1].

There is consistent evidence of a positive association between overweight, obesity, and all-cause morbidity and mortality [2]. A high BMI, poor diet, and lack of physical activity are identifiable risk factors for cancer development, and in cancer survivors, these factors can increase the risk of a secondary cancer or a subsequent primary cancer [3,4]. Cancer and its treatment can result in fatigue, physical inactivity, and loss of muscular strength [5]. Approximately 50% of cancer survivors have overweight or obesity [6], and research has linked obesity to a 46% increased risk of developing distant metastases in women [7]. Considering the consequences of morbidity and mortality, there is a need to facilitate rehabilitation of cancer survivors to reduce BMI, improve diet, and increase physical activity.

Health behavior change interventions can improve physical health outcomes, such as weight and BMI, as well as health behavior (eg, physical activity) and psychological health (eg, quality of life and well-being) in both the general population [8] and among cancer survivors [9,10]. The use of mobile technology (eg, apps and wearables) has been associated with significant reductions in weight and BMI [11] and significant increases in physical activity [12,13]. Mobile health (mHealth) interventions may be able to meet the need for cost-effective health behavior change interventions that can be incorporated into oncology services. Although mHealth interventions hold significant potential, adopting a theory- and evidence-based approach to intervention design is critical [14]. The behavior change wheel is a synthesis of 19 frameworks of behavior change [15]. The behavior change wheel, together with the behavior change technique (BCT) taxonomy, a standardized list of the active ingredients of behavior change interventions [16], enables researchers to develop and describe complex interventions in a systematic and rigorous manner.

Systematic review evidence suggests that the use of relevant BCTs significantly increased the success of weight loss programs [17]. A systematic review of existing healthy eating and physical activity interventions identified the BCTs *self-monitoring* in combination with *goal setting* and *feedback* as the most effective [18]. Furthermore, a meta-analysis of 30 randomized controlled trials (RCTs) that focused on increasing physical activity among cancer survivors reported that the BCTs *prompts*, *social rewards*, and *graded tasks* were associated with larger increases in physical activity in this population [19]. Consequently, these BCTs should be considered for inclusion in interventions aimed at increasing healthy eating and physical activity behaviors among cancer survivors.

Studies have found that both mHealth tools and relevant BCTs can lead to positive health behavior changes and weight loss; therefore, the delivery of BCTs through mHealth tools may be particularly effective. Digital interventions that included a greater number of BCTs were found to have larger effects on health behavior change than interventions with fewer BCTs [20]. A review and meta-analysis of studies using activity monitors found that in people with obesity, physical activity increases were greatest when the BCTs *goal setting* and *feedback* were incorporated in the mHealth intervention [21]. A systematic content analysis of the BCTs provided by wearable activity monitors concluded that most activity monitors included *self-monitoring*, *goal setting*, and *feedback* [22]. Incidentally, a review by Michie et al [18] found these to be the most effective BCTs for promoting healthy diet and physical activity overall.

mHealth interventions incorporating relevant BCTs have the potential to improve health and well-being outcomes. However, there are a limited number of mHealth interventions for cancer survivors that describe content in terms of BCTs. A recent systematic review identified 15 digital health behavior change interventions for cancer survivors, concluding that digital interventions may improve physical activity and reduce BMI; however, findings regarding dietary behavior and well-being outcomes are mixed [23]. Although many of the included studies were pilot or feasibility trials, they highlighted the potential for mHealth interventions to improve health behavior and health outcomes among cancer survivors. All but one study included in this review relied on self-report measures of physical activity [24]. Self-report measures have known limitations, such as recall bias; misinterpretation of items; and overestimation of activity relative to direct measures, such as accelerometer devices [25-28]. Relative to accelerometers, error rates between 35% and 79% have been observed on self-report measures [29,30]. Furthermore, it is noteworthy that only two of the included digital interventions purposively sampled cancer survivors with overweight or obesity [31,32]. This population is arguably most in need of intervention, and health behavior change interventions using nondigital mode of delivery (MOD) can improve outcomes

for this cohort of survivors [33,34]. Overall, there is a need for more large-scale RCTs to provide high-quality evidence regarding the impact of mHealth interventions on objectively measured outcomes in cancer survivors with overweight or obesity.

Aims and Objectives

The aim of this study is to investigate the impact of a personalized mHealth behavior change intervention on physical and psychological health outcomes in a group of cancer survivors with overweight or obesity. More specifically, this project examined the impact of lifestyle education and personalized goal setting, compared with standard medical care, on physical activity (step count) as well as other behavioral, clinical, and psychological outcomes.

Methods

Overview

The full methodological details of the trial, including a detailed description of the development of the intervention, are reported in the study protocol [35] and are summarized below. We used the eHealth extension of the CONSORT (Consolidated Standards of Reporting Trials) statement when writing this paper [36].

Trial Design

A 2-arm, parallel, open-label RCT design was used to investigate the impact of the intervention versus standard care on clinical, psychological, and health behavior outcomes.

Sample Size

The statistical program G*Power was used to conduct power analysis. With 2 groups (intervention and control), 3 measurements (baseline, time 1, and time 2), an assumed correlation among repeated measures of 0.3, a small-medium effect size, and a power of 0.8, the recommended sample size for repeated measures analysis of variance (ANOVA) was 102. A final sample size of 123 was calculated based on an attrition rate of 20%, as observed in similar studies using mobile technology interventions with cancer survivors [37].

Randomization

Participants were randomized to either the intervention or the standard care control condition using a computerized random number generator (enrollment was carried out by MGK and JR, and randomization and group allocation was carried out by JG). The study was not blinded, but step count, one of the main outcome measures, was recorded directly using the Fitbit device (Healthy Metrics Research, Inc).

Study Setting

Recruitment took place offline (by phone), and assessments were carried out face-to-face in a single hospital site, Letterkenny University Hospital, County Donegal, Ireland. Assessments were performed before randomization (T0; baseline), at 12 weeks (T1; intervention end), and at 24 weeks (T2; follow-up).

Ethics Approval

The design of this study was approved by the Research Ethics Committee of the National University of Ireland, Galway, on September 12, 2017 (Ref: 17/MAY/20) and by the Research Ethics Committee at Letterkenny University Hospital on May 2, 2017.

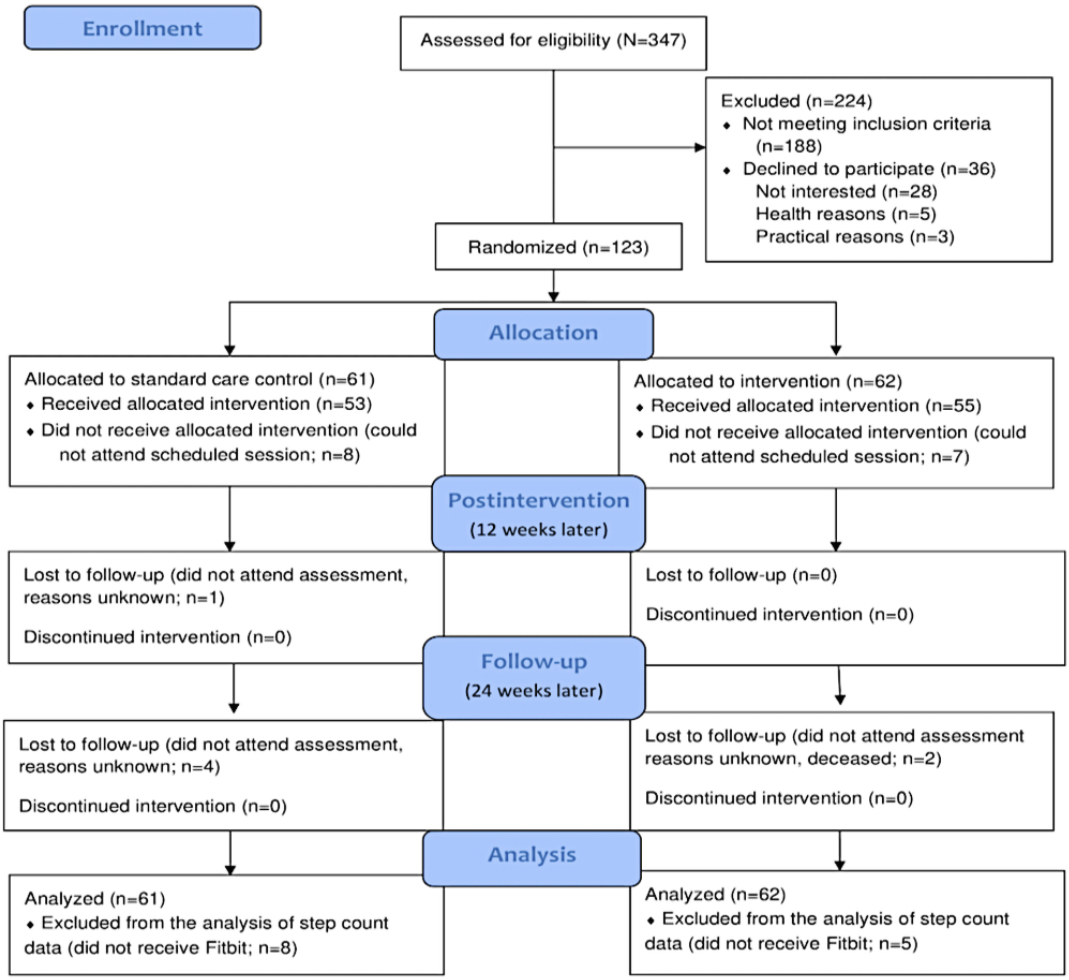
Inclusion Criteria

Adults aged 18-70 years, with a calculated BMI≥25 kg/m^2, with a solid cancer and who had completed active cancer treatment (those continuing on endocrine therapy were permitted inclusion), who attended Oncology Services in Letterkenny University Hospital during the recruitment phase (December 2017 to January 2018), and who were willing to use mobile technology were eligible to participate.

Recruitment

Participants were recruited from the Oncology Services of Letterkenny University Hospital. A total of 159 eligible participants (aged 18-70 years, BMI≥25, and active cancer treatment completed) were identified sequentially from the oncology outpatient waiting list (N=347) by the clinical team. The clinical team contacted these participants by telephone, described the aims and design of the study, and asked if they were willing to use mobile technology. Prospective participants who expressed interest in the study were sent a participant information sheet and consent form (Multimedia Appendix 1). Informed written consent was provided by 77.3% (123/159; response rate) of participants, who then underwent in-person baseline assessments. Of the 36 eligible participants who did not consent to participate (36/159, 22.6%), 28 (78%) were not interested, 3 (8%) were waiting for surgery, 1 (3%) had chronic obstructive pulmonary disease, 1 (3%) was undergoing recurrence workup, 2 (6%) had young children, and 1 (3%) did not drive (Figure 1). A total of 10 eligible participants who were willing to use mobile technology but did not own a smartphone were provided with an Amazon Fire 7 tablet.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants through each stage of the randomized controlled trial.



Intervention

This complex intervention was delivered through mHealth technology and included BCTs that aimed to improve clinical, psychological, and health behavior outcomes. The full details are described in the study protocol [35]. In summary, the intervention consisted of 2 components:

1. A 4-hour lifestyle education and information session (week 1) was delivered by health care professionals (3 physiotherapists, 1 dietician, and 1 clinical psychologist). Physiotherapists demonstrated a series of daily strengthening exercises and recommended schedules for moderate-intensity physical activity. The dietician delivered a comprehensive overview on healthy eating; answered numerous questions that clarified misinformation on nutrition; and specifically advised participants to reduce their caloric intake and reduce the intake of red meat, processed meat, salt, and sugar and increase fruit, vegetable, and fiber intake. The clinical psychologist offered practical strategies for problem solving, identifying barriers to change, and preventing relapse. The BCTs included in this session and the corresponding code from the BCT Taxonomy V1 [16] were *goal setting (outcome) (1.3)*, *provide information on consequences of behavior to the individual (5.1)*, *demonstration of the behavior (6.1)*, *provide instruction on how to perform the behavior (4.1)*, *problem solving (1.2)*, *goal setting (behavior) (1.1)*, and *action planning (1.4)*. These BCTs were applied to both physical activity and dietary behavior changes. The MOD for this component of the intervention was face-to-face human contact in real time with groups of participants. During this session, all participants were provided with a Fitbit Alta.
2. An 8-week goal-setting intervention (weeks 4-12) was delivered using mobile technology (ie, Fitbit Alta accelerometer plus SMS text messaging contact). Participants received weekly text messages with feedback on their average daily step count and a goal of increasing their step count by 10% in the following week. The BCTs included in the personalized goal-setting intervention were *self-monitoring of behavior (2.3)*, *feedback on behavior (2.2)*, *goal setting (behavior) (1.1)*, *graded tasks (8.7)*, *social reward (10.4)*, and *review behavior goal(s) (1.5)*. The MOD for this intervention component was human contact at a distance using nonautomated SMS text messages facilitated using digital wearable technology (ie, Fitbit Alta). Participants continued to wear the Fitbit for the remainder of the study (24 weeks), but the personalized goal-setting intervention ceased at 12 weeks.

Control

Participants randomized to the control condition received standard care and were also provided with a Fitbit Flex 2 to measure physical activity for the 24 weeks of the study. As such, a number of BCTs were also present in the control condition in this study. On being enrolled in the study for meeting eligibility criteria (BMI≥25 kg/m^2), all participants were encouraged to maintain a healthy weight (*goal setting [outcome; 1.3]*). Fitbit accelerometers were distributed at a 15-minute group meeting, where provision of health information was available in the form of leaflets (information on consequences of behavior to the individual [5.1] but not BCT instruction on how to perform the behavior [4.1] or demonstration of the behavior [6.1]). In contrast to the Fitbit Alta distributed to the intervention group, the display panel on the Fitbit Flex 2 does not present summary data (ie, step count), and the app dashboard was modified to not display summary data on the participants' mobile device. The visual display on the device and the app was limited but did not eliminate *self-monitoring of behavior(2.3)* in the control condition. A comparison of the BCTs included in the intervention and control conditions of the study is presented in Textbox 1.

Textbox 1. A comparison of the behavior change techniques included in the intervention and control conditions.

Intervention (Behavior Change Technique, Corresponding Code From the Taxonomy: and Definition):

- Goal setting (outcome; 1.3): "set or agree on a goal defined in terms of a positive outcome of the wanted behavior."
- Provide information on the consequences of behavior to the individual (5.1): "provide information (eg, written, verbal, and visual) about health consequences of performing the behavior."
- Demonstration of the behavior (6.1): "provide an observable sample of the performance of the behavior."
- Provide instruction on how to perform the behavior (4.1): "advise or agree on how to perform the behavior."
- Problem solving (1.2): "analyze, or prompt the person to analyze, factors influencing the behavior and generate or select strategies that include overcoming barriers and/or increasing facilitators."
- Goal setting (behavior; 1.1): "set or agree on a goal defined in terms of the behavior to be achieved."
- Action planning (1.4): "prompt detailed planning of performance of the behavior, must include at least one of context, frequency, duration, and intensity."
- Self-monitoring of behavior (2.3): "establish a method for the person to monitor and record their behavior(s) as part of a behavior change strategy."
- Feedback on behavior (2.2): "monitor and provide informative or evaluative feedback on performance of the behavior and must include one of form, frequency, duration, and intensity."
- Goal setting (behavior; 1.1): "set or agree on a goal defined in terms of the behavior to be achieved."
- Graded tasks (8.7): "set easy-to-perform tasks, making them increasingly difficult, but achievable, until behavior is performed."
- Social reward (10.4): "arrange verbal or nonverbal reward if and only if there has been effort and/or progress in performing the behavior (includes *positive reinforcement*)."
- Review behavior goal(s) (1.5): "review behavior goal(s) jointly with the person and consider modifying the goal(s) or behavior change strategy in light of achievement."

Control (Behavior Change Technique, Corresponding Code From the Taxonomy: and Definition):

- Goal setting (outcome) (1.3): "set or agree on a goal defined in terms of a positive outcome of the wanted behavior."
- Provide information on consequences of behavior to the individual (5.1): "provide information (eg, written, verbal, and visual) about health consequences of performing the behavior."
- Self-monitoring of behavior (2.3): "establish a method for the person to monitor and record their behavior(s) as part of a behavior change strategy."

Materials

All participants were provided with a Fitbit activity tracker for the duration of the study. Each participant was registered with a Fitbit user account. Accounts were set up using a centralized email address corresponding to their study ID number and a randomly generated alphanumeric password. The Fitbit was set up and paired with the participants' mobile devices (ie, smartphone or tablet). The participants were also given an information sheet with instructions on how to synchronize their Fitbit device and app and asked to perform this weekly to prevent loss of data. This sheet also contained the contact details of the research team should they encounter any technical issues or wish to discuss any concerns with their health care providers.

A computer program was developed by the Insight Centre for Data Analytics at the National University of Ireland, Galway, to allow participants' physical activity data to be extracted from the Fitbit server. A member of the research team (JG) logged in to each participant's user account and authorized this third-party program to access their data from Fitbit. The anonymized data for all participants were exported to Excel for analysis.

Fixes

To facilitate the goal-setting intervention, a weighted average for daily step count was calculated for participants with at least five observations per week. Participants who showed no activity for more than 2 days a week were contacted to verify that there

were no technical issues. There were a number of possible reasons for someone to have 0 steps on a given day (eg, the participant did not wear the monitor or the Fitbit failed to record). These reasons were not recorded, and self-reported adherence to monitor wear was not measured. Within 2 weeks of receipt, a number of participants reported challenges using their Fitbit. As a result, all participants were invited to attend 1 of the 2 technical support sessions. A total of 12 participants attended a session and received hands-on support and troubleshooting advice regarding their device from the research team (JG and MGK). Following the implementation of the European Union General Data Protection Regulation (May 2018), participants were automatically logged out of their Fitbit app. However, this was possible to fix at a distance over the phone or via a text message.

Outcomes

Clinical Outcomes

Anthropometric Measurements

Anthropometric measurements included weight in kilograms, BMI, and waist circumference in centimeters.

Functional Exercise Capacity

The 6-minute walk test measures the distance walked in 6 minutes on a hard, flat surface. Systolic blood pressure, diastolic blood pressure, heart rate, blood oxygen saturation, subjective fatigue, and dyspnea were measured pretest (ie, resting), posttest, and 4 minutes later (ie, recovery).

Psychological Outcomes

Health-related quality of life was measured using the Medical Outcomes Survey Short Form (RAND-36) [38]. Other measures of psychological well-being outcomes include the 3-item Loneliness Scale [39], the Brief Fatigue Inventory [40], and the General Self-Efficacy Scale [41]. Exercise self-efficacy [42] and social support for physical activity [43] were also measured.

Health Behavior Outcomes

Self-reported physical activity was measured using the Godin Leisure-Time Exercise Questionnaire [44], and physical activity level (ie, average daily step count) was measured directly using the Fitbit activity tracker. Dietary data were collected using the European Prospective Investigation into Cancer and Nutrition Norfolk Food Frequency Questionnaire [45].

All outcomes were measured at baseline (T0), 12 weeks (T1; intervention end), and 24 weeks (T2; follow-up). The measures are described in full in the trial protocol [35].

Statistical Methods

Missing Data

To maximize power and conform to intent-to-treat analysis, missing data were handled using the expectation-maximization (EM) algorithm. A nonsignificant MCAR test [46] showed that the data were missing completely at random ($\chi^2_{30,080}$=113.3 *P*=.99); therefore, data substitution methods were deemed appropriate. For step count data specifically, EM data substitution was applied only to the 107 participants who received a Fitbit. The 16 participants in the intent-to-treat group (ie, those who could not attend the initial session where Fitbits were distributed) were not included in the missing value analysis and EM data substitution for the analyses of group differences in step count.

Analysis

A series of 3 (time: baseline [T0], 12 weeks [T1], and 24 weeks [T2])×2 (group: control and intervention) mixed ANOVAs were performed to determine the effect of the intervention on clinical, psychological, and health behavior outcomes. In the case of a significant interaction effect, follow-up two-tailed independent sample *t* tests were conducted to investigate between-group differences at each time point, and one-way ANOVAs were conducted to identify within-group differences across time points. Independent samples *t* tests were used to analyze group differences (control and intervention) in average daily step count across the 24 weeks of the study.

Results

Participant Flow

A flow diagram of the progress through each phase of this 2-group parallel randomized trial is shown in Figure 1. A total of 123 eligible participants underwent baseline assessments. The participants were then randomized to the control or intervention arm. Of the 123 participants, 62 (50.4%) were assigned to the intervention group and were invited to attend a lifestyle education and information session where they would also receive their Fitbit activity monitor, and out of these, 55 (89%) participants were able to attend the session. The remaining 61 participants assigned to the control group were invited to an appointment where they were provided with a Fitbit activity monitor, and out of these, 53 (87%) participants were able to attend the session. All participants were invited to a postintervention assessment (12 weeks later) to determine the impact of the lifestyle education and information session and personalized goal-setting mHealth intervention on improving clinical, psychological, and health behavior outcomes. A total of 55 participants in the intervention condition and 52 participants in the control condition attended the assessment. Finally, to determine whether any effects of the intervention were maintained 3 months later, all participants were invited to a follow-up assessment (24 weeks after baseline assessment). In total, 53 participants in the intervention group and 48 participants in the control group attended the follow-up assessment. This resulted in an overall attrition rate of 21% (13/61) in the control arm and 14% (9/62) in the intervention arm.

Baseline Data

Participants' characteristics are described in Table 1. Randomization resulted in an intervention group that was younger, had lower weight and BMI, and had a higher number of males.

Table 1. Participants' characteristics at baseline assessment (N=123).

Characteristics	Control (n=61)	Intervention (n=62)
Age (years), mean (SD)	59.24 (7.65)	55.61 (8.05)
Weight (kg), mean (SD)	87.10 (16.32)	84.18 (13.98)
BMI (kg/m^2), mean (SD)	32.64 (5.41)	30.33 (3.99)
Gender, n		
Female	49	42
Male	4	12
Cancer diagnosis, n (%)		
Breast	49 (80)	50 (81)
Prostate	1 (2)	1 (2)
Lung	0 (0)	1 (2)
Colorectal	9 (15)	8 (13)
Testicular	2 (3)	2 (3)
Have you ever been told by a doctor that you have or have had any of the following conditions? n (%)		
Angina	1 (2)	2 (4)
Heart attack	3 (6)	1 (2)
High blood pressure	19 (36)	18 (33)
Stroke	3 (6)	1 (2)
Diabetes	5 (9)	6 (11)
High cholesterol	21 (40)	20 (37)
Depression	12 (23)	9 (17)
Anxiety	12 (23)	12 (22)

Clinical Outcomes

Anthropometric Measurements

Means and SDs for all anthropometric measurements are presented in Table 2.

Table 2. Anthropometric measurements.

Outcome	Weight, mean (SD)			BMI, mean (SD)			Waist circumference, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2	T0	T1	T2
Control	86.9 (15.2)	85.99 (14.94)	86.16 (14.76)	32.44 (5.07)	32.26 (5.02)	32.33 (5.03)	103.35 (9.28)	101.82 (9.41)	101.53 (9.38)
Intervention	84.17 (13.11)	82.11 (13.41)	82.59 (13.69)	30.47 (3.74)	29.78 (4.04)	29.95 (4.12)	101.15 (11.07)	98.43 (11.72)	98.13 (11.69)

[a]T0: time 0 (baseline).

[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

Weight

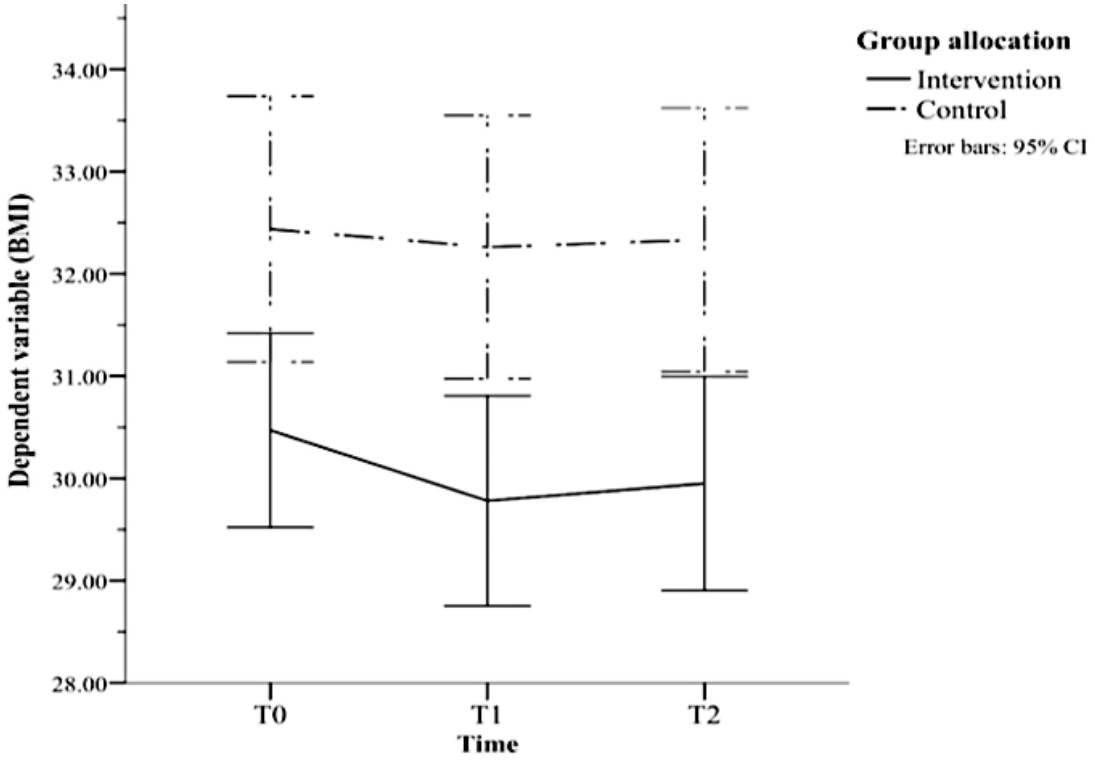
There was no significant interaction effect on weight ($F_{2,242}$=2.615; P=.07). A main effect of time was observed ($F_{2,242}$=18.14; P<.001; ηp^2=0.13). There was no main effect of group ($F_{1,121}$=1.786; P=.18).

BMI

There was a significant interaction between group and time ($F_{2,242}$=4.149; P=.02; ηp^2=0.033) as shown in Figure 2. Follow-up t tests revealed significant group differences in BMI at baseline (t_{121}=2.451; P=.02), at 12 weeks (t_{121}=3.018; P=.003), and at 24 weeks (t_{121}=2.876; P=.005). There was a significant change in BMI across time points in the intervention

group ($F_{2,122}$=12.513; P<.001; ηp^2=0.17). BMI was significantly lower at both 12 weeks (mean difference [MD] −0.689; P<.001) and 24 weeks (MD −0.520; P=.007) than at baseline. In the control group, there was a nonsignificant reduction in BMI ($F_{2,120}$=1.041; P=.36) at 12 weeks (MD −0.18) and 24 weeks (MD 0.11).

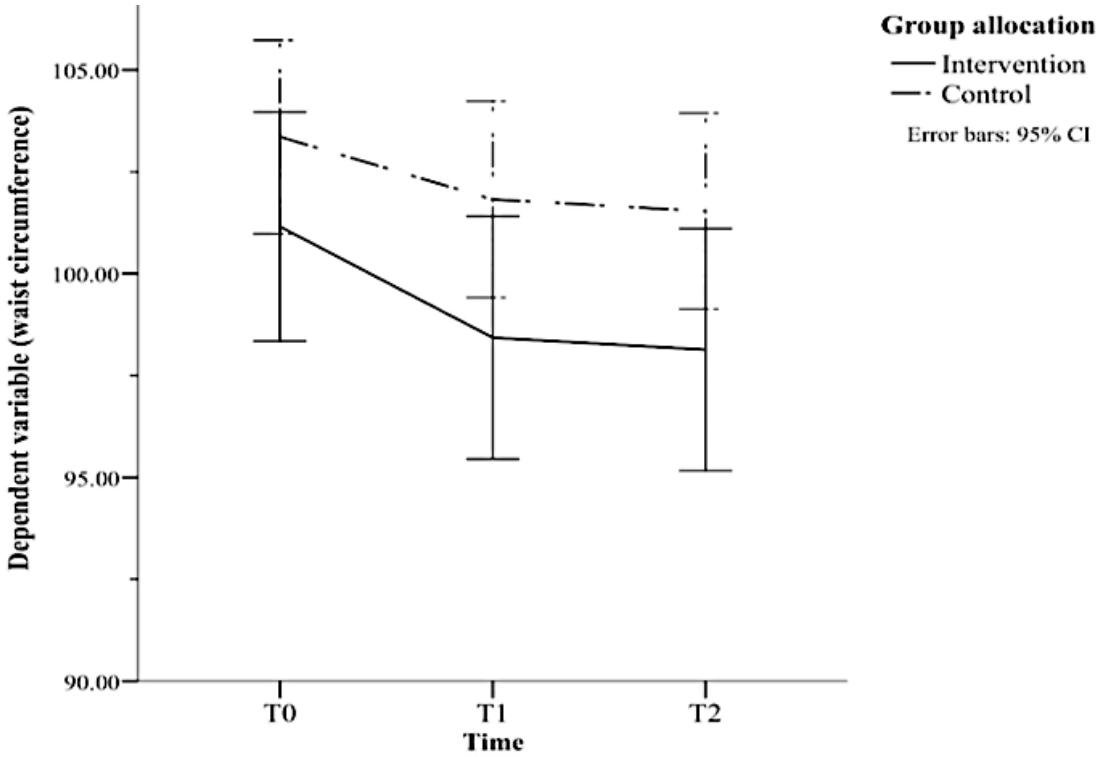
Figure 2. Results of 3×2 mixed analysis of variance showing a significant reduction in BMI for the intervention group only. T0: time 1, baseline; T1: time 1, intervention end (12 weeks); T2: time 2, follow-up (24 weeks).



Waist Circumference

There was a significant interaction effect for waist circumference ($F_{2,242}$=3.342; P=.04; ηp^2=0.027) shown in Figure 3. Post hoc analysis revealed a significant change in waist circumference across time points in both the intervention group ($F_{2,122}$=29.632; P<.001; ηp^2=0.327) and the control group ($F_{2,120}$=20.08; P<.001; ηp^2=0.251). In the intervention group, waist circumference was significantly lower at both 12 weeks (MD −2.725; P<.001) and 24 weeks (MD −3.019; P<.001) than at baseline. This trend was also observed in the control group, with waist circumference significantly lower at both 12 weeks (MD −1.535; P<.001) and at 24 weeks (MD −1.822; P<.001) than at baseline. However, the magnitude of change was greater in the intervention group than in the control group (MD −1.19, SD 0.56; 95% CI −2.31 to −0.06; t_{121}=2.091; P=.04). The difference in waist circumference between 12 and 24 weeks was not significant in the intervention group (MD −0.294; P=.17) or the control group (MD −0.286; P=.21).

Figure 3. Results of 3×2 mixed analysis of variance showing a significant reduction in waist circumference that was maintained at follow-up in both conditions, with a larger reduction in the intervention group. T0: time 1, baseline; T1: time 1, intervention end (12 weeks); T2: time 2, follow-up (24 weeks).



Functional Exercise Capacity

There was no significant interaction effect for distance walked; systolic blood pressure; diastolic blood pressure; heart rate; subjective fatigue; or dyspnea measured before, after, or 3 minutes after the 6-minute walk test (the full set of results are presented in Multimedia Appendix 2). In short, the main effects of time showed that both the groups significantly improved from baseline to T2 in 13 measures of functional exercise capacity (all $P<.001$), and the means and SDs are presented in Table 3.

Table 3. Measures of the 6-minute walk test.

Outcomes	Control, mean (SD)			Intervention, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2
Distance walked	515.99 (67.9)	551.25 (62.44)	566.29 (71.79)	532.57 (69.8)	571.72 (61.68)	590.28 (87.78)
Resting SBP[d]	144.12 (19.57)	135.3 (16.16)	139.45 (13.24)	139.95 (21.04)	129.97 (16.27)	139.99 (33.52)
Posttest SBP	154.48 (20.83)	151.03 (19.18)	150.66 (17.33)	151.67 (23.43)	145.1 (21.07)	152.97 (38.02)
Recovery SBP	142.32 (17.66)	134.31 (17.25)	135.08 (14.74)	138.01 (18.2)	129.38 (15.96)	132.81 (19.19)
Resting DBP[e]	80.89 (7.86)	76.76 (9.2)	80.67 (8.25)	79.56 (9.79)	76.11 (9.12)	77.54 (11.06)
Posttest DBP	82.15 (8.94)	81.78 (12.17)	81.89 (11.66)	81.27 (12.45)	77.39 (10.75)	80.28 (10.83)
Recovery DBP	80.27 (8.06)	77.54 (9)	78.31 (7.45)	78.83 (10.62)	76.35 (9.32)	78.28 (9.74)
Resting HR[f]	79.01 (10.28)	77.36 (8.56)	77.4 (9.27)	78.73 (11.4)	77.56 (15.17)	74.09 (13.30)
Posttest HR	111.67 (17.39)	103.51 (17.94)	113.45 (17.61)	127.06 (143.76)	105.89 (24.44)	113.53 (50.21)
Recovery HR	85.49 (11.88)	82.07 (10.77)	83.5 (10.75)	85.13 (13.32)	82.31 (12.45)	83.02 (15.57)
Resting fatigue	6.39 (0.9)	6.22 (0.91)	5.96 (1.06)	6.35 (0.79)	6.17 (1.03)	6.05 (1.39)
Posttest fatigue	9.83 (2.36)	10.73 (1.98)	10.35 (2.72)	9.9 (2.41)	10.54 (2.45)	10.2 (3.19)
Recovery fatigue	6.93 (1.6)	6.4 (0.98)	6.09 (1.42)	6.88 (1.23)	6.56 (1.38)	6.08 (2.71)
Resting dyspnea	1.24 (0.73)	1.12 (0.42)	1.58 (1.19)	1.19 (0.65)	1.15 (0.67)	1.35 (0.97)
Posttest dyspnea	4.02 (1.67)	3.74 (1.47)	5.06 (4.61)	3.5 (1.33)	4.04 (1.75)	4.25 (2.4)
Recovery dyspnea	1.81 (0.97)	1.24 (0.47)	1.88 (1.88)	1.64 (0.94)	1.40 (0.87)	1.45 (1.51)

[a]T0: time 0 (baseline).

[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]HR: heart rate.

Psychological Outcomes

Quality of Life

No significant interaction effects were observed for health-related quality of life (measured by the RAND36 Medical Outcomes Survey). Means and SDs are presented in Table 4, and the results of the ANOVA are presented in Multimedia Appendix 3.

Table 4. Subscales of RAND-36 Medical Outcomes Survey.

Outcome	Control, mean (SD)			Intervention, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2
Physical functioning	77.87 (17.06)	81.81 (14.52)	81.15 (17.28)	73.87 (20.33)	80.08 (17.28)	77.66 (19.83)
Role limitations: physical health	69.67 (39.82)	78.28 (29.39)	77.05 (35.15)	69.76 (37.64)	85.89 (30.57)	79.03 (33.63)
Role limitations: emotional health	83.06 (33.12)	89.62 (23.99)	83.06 (32.56)	72.58 (40.72)	87.37 (29.37)	83.33 (31.22)
Pain	73.24 (21.85)	78.11 (21.65)	74.22 (23.63)	73.47 (23.16)	79.28 (19.62)	75.57 (21.52)
Emotional well-being	77.38 (17.57)	81.77 (13.59)	83.74 (13.28)	73.03 (18.66)	79.87 (16.32)	78.77 (18.04)
Social functioning	78.48 (23.4)	91.8 (15.79)	88.31 (18.8)	78.63 (23.01)	89.31 (17.8)	85.88 (20.69)
Energy	61.39 (20.21)	70.99 (16.38)	68.69 (19.68)	53.39 (21.27)	67.02 (17.15)	63.06 (18.07)
General health	51.73 (20.17)	57.6 (15.87)	55.87 (16.4)	49.5 (16.54)	56.98 (12.94)	53.04 (15.74)

[a]T0: time 0 (baseline).

[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

Other Psychological Outcomes

There were also no significant interaction effects for loneliness, self-efficacy, exercise self-efficacy, or exercise social support (full results, including nonsignificant findings, are presented in Multimedia Appendix 4). The means and SDs are presented in Table 5.

Table 5. Psychological outcome measures.

Outcomes	Control, mean (SD)			Intervention, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2
Loneliness	4.23 (1.64)	4.1 (1.57)	4.34 (1.63)	3.63 (1.02)	3.69 (0.9)	3.74 (1.01)
Fatigue (global)	35.47 (20.47)	25.18 (20.79)	31.05 (20.3)	23.36 (19.42)	20 (16.45)	21.49 (15.95)
Fatigue severity	9.77 (4.92)	8.05 (4.88)	8.9 (4.59)	6.92 (3.89)	6.38 (3.55)	6.90 (3.47)
Fatigue interference	20.63 (14.43)	13.28 (14.64)	17.76 (14.21)	13.1 (13.47)	10.53 (11.33)	11.48 (10.99)
Self-efficacy	20.56 (4.45)	22.27 (4.26)	21.96 (4.62)	21.69 (4.19)	22.16 (3.92)	22.01 (4.02)
Exercise: self-efficacy	22.33 (4.14)	21.93 (4.56)	21.08 (4.98)	23.14 (2.49)	22.69 (3.85)	22.41 (3.63)
Exercise: social support	12.36 (5.39)	13.39 (4.39)	12.45 (4.8)	12.33 (4.92)	13.03 (4.31)	12.1 (5.01)

[a]T0: time 0 (baseline).

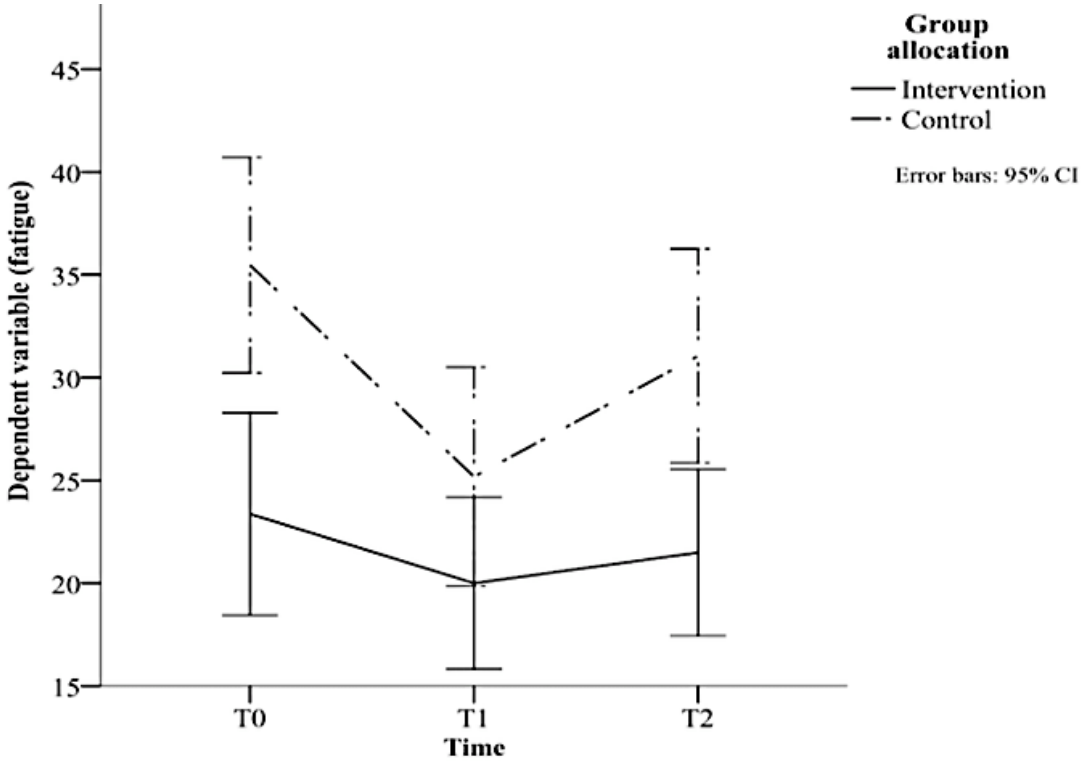
[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

Fatigue

As shown in Figure 4, there was a significant interaction between groups and time ($F_{2,242}$=3.199; P=.04; ηp^2=0.026). Independent samples t tests revealed significant differences between the intervention and control groups at baseline (t_{121}=3.365; P=.001) and at 24 weeks (t_{121}=2.908; P=.004) but not at 12 weeks (t_{121}=1.534; P=.13). Fatigue remained stable in the intervention group ($F_{2,122}$=1.815; P=.17). The change in global fatigue in the control group was significant ($F_{2,120}$=11.701; P<.001; ηp^2=0.163). Fatigue was significantly lower at 12 weeks than at baseline (MD −10.289; P<.001) and was significantly higher at 24 weeks than at 12 weeks (MD 5.872; P=.001). The difference in fatigue between baseline and 24 weeks was not significant (MD −4.417; P=.052).

Figure 4. Results of a 3×2 mixed analysis of variance showing a significant reduction in fatigue at 12 weeks and a nonsignificant increase at 24 weeks in the control group only. T0: time 1, baseline; T1: time 1, intervention end (12 weeks); T2: time 2, follow-up (24 weeks).



Health Behavior Outcomes

Dietary Behavior

Dietary data were collected using the European Prospective Investigation into Cancer and Nutrition Norfolk Food Frequency Questionnaire 45. There were no significant interaction effects for any of the 10 food groups assessed (Multimedia Appendix 5), and the means and SDs are presented in Table 6.

Table 6. Dietary behavior.

Outcomes	Control, mean (SD)			Intervention, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2
Fiber	16.25 (4.06)	16.33 (4.78)	16.64 (4.14)	17.63 (6.62)	17.45 (5.32)	18.6 (6.18)
Kilocalorie	1846.14 (538.65)	1681.15 (562.66)	1687.54 (455.96)	2030.77 (664.09)	1760.36 (569.79)	1828.44 (687.56)
Sodium	2940.68 (891.54)	2547.2 (785.47)	2685.03 (759.71)	3131.63 (1021.24)	2586.32 (834.46)	2840.5 (1075.67)
Saturated fats	29.06 (13.26)	24.91 (10.35)	26.30 (10.29)	31.95 (14.17)	24.80 (13.1)	27.52 (15.98)
Fruit	243.29 (138.28)	291.02 (153.8)	269.74 (160.84)	259.69 (179.98)	310.94 (214.98)	312.33 (329.19)
Meat	126.40 (54.37)	103.33 (48.28)	119.18 (49.56)	125.58 (62.87)	106.49 (78.39)	120.22 (59.21)
Sugar	48.65 (45.42)	31.84 (28.06)	35.55 (28.4)	55.51 (57.16)	31.09 (22.25)	36.89 (36.42)
Vegetables	231.91 (101.65)	244.62 (95.77)	263.57 (105.37)	248.9 (127.94)	264.88 (106.22)	303 (130.75)
Alcohol	2.02 (2.67)	1.75 (2.46)	2.15 (2.6)	3.62 (9.97)	2.80 (7.23)	3.28 (7.95)
Alcoholic beverages	22.53 (27.63)	20.67 (28.95)	27.79 (39.86)	44.97 (132.53)	31.81 (91.21)	40.67 (108.43)

[a]T0: time 0 (baseline).

[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

Self-reported Physical Activity Level

There was no significant main effect of time ($F_{2,242}$=1.56; P=.21). There was no main effect of condition ($F_{1,121}$=0.073; P=.79) and no significant interaction effect ($F_{2,242}$=0.260; P=.77). The means and SDs are presented in Table 7.

Table 7. Scores on the Godin Leisure-Time Exercise Questionnaire.

Outcome	Control, mean (SD)			Intervention, mean (SD)		
	T0[a]	T1[b]	T2[c]	T0	T1	T2
Weekly leisure activity	31.03 (17.25)	34.68 (26.12)	33.03 (29.39)	31.14 (20.52)	34.75 (18.34)	30.38 (16.71)

[a]T0: time 0 (baseline).

[b]T1: time 1 (intervention end; 12 weeks).

[c]T2: time 2 (follow-up; 24 weeks).

Direct Physical Activity Level

The step count data were collected continuously using Fitbit. Daily step count totals were summed, and an average daily step count was calculated for each week of the 24-week study. Means and SDs are presented in Table 8, along with the results of independent *t* tests comparing group differences in step count. The intervention group had a significantly higher average daily step count on 13 of the 24 weeks of the study (ie, weeks 3, 5-9, 11, 12, 14-17, and 21), contributing to an additional 1689-2500 steps per day (mean 2032, SD 270).

An analysis of the personalized goal-setting intervention demonstrated that 69% (37/54) of participants in the intervention group met at least 50% (4/8) of their step count goals. However, the goal success rate was not significantly correlated with any of the study outcome variables. A further analysis of prescribed step count goals within the context of goal achievement indicates that success was highest in the earlier stages of the goal-setting intervention when step count goals were below 10,000 steps (Table 9).

Table 8. Average daily step count for weeks 1-24 (N=107).

Group	Control, mean (SD)	Intervention, mean (SD)	*t* test (*df*)	*P* value
Week 1	8792.11 (3990.17)	8775.34 (8946.51)	0.012 (105)	.99
Week 2	8434.01 (4135.28)	8978.65 (6063.41)	−0.542 (105)	.59
Week 3	8622.72 (3126.16)	10,401.04 (3659.81)	−2.7 (105)	.01
Week 4 (SMS 1)	8638.9 (3768.52)	9833.51 (3737.18)	−1.646 (105)	.10
Week 5 (SMS 2)	8359.62 (3414.64)	10,621.07 (4020.28)	−3.133 (105)	<.001
Week 6 (SMS 3)	8308.83 (3523.38)	10,265.19 (4345.32)	−2.555 (105)	.01
Week 7 (SMS 4)	8519.34 (3318.01)	10,982.03 (4708.59)	−3.122 (105)	<.001
Week 8 (SMS 5)	9940.19 (4317.15)	12,247.11 (5842.16)	−2.319 (105)	.02
Week 9 (SMS 6)	8819.53 (4088.03)	10,744.1 (4681.16)	−2.263 (105)	.03
Week 10 (SMS 7)	9065.88 (3755.52)	10,262.84 (4780.74)	−1.438 (105)	.15
Week 11 (SMS 8)	8576.29 (3865.43)	10,610.33 (5441.93)	−2.225 (105)	.03
Week 12	9067.76 (3838.9)	10,915.87 (4804.31)	−2.196 (105)	.03
Week 13	9883.54 (3462.85)	11,080.53 (5621.11)	−2.113 (105)	.19
Week 14	9219.04 (3711.7)	10,908.43 (4703.33)	−2.06 (105)	.04
Week 15	9185.22 (3715.93)	11,260.69 (4871.56)	−2.474 (105)	.01
Week 16	7750.5 (3536.26)	9553.78 (5119.87)	−2.116 (105)	.04
Week 17	8726.35 (3938.05)	10,513.67 (4666.4)	−2.139 (105)	.04
Week 18	9020.32 (3487.54)	10,255.48 (3446.08)	−1.843 (105)	.07
Week 19	9480.41 (3820.46)	10,901.42 (4965.92)	−1.657 (105)	.10
Week 20	8624.73 (3426.48)	10,108.47 (4817.34)	−1.833 (105)	.07
Week 21	7700.75 (4584.27)	10,201.34 (4386.2)	−2.883 (105)	<.001
Week 22	7520.26 (11826.73)	10,171.88 (7265.27)	−1.4 (105)	.16
Week 23	8224.2 (5145.45)	9481.54 (5442.37)	−1.228 (105)	.22
Week 24	8483.63 (4808.37)	9700.24 (5860.72)	−1.173 (105)	.24

Table 9. Average daily step count goal for each week and participants' rate of success in achieving step count goals in weeks 5-12 (n=54).

Time	Goal 1: week 5	Goal 2: week 6	Goal 3: week 7	Goal 4: week 8	Goal 5: week 9	Goal 6: week 10	Goal 7: week 11	Goal 8: week 12
Step count goal, mean (SD)	7541.75 (4046.05)	7907.88 (5186.28)	10,601.11 (4874.16)	11,598.83 (5337.32)	9396.26 (5745.74)	10,922.66 (6074.49)	11,145.08 (5479.19)	11,294.56 (6039.08)
Did achieve goal, n (%)	41 (76)	39 (72)	25 (46)	35 (65)	35 (65)	18 (33)	17 (32)	24 (44)
Did not achieve goal, n (%)	13 (24)	15 (28)	29 (54)	19 (35)	19 (35)	36 (67)	37 (68)	30 (56)

Discussion

Principal Findings

The aim of this trial is to examine the impact of a personalized mHealth behavior change intervention on clinical, psychological, and health behavior outcomes among a group of cancer survivors with overweight or obesity. The results show that the intervention yielded several significant benefits over and above that shown in the standard care control group. The intervention group had a significantly greater reduction in BMI than the control group. This reduction in BMI was maintained at the 24-week follow-up. Relative to the control group, there was a significantly greater reduction in waist circumference in the intervention group. At follow-up, there was a modest reduction in BMI (0.52) and waist circumference (3.02 cm) with small to medium effect sizes. In relation to behavioral outcomes, participants in the intervention group had significantly higher physical activity during both the intervention phase (8 out of the 12 weeks) and the follow-up phase (5 out of the 12 weeks) than those in the control group. Participants in the intervention averaged approximately 2000 extra steps per day (the equivalent of 1 mile or 20 minutes of physical activity [47]). However, there were no significant changes in functional exercise capacity, dietary behavior, or psychological outcomes.

The design of this intervention is aligned with the National Institute for Health and Care Excellence guidelines for weight

management in people with obesity [48]. They recommend multicomponent interventions that include effective behavior change strategies aimed at increasing physical activity or decreasing sedentary behavior, reducing energy intake, and improving diet quality. Although the statistically significant changes in anthropometric measures observed in this study fall short of the 5% change deemed clinically significant, the clinical guidelines for obesity management acknowledge that more modest losses can also yield significant health benefits [48]. The magnitude of change observed here is in line with the results of traditional behavioral interventions for people with overweight or obesity [49] and cancer [50]. They are also consistent with interventions using digital MOD for people with chronic conditions, such as obesity [51] and cancer [23]. Cancer survivorship is characterized by ongoing physical and psychological challenges, making behavior change and weight management more difficult [52], and any change observed is of paramount clinical importance for this population. Furthermore, the results of a mixed methods investigation nested within this RCT and published separately concluded that this intervention is acceptable to participants [53], in addition to being effective.

A key element of this intervention was the use of personalized goals delivered through mobile technology (ie, Fitbit and SMS text messaging contact). This aimed to enhance participants' motivation to increase their overall levels of physical activity. In addition to significant increases in physical activity (step count), the results show a good level of goal achievement, suggesting that goal-setting intervention did influence participants' motivation to increase activity. Although goals were personalized (ie, increase daily step count by 10% per week), goal attainment was highest in the initial stages when step count targets were lower, suggesting there may be a threshold after which increasing step count become unattainable. This is not surprising, and recent evidence suggests that significant health benefits are achievable at much lower levels of physical activity (ie, 4400-7500 steps per day) [54]. This is also in line with existing studies, which found that setting a higher physical activity goal leads to higher physical activity levels but concurrently lower goal achievement [55]. This suggests that goal setting may be more effective in the early stages of the physical activity intervention, but it becomes less critical to attaining a higher physical activity level once a certain threshold is consistently achieved or perhaps when health behaviors have been consolidated. If one assumes that a low level of goal achievement is deleterious over time (eg, decrease in motivation) [56], the results do not rule out the existence of an *optimal goal-setting zone* for some participants. This was the approach taken in this study, where targets were based on current performance. Swann et al [56] argued that achievement of goals is actually not the primary aim of goal-setting theory; instead, goal setting is simply a mechanism for enhancing task performance regardless of whether the goal is achieved. Within this context, a higher step count is a positive outcome, despite failure to achieve step count targets at the higher end of the spectrum. It is likely that this component of the intervention, including the highly effective BCTs of *goal setting*, *feedback*, *review*, and *self-monitoring* [17-19] delivered via mobile technology, facilitated the higher levels of physical activity in the intervention participants, thus contributing to the significant reductions in BMI and waist circumference.

Although the results did not demonstrate any significant improvements in dietary behavior, it is clear that changes in lifestyle (primarily increased physical activity) contributed to significantly greater benefits in key clinical outcomes for the intervention group. An emerging body of research with cancer survivors suggests that digital interventions have positive effects on BMI and physical activity, but the findings are less consistent for diet [23]. For example, the digital health behavior change intervention by O'Carroll Bantum et al [57] also found increased physical activity but no increase in fruit or vegetable intake among cancer survivors. As such, the results of this study are in line with those of previous studies. On reflection it is not wholly unexpected that no significant dietary behavior changes were found. Participants in the intervention group attended a 4-hour lifestyle education and information session that included BCTs related to dietary behavior change delivered by a dietician, in addition to a goal-setting intervention that focused exclusively on physical activity. This may not have been sufficient to change dietary behavior. In postintervention interviews, participants indicated that additional behavioral support was needed to change their diet [53]. The success of the goal-setting intervention for increasing step count in this study is encouraging, and future digital interventions should consider goal setting in relation to dietary behavior in combination with physical activity goals.

This intervention had 2 behavioral targets (increase physical activity and improve diet) to improve health and well-being outcomes. Systematic review evidence has found that health behavior change interventions focusing on physical outcomes improve well-being in the general population [8] and among cancer survivors [9,10]. Therefore, it was unexpected that despite increased physical activity levels, the intervention group did not report significant improvements in any of the psychological well-being outcomes over the course of the study. These findings are consistent with other lifestyle interventions for cancer survivors that also found no significant group differences in quality of life [32,58,59], well-being [23], or fatigue [20,60]. One possibility for the lack of significant effects is the relatively high levels of well-being in participants at baseline. Furthermore, although participants received a comprehensive presentation from a clinical psychologist at the lifestyle education and information session, this presentation focused on behavior (eg, action planning and problem solving); thus, there was no aspect of the intervention that deliberately targeted increased well-being. Future interventions may wish to incorporate techniques and strategies aimed at improving well-being more directly. Nevertheless, qualitative data show that participants perceived the aim of the intervention to be *moving on* psychologically from cancer and reported emotional and psychological improvements as a result of participating [53]. These self-reported improvements did not translate into statistically significant interaction effects in this trial.

Strengths and Limitations

Although the effect sizes were small to medium, the large sample size and high retention rate means that the study was

adequately powered to detect such effects. Participants were randomized to conditions to reduce selection bias, and the use of intention-to-treat analysis limited the impact of attrition bias. However, it was not possible to blind participants or outcome assessors in this study, a limitation common to many digital health interventions [36]. This may have introduced a performance or detection bias. Furthermore, the analyses were not adjusted for baseline differences that occurred due to chance. Unadjusted analyses from randomized trials provided valid estimates of effects [61], and baseline levels of variables that are likely to be prognostic (ie, anthropometric measures) were accounted for in the analyses. Nevertheless, these results should be interpreted with caution. The participants were mostly female breast cancer survivors. This is representative of trends in the wider literature [62] but may limit the generalizability of the findings to people with other types of cancer and to people of other genders in the wider population. Furthermore, inclusion criteria required participants to be willing to use mobile technology. This may have contributed to a digital divide, in the sense that prospective participants who may have benefited from the intervention were excluded because they were inexperienced or uncomfortable using digital technologies. It is worth noting that no one was excluded due to a lack of mobile technology; participants who did not have access to but were willing to engage with mobile technology were provided with a mobile device (Amazon Fire tablet) by the research team. Finally, the participants in the trial had access to technical support from the research team, as needed. This was front loaded, as support was needed more frequently at commencement when participants were becoming familiar with the technology. This may not be feasible in standard oncology care, limiting the applicability of the findings outside of an RCT setting. A custom-built front-end software was used to bulk export participants' step count data to facilitate the weekly goal-setting intervention; as such, something similar may be needed if applying the same intervention in a health care setting [63]. Future research will be needed to identify potential implementation issues for delivering this intervention in clinical settings.

A notable strength of this study is the use of a direct measure of physical activity (ie, Fitbit accelerometer). In a review of 15 digital health behavior change interventions with cancer survivors [23], no study had used a direct measure of physical activity to evaluate the impact of the intervention. The majority of studies rely on self-report measures that are biased in a number of ways and provide less accurate estimates than accelerometers [25-28,30]. Consumer-based wearable accelerometers offer similar accuracy to criterion measures in controlled settings [64,65] and in free-living settings [66,67]. However, relative to research-grade accelerometers, some studies have reported underestimation [68] and overestimation of steps [69], especially at faster ambulatory speeds. In addition to the high cost of these devices (eg, ActivPAL and ActiGraph), there are a number of practical constraints to their use in studies with large samples and over longer time frames [70,71]. To our knowledge, no study has examined the potential differences in precision between the 2 Fitbit models used in this study. Reassuringly, the Fitbit has been consistently rated among the most accurate consumer-based wearable activity monitor for measuring step count [64,68,72]. According to the authors' knowledge, this trial is the first to evaluate a digital health behavior change intervention using a direct measure of physical activity with a sample of cancer survivors with overweight or obesity. That being said, self-report measures of dietary behavior were used in this study, although interviews by health care professionals (ie, 24-hour dietary recall) would have been superior, particularly for accurately measuring caloric intake.

Conclusions

Cancer survivors who have overweight or obesity require additional support to self-manage their health behaviors. mHealth technology may provide a cost-effective solution within modern oncology care. mHealth has enormous potential for improved health care delivery, but evidence from this group currently lacks a strong base [58,62]. The results of this study represent a promising contribution to the field. This mHealth intervention significantly reduced BMI and waist circumference and increased physical activity levels, but it was consistent with an emerging body of research with cancer survivors [23], which demonstrates limited impact on diet or well-being. Future research is needed to continue evaluating and refining mHealth behavior change interventions to improve health and well-being outcomes for the growing number of cancer survivors.

Acknowledgments

The authors would like to acknowledge the support of the nursing staff Tereze Toby, Mary McCollum, Noreen Rodgers, and Caroline Nee; physiotherapists Tommy Kerr, Aoife O'Donnell, and Eimear Masterson; dietician Nina Singaroyan at Letterkenny University Hospital; clinical psychologist Charlene Haughey at Cancer Care West and Letterkenny University Hospital; the Irish Cancer Society Daffodil Centre at Letterkenny University Hospital; the Insight Center for Data Analytics at the National University of Ireland, Galway; and all the participants for their time. This research was funded by a grant awarded to JR and JCW by the Irish Cancer Society with support from Relay for Life Donegal.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant information sheet and consent form.
[DOC File , 410 KB - mhealth_v9i7e24915_app1.doc]

Multimedia Appendix 2
Results of 3×2 analysis of variance on measures of the 6-minute walk test.
[DOCX File , 16 KB - mhealth_v9i7e24915_app2.docx]

Multimedia Appendix 3
Results of 3×2 analysis of variance on subscales of the RAND-36 Medical Outcomes Survey: Short Form.
[DOCX File , 14 KB - mhealth_v9i7e24915_app3.docx]

Multimedia Appendix 4
Results of 3×2 analysis of variance on other psychological outcomes.
[DOCX File , 14 KB - mhealth_v9i7e24915_app4.docx]

Multimedia Appendix 5
Results of 3×2 analysis of variance on dietary behavior.
[DOCX File , 15 KB - mhealth_v9i7e24915_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1206 KB - mhealth_v9i7e24915_app6.pdf]

References

1. Cancer in Ireland 1994-2017 with Estimates for 2017-2019. Annual Report of the National Cancer Registry. 2019. URL: https://tinyurl.com/2pe8vfxx [accessed 2021-06-16]
2. Flegal KM. Body-mass index and all-cause mortality. Lancet 2017 Jun;389(10086):2284-2285. [doi: 10.1016/s0140-6736(17)31437-x]
3. Stewart BW, Wild CP. World Cancer Report. International Agency for Research on Cancer. 2014. URL: https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014 [accessed 2019-09-16]
4. Bray F, Soerjomataram I. The changing global burden of cancer: transitions in human development and implications for cancer prevention and control. In: Cancer: Disease Control Priorities. London, UK: The International Bank for Reconstruction and Development / The World Bank; 2015.
5. Schneider C, Hsieh C, Sprod L, Carter S, Hayward R. Cancer treatment-induced alterations in muscular fitness and quality of life: the role of exercise training. Ann Oncol 2007 Dec;18(12):1957-1962 [FREE Full text] [doi: 10.1093/annonc/mdm364] [Medline: 17804476]
6. Reeves MM, Terranova CO, Eakin EG, Demark-Wahnefried W. Weight loss intervention trials in women with breast cancer: a systematic review. Obes Rev 2014 Sep;15(9):749-768. [doi: 10.1111/obr.12190] [Medline: 24891269]
7. Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer--mechanisms underlying tumour progression and recurrence. Nat Rev Endocrinol 2014 Aug;10(8):455-465 [FREE Full text] [doi: 10.1038/nrendo.2014.94] [Medline: 24935119]
8. Dale H, Brassington L, King K. The impact of healthy lifestyle interventions on mental health and wellbeing: a systematic review. Mental Health Rev J 2014 Mar 5;19(1):1-26. [doi: 10.1108/mhrj-05-2013-0016]
9. Fong DY, Ho JW, Hui BP, Lee AM, Macfarlane DJ, Leung SS, et al. Physical activity for cancer survivors: meta-analysis of randomised controlled trials. Br Med J 2012 Jan 30;344:e70 [FREE Full text] [doi: 10.1136/bmj.e70] [Medline: 22294757]
10. Mishra S, Scherer R, Geigle P. Exercise interventions on health-related quality of life for cancer survivors. Cochrane Database Syst Rev 2012;8:1-295. [doi: 10.1002/14651858.cd007566]
11. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
12. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
13. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mhealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
14. Walsh J, Groarke J. Integrating behavioral science with mobile (mhealth) technology to optimize health behavior change interventions. Eur Psychol 2019 Jan;24(1):38-48 [FREE Full text] [doi: 10.1027/1016-9040/a000351]
15. 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]

16. 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]
17. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609 [FREE Full text] [doi: 10.1111/obr.12165] [Medline: 24636238]
18. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
19. Finne E, Glausch M, Exner A, Sauzet O, Stölzel F, Seidel N. Behavior change techniques for increasing physical activity in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Cancer Manag Res 2018;10:5125-5143 [FREE Full text] [doi: 10.2147/CMAR.S170064] [Medline: 30464612]
20. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
21. de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
22. 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]
23. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
24. Hoffman AJ, Brintnall RA, Brown JK, von Eye A, Jones LW, Alderink G, et al. Virtual reality bringing a new reality to postthoracotomy lung cancer patients via a home-based exercise intervention targeting fatigue while undergoing adjuvant treatment. Cancer Nurs 2014;37(1):23-33. [doi: 10.1097/ncc.0b013e318278d52f]
25. Steene-Johannessen J, Anderssen SA, Van Der Ploeg HP, Hendriksen IJM, Donnelly AE, Brage S, et al. Are self-report measures able to define individuals as physically active or inactive? Med Sci Sports Exer 2016;48(2):235-244. [doi: 10.1249/mss.0000000000000760]
26. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 6;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
27. Loney T, Standage M, Thompson D, Sebire S, Cumming S. Self-report vs objectively assessed physical activity: which is right for public health? J Phys Act Health 2011 Jan;8(1):62-70. [doi: 10.1123/jpah.8.1.62] [Medline: 21297186]
28. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun;71(Suppl 2):1-14. [doi: 10.1080/02701367.2000.11082780] [Medline: 25680007]
29. Ward D, Evenson K, Vaughn A, Rodgers A, Troiano R. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S582-S588. [doi: 10.1249/01.mss.0000185292.71933.91] [Medline: 16294121]
30. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3(1):e000254 [FREE Full text] [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
31. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A Web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
32. 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]
33. Demark-Wahnefried W, Morey MC, Sloane R, Snyder DC, Miller PE, Hartman TJ, et al. Reach out to enhance wellness home-based diet-exercise intervention promotes reproducible and sustainable long-term improvements in health behaviors, body weight, and physical functioning in older, overweight/obese cancer survivors. J Clin Oncol 2012 Jul 1;30(19):2354-2361. [doi: 10.1200/jco.2011.40.0895]
34. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. J Am Med Assoc 2009 May 13;301(18):1883-1891 [FREE Full text] [doi: 10.1001/jama.2009.643] [Medline: 19436015]
35. Groarke JM, Richmond J, Kelly MG, McSharry J, Groarke A, Kerr T, et al. Examining the impact of a personalized self-management lifestyle program using mobile technology on the health and well-being of cancer survivors: protocol and rationale for a randomized controlled trial (the moving on study). JMIR Res Protoc 2019 Aug 23;8(8):e13214. [doi: 10.2196/13214]

36. 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]
37. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
38. Ware JE, Gandek B. Overview of the sf-36 health survey and the international quality of life assessment (IQOLA) project. J Clin Epidmol 1998 Nov;51(11):903-912. [doi: 10.1016/s0895-4356(98)00081-x]
39. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
40. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer 1999 Mar 1;85(5):1186-1196. [doi: 10.1002/(sici)1097-0142(19990301)85:5<1186::aid-cncr24>3.0.co;2-n] [Medline: 10091805]
41. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Measures in Health Psychology: A User's Portfolio. Windsor, UK: Nfer-Nelson; 1995:35-37.
42. Armitage CJ. Can the theory of planned behavior predict the maintenance of physical activity? Health Psychol 2005 May;24(3):235-245. [doi: 10.1037/0278-6133.24.3.235] [Medline: 15898858]
43. Molloy G, Dixon D, Hamer M, Sniehotta F. Social support and regular physical activity: does planning mediate this link? Br J Health Psychol 2010;15:859-870. [doi: 10.1348/135910710x490406]
44. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
45. Bingham SA, Welch AA, McTaggart A, Mulligan AA, Runswick SA, Luben R, et al. Nutritional methods in the European prospective investigation of cancer in Norfolk. Public Health Nutr 2001 Jun;4(3):847-858. [doi: 10.1079/phn2000102] [Medline: 11415493]
46. Little R. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
47. O'Brien MW, Kivell M, Wojcik W, d'Entremont G, Kimmerly D, Fowles J. Step rate thresholds associated with moderate and vigorous physical activity in adults. Int J Environ Res Public Health 2018 Nov 03;15(11) [FREE Full text] [doi: 10.3390/ijerph15112454] [Medline: 30400331]
48. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
49. Booth HP, Prevost TA, Wright AJ, Gulliford MC. Effectiveness of behavioural weight loss interventions delivered in a primary care setting: a systematic review and meta-analysis. Fam Pract 2014 Dec;31(6):643-653 [FREE Full text] [doi: 10.1093/fampra/cmu064] [Medline: 25298510]
50. LeVasseur N, Cheng W, Mazzarello S, Clemons M, Vandermeer L, Jones L, et al. Optimising weight-loss interventions in cancer patients-a systematic review and network meta-analysis. PLoS One 2021;16(2):e0245794 [FREE Full text] [doi: 10.1371/journal.pone.0245794] [Medline: 33539414]
51. Franssen WM, Franssen GH, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
52. Stein K, Syrjala K, Andrykowski M. Physical and psychological long-term and late effects of cancer. Cancer 2008 Jun 1;112(11 Suppl):2577-2592 [FREE Full text] [doi: 10.1002/cncr.23448] [Medline: 18428205]
53. Groarke JM, Richmond J, Mc Sharry J, Groarke A, Harney OM, Kelly MG, et al. Acceptability of a mobile health behavior change intervention for cancer survivors with obesity or overweight: nested mixed methods study within a randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 16;9(2):e18288 [FREE Full text] [doi: 10.2196/18288] [Medline: 33591290]
54. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 1;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
55. Chevance G, Baretta D, Golaszewski N, Takemoto M, Shrestha S, Jain S, et al. Goal setting and achievement for walking: a series of N-of-1 digital interventions. Health Psychol 2021 Jan;40(1):30-39 [FREE Full text] [doi: 10.1037/hea0001044] [Medline: 33252961]
56. Swann C, Rosenbaum S, Lawrence A, Vella SA, McEwan D, Ekkekakis P. Updating goal-setting theory in physical activity promotion: a critical conceptual review. Health Psychol Rev 2021 Mar;15(1):34-50. [doi: 10.1080/17437199.2019.1706616] [Medline: 31900043]

57. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014 Feb 24;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
58. Courneya KS, Friedenreich CM, Quinney HA, Fields AL, Jones MW, Fairey AS. A randomized trial of exercise and quality of life in colorectal cancer survivors. Eur J Cancer Care (Engl) 2003 Dec;12(4):347-357. [doi: 10.1046/j.1365-2354.2003.00437.x] [Medline: 14982314]
59. Cadmus LA, Salovey P, Yu H, Chung G, Kasl S, Irwin ML. Exercise and quality of life during and after treatment for breast cancer: results of two randomized controlled trials. Psychooncology 2009 Apr;18(4):343-352 [FREE Full text] [doi: 10.1002/pon.1525] [Medline: 19242918]
60. Shelton ML, Lee JQ, Morris GS, Massey PR, Kendall DG, Munsell MF, et al. A randomized control trial of a supervised versus a self-directed exercise program for allogeneic stem cell transplant patients. Psychooncology 2009 Apr;18(4):353-359. [doi: 10.1002/pon.1505] [Medline: 19117328]
61. Senn S. Seven myths of randomisation in clinical trials. Stat Med 2013 Apr 30;32(9):1439-1450. [doi: 10.1002/sim.5713] [Medline: 23255195]
62. Courneya KS. Physical activity in cancer survivors: a field in motion. Psychooncology 2009 Apr;18(4):337-342. [doi: 10.1002/pon.1546] [Medline: 19306338]
63. Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28 [FREE Full text] [doi: 10.2196/mhealth.9117] [Medline: 29371177]
64. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016 Jul;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025] [Medline: 27477705]
65. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport 2014 Sep;17(5):496-500. [doi: 10.1016/j.jsams.2013.10.241] [Medline: 24268570]
66. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
67. 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]
68. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current state of commercial wearable technology in physical activity monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
69. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 9;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
70. Migueles JH, Cadenas-Sanchez C, Ekelund U, Nyström CD, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
71. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. Sleep Health 2015 Dec;1(4):275-284. [doi: 10.1016/j.sleh.2015.09.002] [Medline: 29073403]
72. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2018 Jul;52(13):844-850. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]

Abbreviations

ANOVA: analysis of variance
BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials
EM: expectation-maximization
MD: mean difference
mHealth: mobile health
MOD: mode of delivery
RCT: randomized controlled trial

Edited by L Buis; submitted 09.10.20; peer-reviewed by E Lyons, W Demark-Wahnefried, S Mikles; comments to author 15.12.20; revised version received 01.04.21; accepted 07.05.21; published 05.07.21.

Please cite as:
Walsh JC, Richmond J, Mc Sharry J, Groarke A, Glynn L, Kelly MG, Harney O, Groarke JM
Examining the Impact of an mHealth Behavior Change Intervention With a Brief In-Person Component for Cancer Survivors With Overweight or Obesity: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(7):e24915
URL: https://mhealth.jmir.org/2021/7/e24915
doi:10.2196/24915
PMID:36260394

Original Paper

A Clinician-Led, Experience-Based Co-Design Approach for Developing mHealth Services to Support the Patient Self-management of Chronic Conditions: Development Study and Design Case

Ting Song[1,2], MSc; Ping Yu[1,2,3*], PhD; Vida Bliokas[2,4], PhD; Yasmine Probst[2,5], PhD; Gregory E Peoples[5], PhD; Siyu Qian[1,2,6], PhD; Lauren Houston[2,5], PhD; Pascal Perez[2,3], PhD; Mehrdad Amirghasemi[3], PhD; Tingru Cui[7], PhD; Nadeesha Pathiraja Rathnayaka Hitige[1], BSc; Natalie Anne Smith[2,8*], MBBS

[1]Centre for Digital Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia

[3]Smart Infrastructure Facility, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[4]School of Psychology, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia

[5]School of Medicine, Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, Australia

[6]Illawarra Shoalhaven Local Health District, Wollongong, Australia

[7]School of Computing and Information Systems, Faculty of Engineering and Information Technology, University of Melbourne, Melbourne, Australia

[8]Department of Anaesthesia, Wollongong Hospital, Wollongong, Australia

[*]these authors contributed equally

Corresponding Author:
Natalie Anne Smith, MBBS
Department of Anaesthesia
Wollongong Hospital
Crown Street
Wollongong, 2500
Australia
Phone: 61 2 4225 1554
Fax: 61 2 4225 1561
Email: NatalieAnne.Smith@health.nsw.gov.au

Abstract

Background: Despite the increasing use of mobile health (mHealth) services, such as mHealth apps or SMS text messaging services, that support the patient self-management of chronic conditions, many existing mHealth services lack theoretical guidance. In addition, although often the target audience for requirement acquisition at the initial mHealth app design stage, it is a common challenge for them to fully conceptualize their needs for mHealth services that help self-manage chronic conditions.

Objective: This study proposes a novel co-design approach with the initial requirements for mHealth services proposed by clinicians based on their experiences in guiding patients to self-manage chronic conditions. A design case is presented to illustrate our innovative approach to designing an mHealth app that supports the self-management of patients with obesity in their preparation for elective surgery.

Methods: We adopted a clinician-led co-design approach. The co-design approach consisted of the following four cyclic phases: understanding user needs, identifying an applicable underlying theory, integrating the theory into the prototype design, and evaluating and refining the prototype mHealth services with patients. Expert panel discussions, a literature review, intervention mapping, and patient focus group discussions were conducted in these four phases.

Results: In stage 1, the expert panel proposed the following three common user needs: motivational, educational, and supportive needs. In stage 2, the team selected the Social Cognitive Theory to guide the app design. In stage 3, the team designed and developed the key functions of the mHealth app, including automatic push notifications; web-based resources; goal setting and monitoring; and interactive health-related exchanges that encourage physical activity, healthy eating, psychological preparation,

and a positive outlook for elective surgery. Push notifications were designed in response to a patient's risk level, as informed by the person's response to a baseline health survey. In stage 4, the prototype mHealth app was used to capture further requirements from patients in the two focus group discussions. Focus group participants affirmed the potential benefits of the app and suggested more requirements for the function, presentation, and personalization needs. The app was improved based on these suggestions.

Conclusions: This study reports an innovative co-design approach that was used to leverage the clinical experiences of clinicians to produce the initial prototype app and the approach taken to allow patients to effectively voice their needs and expectations for the mHealth app in a focus group discussion. This approach can be generalized to the design of any mHealth service that aims to support the patient self-management of chronic conditions.

(JMIR Mhealth Uhealth 2021;9(7):e20650) doi:10.2196/20650

KEYWORDS

mHealth; smartphone; mobile apps; chronic disease; surgery; obesity; theory; community-based participatory research; mobile phone

Introduction

Background

With the ubiquity of smartphones and the internet, there has been an increasing use of mobile health (mHealth) services in health care systems worldwide to support the patient self-management of chronic conditions, such as hypertension, diabetes, and obesity [1-4]. This has become even more salient since 2020, as the demand for and use of telemedicine services has increased due to the COVID-19 pandemic [4,5]. The use of mobile apps to collect various patient data, such as ecological momentary assessment or experience sampling, has gained momentum to support self-monitoring, develop self-awareness, and promote behavioral change [6,7].

The design techniques of mHealth services have also been continuously evolving, from traditional system design methods focusing on their appearance, functionality, and values to interactive design methods focusing on the way users interact with mHealth services. Common interactive design methods include user-centered design [8-10], activity-centered design [11], and goal-directed design [3,12], focusing on the object, process, and outcome of product use, respectively. To increase user engagement and meet the needs of multidisciplinary collaboration, the development team can also invite all stakeholders, such as domain experts, users, and researchers, to participate in the design and development, known as co-design [13,14]. This is especially relevant when addressing specific diseases or improving physical and mental well-being.

Behavioral change theories and models, such as the Social Cognitive Theory (SCT) [15] and Health Belief Model (HBM) [16], drawn from psychology and sociology [17-19], focus on predicting and explaining human behavior and the wide range of factors that affect these behaviors, such as emotions, habits, and daily routines. These theories provide a roadmap for scientific research and practice [20] and are useful in guiding the implementation of successful mHealth services [21]. mHealth services based on sound behavioral change theories are more likely to lead to positive changes in health behavior [22,23], that is, to lead to successful changes in physical activity and healthy eating [24,25]. Behavioral change techniques include monitoring, intention building, goal setting and planning, progress reflecting, and performance reporting [26,27]. They can be implemented in more interactive and dynamic functionalities in mHealth apps to motivate patients [28].

Despite their increasing popularity, the reported effectiveness of mHealth services for self-management of chronic conditions has been mixed [1,2,29-32]. The frequency with which large numbers of such services enter the market, coupled with the limited time that professional clinicians have available, has inhibited clinician participation in mHealth app design or evaluation [1]. Heterogeneity in mHealth design and purpose also leads to different levels of app usage. For example, apps only being designed as data collection tools instead of comprehensive interventions might not lead to positive changes in health behavior [28]. Similarly, merely providing health information on a regular basis is proven to be ineffective unless reinforcement and motivation are provided [33,34]. As mentioned earlier, inadequate application of the behavioral change theory to guide the design and implementation of mHealth services also leads to unsatisfactory intervention effects [23,31,35,36]. In addition, although often the target audience for requirement acquisition at the initial mHealth app design stage, it is a common challenge for patients to fully conceptualize their needs for mHealth services that support them to self-manage chronic conditions [32]. These may lead to the ineffectiveness of mHealth services [3,9,10,13,14].

Research Aim

To address the abovementioned limitations for the design of effective mHealth services that support the patient self-management of chronic conditions, this study proposes a novel approach with the initial requirements for mHealth services proposed by clinicians based on their experiences in guiding patients to self-manage chronic conditions. We adopted a co-design approach with multidisciplinary collaboration to improve knowledge about patients' need for mHealth services.

A Design Case of mHealth Services for Preoperative Obesity Management

Obesity has increasingly become a global public health challenge, with 5.9 million Australians (31.3%) having a BMI ≥30 kg/m^2 in 2017-2018 [37]. Obesity can complicate procedures such as siting intravenous cannulae and inserting endotracheal tubes. It may affect weight-based decisions such as ventilator settings or drug doses and can also make surgical

access more difficult [38,39]. Obesity is also a risk factor for short-term postoperative complications such as infection, deep vein thrombosis, poor wound healing, blood loss, respiratory problems, and myocardial infarction [38,40,41]. In Australia, the average waiting time for elective surgery in public hospitals is 41 days [42]. Losing weight and improving preoperative fitness through lifestyle changes during this period, known as prehabilitation, is gaining increasing attention [43,44].

Common methods of prehabilitation include engagement in regular physical activities, dietary optimization, and psychological support [45,46]. Current evidence suggests that higher preoperative fitness can lead to fewer last-minute cancelations, better postoperative outcomes, and shorter patient waiting times [47-49]. However, many local health care systems do not have enough resources to help patients with obesity improve their preoperative fitness for surgery, even when the period between the booking and performance of surgery can be up to 12 months. Moreover, dietary modifications and changes in physical activity are difficult to maintain [50,51]. Therefore, innovative methods are needed to encourage and motivate patients with obesity to improve their physical fitness, dietary habits, and mental well-being before elective surgery.

To date, the use of mHealth services to improve preoperative fitness is in its infancy [52,53]. There are bariatric surgery-specific apps available in app stores; however, no studies have reported the use of a behavioral theory to guide the development of mHealth services to deliver health and weight management coaching to patients with obesity before elective surgery [54,55]. Considering that there is still room for further research on embedding a relevant behavioral theory within mHealth apps to improve effectiveness, a case is presented to illustrate our innovative approach to designing an mHealth app that supports self-management of patients with obesity in their preparation for elective surgery.

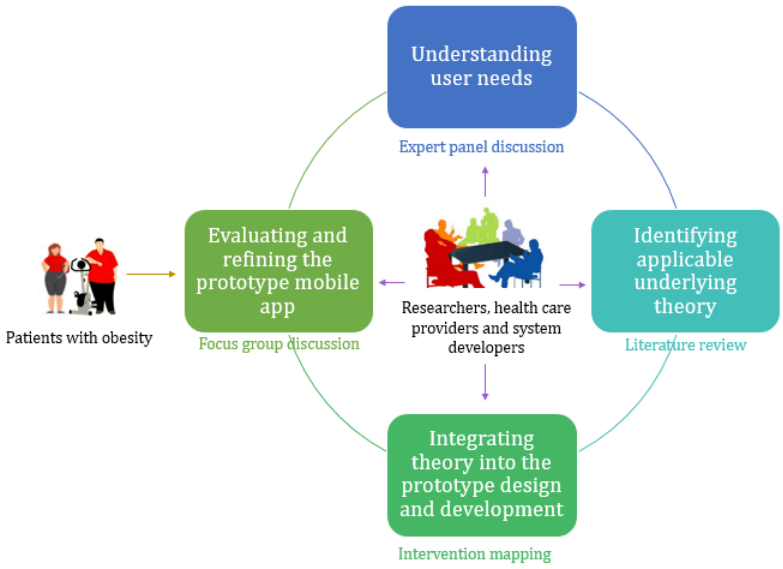
Methods

Overview

A clinician lead co-design approach was undertaken by a multidisciplinary team for designing mHealth services to support the patient self-management of chronic conditions. The experience-based co-design principle [56] and the guidelines for developing complex interventions to improve health and health care published by the UK Medical Research Council [57] have been followed to design the mHealth app. Experience-based co-design claims that all stakeholders, including researchers, developers, and service users, participate in the design process and develop a set of feasible service plans or care paths by gathering their experiences [56,58,59]. The Medical Research Council framework defines a series of actions for intervention development [57]. On the basis of these references, we formulated four iterative phases for the prototype mHealth app design: understanding user needs, identifying applicable underlying theories, integrating theory into the prototype design and development, and evaluating and refining the prototype mobile app (Figure 1 [57]).

In phase 1, we aimed to understand user needs through an expert panel discussion. The panel included domain experts in the medical field and health information systems. Domain (medical) experts can put forward specific challenges that patients may face in the self-management of chronic diseases based on their clinical experience and can also provide various targeted professional assessments and interventions for integration into the mHealth service. Health information system experts can effectively transform this information into system requirements and design considerations to achieve an optimal design solution. In phase 2, a literature review was conducted to compare the relevant theories and to select the optimal one, as suggested by the empirical studies to guide the mHealth service design, in accordance with the guidance of Nash and Barnier [20] and Yang and Van Stee [21]. In phase 3, the selected theory was implemented in the prototype mHealth design and development using the intervention mapping approach, that is, using theory and the corresponding constructs to propose the relevant functions of the mHealth service, which is widely used for the development of theory-based health promotion programs [60]. In phase 4, a focus group discussion was conducted with patients. The patients were provided with the opportunity to interact with a living prototype mHealth service, which facilitates the identification of their needs and their desired functions of the mHealth service. The tangible feedback provided the research team with evidence to further improve the design of mHealth services.

Figure 1. Stakeholders, procedures, and methods of the intervention development process.



Phase 1: Understanding User Needs

A multidisciplinary team of 12 experts conducted panel discussions at the University of Wollongong, Australia, to understand potential user needs; 8 team members hold a PhD and 4 hold master's or bachelor's (with Honors) degrees as their highest qualification. In total, 9 panel members had more than 5 years of research experience related to health sciences.

A clinical anesthetist (NAS), with a 20-year clinical experience in Wollongong Hospital, described the issues related to obesity, anesthesia, and surgery, which provided information on user needs from a clinical perspective. She noted that the challenge of obesity is particularly problematic in the local area, the Illawarra Shoalhaven Local Health District, New South Wales, Australia, in which the prevalence of obesity is higher than the national figure (36% compared with 28%) [61]. According to local audit data, one-third of patients scheduled for elective surgery waited more than three months between booking and operation, and about half of them gained weight while on the waiting list. In Wollongong Hospital, approximately 55% of the 6000 patients who undergo elective surgery each year have obesity [62].

Together with an eHealth researcher (PY), these authors proposed that an mHealth intervention could potentially address the needs of this population. An accredited clinical psychologist (VB), accredited practicing dietitian (YP), and accredited exercise physiologist (GEP) provided specific input in their areas of expertise. Researchers in health information systems and software engineering were involved in designing technical solutions for delivering the intervention.

Phase 2: Identifying Applicable Underlying Theory

After understanding the preliminary user needs, the researchers conducted a literature review to identify applicable underlying theories to guide the app design. Three interdisciplinary databases (Scopus, PubMed, and PMC) were searched, which allowed the inclusion of peer-reviewed English-language journal articles. Terms, that is, Medical Subject Headings and their variants, applied were *theory or model*, *intervention or program*, and *behavioral change*. Empirical studies and systematic reviews that reported the explicit use of theory to guide lifestyle-related behavioral changes were included in the study. Studies that did not report why or how the theory was used were excluded. The data were extracted in a Microsoft Excel spreadsheet for constant comparison and analysis. The team then selected the most suitable theory to guide the prototype design for this study.

Phase 3: Integrating Theory Into the Prototype Design and Development

After selecting the SCT as the most applicable theory, we used it as a framework to integrate the prototype design, following the intervention mapping protocol [60]. First, we listed all user needs based on their understanding of health issues, risk groups, behavioral and environmental determinants, and available resources. Second, we listed the constructs of the selected theory and the evidence-based intervention techniques used to guide the behavioral changes that fit the abovementioned intervention context. Third, we mapped these intervention techniques to the app design to build different functional modules. Fourth, we designed the system architecture, functional modules, user interface, and database and integrated these components into a coherent program (ie, that of a prototype mobile app), through several iterations. React Native, an open-source framework that uses JavaScript and React to develop native, iOS, and Android mobile apps, was used in the development [63].

Phase 4: Evaluating and Refining the Prototype Mobile App

The prototype mobile app was pilot used by patients with obesity to gauge their perceptions of its usefulness and usability. The pilot trial was conducted in November 2019 in two focus group discussions with 6 people per group. Participants were recruited via purposive sampling of patients with obesity who were undergoing weight loss treatment at a hospital in the South-Western Sydney Local Health Service via existing networks. The inclusion criteria were patients who (1) were

aged ≥18 years, (2) had a BMI ≥35 kg/m^2, (3) were English-speaking with self-elected adequate reading skills, and (4) provided informed consent to participate in the study. A clinic nurse initially discussed the project and sought verbal consent from each patient to participate in the focus group discussion. One week later, patients were sent a text message asking for confirmation of their verbal consent. Ethical approval for the study was obtained from the University of Wollongong and Illawarra and Shoalhaven Local Health District Health and Medical Human Research Ethics Committee (2018/175; HREC/18/WGONG/64).

One researcher (LH), with prior training in research theory and experience in conducting and observing qualitative research, moderated the semistructured focus group discussions. For each group, she gave a brief introduction to the participants and asked them to sign a written consent form. She then distributed 3 mobile phones and 3 mobile tablets with the app preinstalled for all participants and instructed them on its use. After the time spent with the app, a semistructured list of questions was deliberated in the group, discussing the relevant functions of the app (Multimedia Appendix 1). The conversation continued until all of the relevant issues and opinions were openly raised and discussed, beyond answering the interview questions.

The focus group discussions were audio-recorded and transcribed verbatim; 3 other researchers were present as observers and took notes to record the nonverbal characteristics of the focus groups, such as gender. The total discussion time for each group was approximately 1 hour and 30 minutes.

The transcripts were analyzed using a content analysis approach to capture patient feedback [64]. Each original sentence was judged by 2 researchers independently to see if it contained a suggestion that would be useful for improving app functionality. These suggestions were listed in point form and circulated to the expert team. The developers discussed the approved and feasible suggestions for further modification.

Results

Phase 1: Understanding User Needs

The expert panel discussion proposed three kinds of potential needs for patients with obesity to improve fitness before surgery: motivational needs, educational needs, and supportive needs.

Motivational Needs

Many individuals find it difficult to maintain sufficient motivation to lose weight over time. Many have repeated failed experiences, often with initial weight loss followed by regaining weight, which can further decrease confidence and motivation [50,65]. It is well recognized that supporting motivation is essential for sustainable behavioral change [3,27,66,67].

Educational Needs

Educating patients with obesity about the general health risks associated with obesity remains important. Most studies have not considered risks that are specifically associated with anesthesia and surgery, and many do not realize that obesity itself poses an additional perioperative risk. Therefore, we felt it important to educate patients faced with upcoming surgery about the related risks in a manner that was personally tailored to their specific situation. The aim of increasing awareness was to capitalize on the potential *teachable moment* of surgical booking to encourage behavioral change [68-70].

Supportive Needs

Even if patients are aware of obesity-related health risks, many need ongoing guidance about strategies and encouragement to stay motivated in changing their lifestyles to lose weight. Therefore, it is essential to provide timely, relevant information and regular interactions to improve patient skills and encourage a positive attitude toward behavioral change. Regular reminders are an effective means of providing personalized, targeted support [66,67].

Phase 2: Identifying Applicable Underlying Theory

In addition to the SCT and HBM, theories relating to health behavior include the Theory of Planned Behavior [71], Self-determination Theory [72], Transtheoretical Model [73], Community Organization Model [74], and Diffusion of Innovation Theory [75] (see the detailed comparison in Multimedia Appendix 2).

As the HBM and Theory of Planned Behavior focus only on rational reasoning, ruling out unconscious, spontaneous behavior and its emotional effects [53,76], and the Self-determination Theory is confined to explaining only behavioral motivations, none of these was considered suitable as a guide for the development of interventions that provide comprehensive health care support, such as health education and reminders [72]. The Transtheoretical Model cannot explain how an individual thinks he or she is ready to cope with a change or not, which would have caused difficulty in mapping intervention techniques to the behavioral determinant factors, thus weakening the explanatory capacity of the theory [77]. The Community Organization Model and Diffusion of Innovation Theory are both focused on initiatives to support community health promotion at the population level; therefore, they are not intended to guide the development of interventions to support behavioral change for individual patients [78]. Finally, the SCT was selected to guide the design and development of the app.

The SCT explains an individual's behavior through a reciprocal model of interactions among behavior, personal factors, and the social environment. It is a theory that synthesizes a wide range of behavioral, cognitive, and environmental determinants of behavioral change, such as self-efficacy, observational learning, outcome expectations, and additional reinforcement [46]. This theory not only explains the behavior of individuals under rational circumstances but also describes the influence and interaction of internal cognitive and external environmental influences on human behavior [15], so it can be applied to guide the design of complex interventions to support the management of chronic conditions. Moreover, the SCT considers that people learn not only through their own experience but also through imitating behaviors and the results of these behaviors. We felt that this was consistent with the purpose of this study to provide professional coaching to guide patients with obesity in terms of preoperative weight loss and fitness improvement. Therefore,

the SCT was chosen as the most appropriate theory to guide the intervention design in our study.

Phase 3: Integrating Theory Into the Prototype Design and Development

The SCT contains seven significant constructs: self-control, self-efficacy, expectations, expectancies, reinforcement, behavior capacity, and observational learning. All of these constructs were used to guide the design of conceptual intervention techniques and functions in the app (Table 1).

Our app, Fitness4Surgery, consists of two interfaces (Figure 2 and Multimedia Appendix 3): a mobile interface for patients to self-manage their obesity and a web-based portal for the health care administrators to edit, modify, and update the content of push notifications, view patients' interactions with the mobile app, and formulate text interventions.

At the initial log-in to the app, patients are requested to answer the questionnaire surveys about their mobile phone usage experience, level of physical activity, diet, psychological well-being, and preoperative health. The system classifies their level of function as *high* or *low* based on these data and will provide corresponding push notifications automatically; 96 push notifications were designed by four domain experts, that is, the clinician, the clinical psychologist, the practicing dietitian, and the exercise physiologist, based on advice from the Australian Dietary Guidelines and Australia's Physical Activity and Sedentary Behavior Guidelines and the research evidence [79-81] (see exemplar notifications in Table 2).

Table 1. User needs, theoretical construct, conceptual intervention techniques, and app functionalities.

User needs and theoretical construct	Conceptual intervention techniques	App functionalities
Motivational needs		
Self-control	• Let users set goals and encourage them to monitor their own behavior toward the achievement of these goals	• Build a functional module named "My Goals"
Self-efficacy	• Let users start by setting small, progressive, and realistic goals	• Provide some simple exemplar goals in "My Goals"
Expectations	• Let users know about the benefits of fitness for surgery	• Provide an introductory video about the significance of fitness for surgery
Expectancies	• Allow users to track and monitor their own changes in weight, diet, and physical activity. Provide feedback and let users evaluate what they value	• Build a functional module named "My Steps" to record the number of steps each day • Build a functional module named "Surveys" and ask the users to complete these surveys • Send different push notifications according to different responses for feedback
Reinforcements	• Let users recognize and praise their achievements by specifying rewards	• Make a trophy pop out once a user achieves their goal
Educational needs		
Behavior capacity	• Teach users how to self-manage diet, physical activity, mood, and medical conditions	• Translate relevant educational information into push notifications and send them to the users
Observational learning	• Let users watch some actions and outcomes of others' behavior	• Provide an introductory video about how to achieve fitness for surgery
Supportive needs		
Reinforcements	• Remind users to perform the behavioral change toward fitness for surgery • Provide users with toolkits and resources that make the new behaviors easier to perform	• Send push notifications • Build a functional module named "My Resources"

Figure 2. Working mechanisms of Fitness4Surgery.

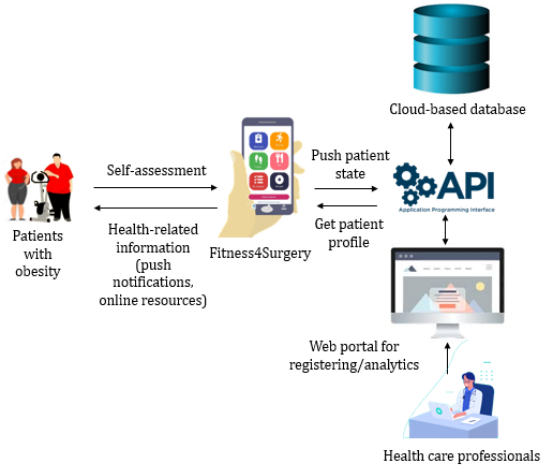


Table 2. Exemplar push notifications in the four domains.

Domain and content	Risks for the group of patients
Physical activity	
"Limit sitting time to a maximum of 30 minutes."	Low
"Limit sitting time to a maximum of 20 minutes."	High
Diet	
"Each day this week take a photo of your meal."	High or low
Psychology	
"You've got this! Stick with it. Set goals and take them one step at a time. It will be worth it!"	High or low
Medical advice	
"Reducing weight before your surgery can benefit your recovery and prevent unwanted complications. Work out your target weight by speaking to your general practitioner or head to Get Healthy NSW[a] by phone or on the web for more help."	High or low

[a]NSW: New South Wales.

Patients with obesity are asked to set goals on the app. Once a goal is reached, a trophy pop will be displayed on the screen as a reward. They will be presented with links to existing health resources, such as the Heart Foundation [82] and Get Healthy New South Wales [83]. The patient-only use of the app keeps the data entered by the patient confidential. Patients can show the records to their doctors for discussion if they wish. There is a large variation in each clinician's approach to managing obesity [84]; therefore, the app did not cover this function. A notification will be automatically sent to patients every month to ask them to update their responses to the surveys.

Phase 4: Evaluating and Refining the Prototype Mobile App

Evaluation

Overview

The focus group participants reported two perceived benefits of the app: usefulness and ease of use. They also discussed areas that needed improvement for the four functional modules, that is, Survey, My Goals, My Resources, and push notifications (Table 3).

Table 3. Data analysis of the focus group discussion.

Theme, category, and topic	Exemplar quote
Perceived benefits	
Usefulness	
Evaluate health conditions and health literacy before surgery	"Well, yes it's a good app, because it's got a lot of benefits; our progress, our goals, resources we can use, so, yes, it's probably good for everybody here." [group 1-05, female]
Access to health-related knowledge, skills, and referrals	"Looking at the resources that they've offered, it's a personal lifestyle app." [group 2-05, male]
Improving the patients' motivation	"If it's just going to be a tool for yourself, perhaps an inspirational tool?" [group 1-01, male]
Ease of use	
Easy to use the system	"Yes, I think it's good, it's quite easy to follow, because I'm not very good at this, but I found it quite easy." [group 1-02, female]
Improvement	
Survey	
Inflexibility	"I physically cannot walk ten minutes, so my answer to the question is zero, but when I tried to complete the page, it's telling me it's incomplete, so all the information I put in is going to be junked (deleted)." [group 2-03, female]
Ambiguity	"Click on what? No, something's wrong." [group 1-02, female]
My Goals	
Insufficient feature	"When you complete the survey at the start about you, yes, you've got a baseline, but over a period of time, the app is not collecting anything." [group 2-02, male]
Suboptimal interface design	"What about a graph about filling out or achieving the goals and seeing...something visual." [group 1-05, female]
Lack of tracking	"...is there a place where you can monitor your weight loss as you go?" [group 1-06, female]
My Resources	
Insufficient information and referrals	"...it does need more resources in it" [group 2-02, male]
Push notifications	
Personalization	"Maybe there's an option that you can turn that setting on or off...So, I like to stipulate what time (messages arrive)." [group 1-06, female]

Perceived Benefits

Overall, the participants reported that the app could be useful, and they were looking forward to using this product for their surgical preparation. They felt that the app would help them (1) evaluate their health conditions and health literacy before surgery; (2) access health-related knowledge, skills, and referrals; and (3) improve their motivation by setting goals and rewarding their achievements. Some praised the ease of use of the app, even if they were not proficient in using smartphones.

Improvement

Survey

The participants raised two issues regarding the module. First, the operability of the system was not sufficiently flexible. They found that some questions in the questionnaire were irrelevant to their own situation. As all questions were mandatory, it was difficult to continue entering the data. Therefore, they suggested that some questions should be optional, which would allow them to skip the questions that were irrelevant to them or move through this section more quickly if they wished to. The second was semantic ambiguity. A few participants were confused about the meaning of certain questions and did not know how they should answer. They suggested modifying the expression of some questions to make them easier to understand.

My Goals

The participants raised three issues regarding the module. The first issue was monotony. They stated that it was boring because of its simplistic function and presentation, with an inadequate interface design. They suggested having some preset common goals as examples for users to choose from while retaining the flexibility to set their own goals. In particular, they felt that the color scheme was uninspired, with insufficient incentives provided to achieve their preset goals. They strongly requested the use of different colors, shapes, and icons to enrich the interface, with the provision of visual rewards once goals were completed, such as the appearance of animated trophies or fireworks. The second issue was functional. The participants felt that the timeframe for goals was important but not currently well defined. The third issue was the lack of a tracking mechanism. They felt disappointed because they could not monitor their progress toward achieving their goals. They also described several functions that they expected or found in other apps.

My Resources

The participants were generally satisfied with this module, except that some requested the provision of more psychological support. One participant requested a recipe section in the app.

Push Notifications

The main focus of this discussion was on the optimal delivery time and frequency of these notifications. Some felt that daily notifications would help remind them of healthy routines, whereas others felt that excessive reminders could be overwhelming. One suggestion that was supported by the participants in both groups was that the app should allow users to set the timing that suited them to receive push notifications and that they could turn the reminder on or off themselves.

After Focus Group Refinements

The refined design of the interface was simplified into three pages: *Home*, *Messages*, and *Settings* (Figure 3). The Home page consists of two major parts: user information and four buttons—My Goals, My Weight, My Surveys, and My Resources. The user information includes the name, profile, step count, and weight record. The display of weight at multiple points over time encourages users to progress toward their target weight.

In the My Goals module, the ability to select preset goals from a list was added to the free-text space. Once a goal is achieved, the user clicks once to record it and receives a star as a reward; 10 stars can be exchanged for a trophy and 10 trophies for a firework. A progress tracking function was also incorporated into the module.

A new module, My Weight, was added to the Home page, where users can read the changes in weight and BMI, both numerically and graphically. This allows users to track and monitor their own changes in weight, which addresses the application of expectancies in the SCT.

Apart from color and layout adjustments, the functions of My Surveys and My Resources remained unchanged. Users were able to answer the surveys at any time, and the results were recorded in a new area accessible for later review.

Figure 3. Screenshots of key functional modules of the refined Home page of the Fitness4Surgery app. (A) Home page, (B) goal selection, (C) achieved goals record, (D) weight and BMI record, (E) survey record, and (F) web-based resources.

Discussion

Principal Findings

To date, limited evidence exists regarding the use of a behavioral theory to guide the development of mobile services to support the patient self-management of chronic conditions, particularly in the context of prehabilitation for patients with obesity who are awaiting surgery. Guided by an existing framework for developing complex interventions to improve health and health care, this multidisciplinary study proposed a clinician-led, experience-based co-design approach and implemented it in developing a prototype mobile app, Fitness4Surgery, to provide

guidance and support for patients with obesity to change lifestyle, lose weight, and improve fitness. The approach consisted of four iterative phases: understanding user needs, identifying theory, integrating theory into the design, and evaluating the prototype. In each phase, we engaged as many relevant stakeholders as possible for the consultation and gathered multiple sources of evidence from expert panel discussions, literature review, intervention mapping, and focus group discussions. Therefore, we adopted an evidence-based approach to design our mHealth service by drawing on experiences from clinicians, patients, researchers, and software developers.

To the best of our knowledge, this study is the first to articulate a detailed *co-design* approach that leverages the clinical experiences of clinicians and multidisciplinary teams to produce the initial prototype app. The prototype allowed the patients who participated in the focus group discussions to directly interact with the mobile app and experience its functions. This *hands-on* experience enabled them to draw on their needs and expectations for the mHealth app. The research output is useful for designing innovative digital interventions to provide just-in-time support for patients, which is low cost and easy to access [85]. This provides a useful alternative solution to address the service gap due to a shortage of funding and lack of human resources to provide these services face-to-face to vulnerable patients in the public health care system. mHealth services are also advantageous in the current period of the COVID-19 pandemic when social distancing is required [4,5]. Compared with similar studies published, our research contributes three distinct innovations to advance the design of mHealth apps (Table 4).

Table 4. A comparison of the contribution of this study and the existing literature.

Study	Aim	Design technique	Theory used	Theory mapping	User or clinician design and test
This study	To support patients with obesity to lose weight and improve fitness before surgery	Experience-based co-design	SCT[a]	Yes	Yes, expert panel discussion and focus groups
Smaradottir et al, 2020 [10]	To support chronic pain management	User-centered design	No	N/A[b]	Yes, a cocreation workshop
Wachtler et al, 2018 [8]	To improve treatment allocation for depression	User-centered design	Theory of agent-oriented modeling	No	Yes, two focus groups
Morita et al, 2019 [9]	To support asthma self-management	User-centered design	No	N/A	Yes, semistructured interviews
Duan et al, 2020 [3]	To improve patient compliance with hypertension self-management	Goal-directed design	HBM[c] and technology acceptance model	No	Yes, persona establishment (questionnaire and interview)
Fore et al, 2013 [12]	To support chronic care for pediatric inflammatory bowel disease	Goal-directed design	No	N/A	Yes, semistructured interview
Woods et al, 2019 [13]	To support heart failure self-management	Nurse-led co-design	No	N/A	Yes, interviews and workshops
Martin et al, 2020 [14]	To improve obesity-related health behaviors of adolescents	Co-design	Behavior change wheel, positive psychology, SDT[d], and nudging theory	No	Yes, workshop

[a]SCT: Social Cognitive Theory.

[b]N/A: not applicable.

[c]HBM: Health Belief Model.

[d]SDT: Self-determination Theory.

First, only a few of the mHealth developments for supporting behavioral change in recent years have reported the use of a specific theory [3,8,14]. Our research analyzed and compared common theories related to behavioral changes. In addition to guiding ideology at a high level, the theory-based design also involves the in-depth mining and analysis of all relevant constructs in theory and the mapping of the constructs to each user's needs to conceptualize a series of corresponding intervention techniques. These techniques were then converted to different real functionalities and were built into different function modules in the app, ensuring scientific rigor and practicality. The three proposed types of user needs are consistent with those addressed by similar mHealth interventions in alcohol and HIV areas [66,86]. Huygens et al [87] conducted a comprehensive qualitative focus group discussion with patients with chronic diseases to explore their expectations and needs for using mHealth for self-management purposes. Patients with obesity perceived need fulfillment and disease control as determinants of their willingness to use the app, which reflects the advantages of ecological momentary assessment and intervention [6,87,88].

Second, user interviews are traditionally conducted to gather user needs from the first stage [3,8-10,12-14], but in our study,

clinicians proposed patient needs as the first step toward app content and design. This approach had two advantages. The first advantage was that clinicians' suggestions based on scientific and public health reports and their years of field observation were practical and valuable in guiding the design of intervention content and delivery. The second advantage was that patients are often not fully aware of the scientific background of their health conditions [32]. Providing patients with a prototype for a trial and revising the app based on their feedback made our use of resources as effective and efficient as possible. This agrees with the concept of formative research in which health care researchers or practitioners identify a community of interest, describe the features of the community associated with a specific medical issue, and define the initial needs, which are then tested in the population of interest [89]. The qualitative data collected in this stage can provide rich insights into the use of mHealth technology and the most effective engagement strategies [90]. Although we did not deliberately pursue information saturation as part of our qualitative approach, the ideas gathered from different disciplines had many common and overlapping points that guided app development. Focus group interviews with the patients played an additional role. While suggesting several improvements, their attitude toward usefulness and ease of use of the app was positive, indicating the potential value of this app.

Patient feedback affirmed two indicators for measuring the level of acceptance in the technology acceptance model: usefulness and ease of use [91]. This reflects the scientific rigor of our design. The patients also indicated a limitation of the app, that is, insufficient information content and clarity, which is a known factor to affect the success of mHealth systems [32]. This reminded us about the importance of targeting information delivery to fit patients' health literacy. The main limitation of our prototype was a lack of personalization, which has been identified in previous studies [1,29,30].

Third, most current app-based interventions specifically target bariatric surgery [54,55]. Our research extends the scope of potentially effective mHealth interventions to any elective surgery, making the product much more generalizable to a wider audience.

Limitations and Future Work

This study had some limitations. First, participants in the focus groups were recruited via purposive sampling, so the diverse demographic groups were not evenly distributed. This could have led to a biased finding of the patient's level of acceptance and satisfaction with the mHealth app [92]. In the large scope trial, stratified sampling can be used to avoid this problem. Second, the technology is rapidly changing. The current version of the app is relatively simple, despite meeting our identified requirements. Further development is required to develop intelligent and personalized functions. Third, the reuse of lessons may be limited by the small scope of the study at one site. However, the health informatics experts in our team have extensive experience in the development and evaluation of eHealth solutions. Moreover, the app design is underpinned by a carefully selected theory based on sound literature research, leading to robust functionality. Therefore, the design process is useful for other similar learning initiatives.

Future research will be conducted to evaluate the effectiveness of the app, with measures including user satisfaction and perioperative efficiency and outcomes. The app also has the potential to be used postoperatively and preoperatively to provide ongoing motivation and resources to users. This mHealth platform may be particularly useful when face-to-face health care options are limited, such as in regional and remote communities, and during periods with social distancing restrictions. Moreover, the integration of the app-based system with existing electronic health records or tools used by clinicians in the health system could also be further investigated.

Conclusions

This study reports an innovative co-design approach with clinicians and patients to address the challenges facing participative co-design with patients' mHealth services that support their self-management of chronic conditions. It presents a detailed process to leverage the experiences of clinicians to produce the initial prototype app. *Hands-on* interaction with the prototype mHealth app in focus group discussions allowed the patients to effectively articulate their needs and expectations for the mHealth app. This research also presents a method to integrate theory into mHealth design, which addresses a missing link in the design of mHealth services that support the patient self-management of chronic conditions. The reported design approach can be generalized to the design of any mHealth services that aim to support the patient self-management of chronic conditions.

Acknowledgments

This research was supported by the 2018 Illawarra Health and Medical Research Institute Clinical Translation Grant Scheme. The authors also wish to thank Yunchuan Shi, Ziwen Zhou, Chunhao Li, Yuting Chen and Zhengshu Dai for developing the app, Mikaela Dawking for helping with the early nutrition messages, the research nurses, and the focus group participants.

Authors' Contributions

PY and NAS led this collaboration project, secured funding for the study, and guided the design and pilot of the intervention. NAS and PY conceived the concept of the mHealth intervention. NAS, VB, YP, and GEP designed the contents and rules of the push notifications as well as the built-in assessment questions in the app. PY, TS, and SQ conducted the review, identified the underlying theory for app development, and designed the conceptual functional model. MA, PP, and NPRH designed the architecture

of the app and the supporting back-end program. LH moderated the focus group discussion and provided expert insights into the app refinement. TS and NPRH analyzed the transcripts. TS and SQ managed the project. TS drafted the manuscript. PY, NAS, SQ, VB, YP, GEP, TC, MA, and LH critically commented on and extensively revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group questions.
[PDF File (Adobe PDF File), 100 KB - mhealth_v9i7e20650_app1.pdf]

Multimedia Appendix 2
Comparison of behavioral theories.
[PDF File (Adobe PDF File), 22 KB - mhealth_v9i7e20650_app2.pdf]

Multimedia Appendix 3
Workflow diagram of the prototype.
[PNG File , 346 KB - mhealth_v9i7e20650_app3.png]

References

1. Ghelani DP, Moran LJ, Johnson C, Mousa A, Naderpoor N. Mobile apps for weight management: a review of the latest evidence to inform practice. Front Endocrinol (Lausanne) 2020;11:412 [FREE Full text] [doi: 10.3389/fendo.2020.00412] [Medline: 32670197]
2. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
3. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
4. Vas P, Hopkins D, Feher M, Rubino F, Whyte M. Diabetes, obesity and COVID-19: a complex interplay. Diabetes Obes Metab 2020 Oct;22(10):1892-1896 [FREE Full text] [doi: 10.1111/dom.14134] [Medline: 32627299]
5. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
6. Forman EM, Schumacher LM, Crosby R, Manasse SM, Goldstein SP, Butryn ML, et al. Ecological momentary assessment of dietary lapses across behavioral weight loss treatment: characteristics, predictors, and relationships with weight change. Ann Behav Med 2017 Oct;51(5):741-753. [doi: 10.1007/s12160-017-9897-x] [Medline: 28281136]
7. Schumacher LM, Martin GJ, Goldstein SP, Manasse SM, Crosby RD, Butryn ML, et al. Ecological momentary assessment of self-attitudes in response to dietary lapses. Health Psychol 2018 Feb;37(2):148-152 [FREE Full text] [doi: 10.1037/hea0000565] [Medline: 29172606]
8. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
9. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Lavigne AS, et al. A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956 [FREE Full text] [doi: 10.2196/10956] [Medline: 30688654]
10. Smaradottir BF, Fagerlund AJ, Bellika JG. User-centred design of a mobile application for chronic pain management. Stud Health Technol Inform 2020 Jun 26;272:272-275. [doi: 10.3233/SHTI200547] [Medline: 32604654]
11. Lindgren H. Towards personalized decision support in the dementia domain based on clinical practice guidelines. User Model User-Adap Inter 2011 Jan 9;21(4-5):377-406. [doi: 10.1007/s11257-010-9090-4]
12. Fore D, Goldenhar LM, Margolis PA, Seid M. Using goal-directed design to create a novel system for improving chronic illness care. JMIR Res Protoc 2013 Oct 29;2(2):e43 [FREE Full text] [doi: 10.2196/resprot.2749] [Medline: 24168835]
13. Woods L, Duff J, Roehrer E, Walker K, Cummings E. Design of a consumer mobile health app for heart failure: findings from the nurse-led co-design of Care4myHeart. JMIR Nursing 2019 Sep 23;2(1):e14633. [doi: 10.2196/14633]
14. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]

15. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
16. Champion V, Skinner C. The health belief model. In: Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008:45-65.
17. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015 Sep;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
18. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
19. Darnton A. Practical Guide: An overview of behaviour change models and their uses. In: GSR Behavior Change Knowledge Review. London: Government Social Research Unit; 2008:1-43.
20. Nash M, Barnier A. The Oxford Handbook of Hypnosis: Theory, Research, and Practice. Oxford, United Kingdom: Oxford University Press; 2012:1-802.
21. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
22. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
23. Song T, Qian S, Cui T, Yu P. The use of theory in mobile health interventions for patient self-management of chronic diseases. Stud Health Technol Inform 2019 Aug 21;264:1982-1983. [doi: 10.3233/SHTI190745] [Medline: 31438439]
24. Sarcona A, Kovacs L, Wright J, Williams C. Differences in eating behavior, physical activity, and health-related lifestyle choices between users and nonusers of mobile health apps. Am J Health Promot 2017 Jul 11;48(5):298-305. [doi: 10.1080/19325037.2017.1335630]
25. Lieffers JR, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-239. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
26. 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]
27. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Dec 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
28. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
29. Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
30. Wang E, Abrahamson K, Liu PJ, Ahmed A. Can mobile technology improve weight loss in overweight adults? A systematic review. West J Nurs Res 2020 Sep;42(9):747-759. [doi: 10.1177/0193945919888224] [Medline: 31762402]
31. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
32. Nisha N, Iqbal M, Rifat A, Idrish S. Exploring the role of service quality and knowledge for mobile health services. Int J E-Busi Res 2016;12(2):45-64. [doi: 10.4018/IJEBR.2016040104]
33. Fothergill E, Guo J, Howard L, Kerns JC, Knuth ND, Brychta R, et al. Persistent metabolic adaptation 6 years after "The Biggest Loser" competition. Obesity (Silver Spring) 2016 Aug;24(8):1612-1619 [FREE Full text] [doi: 10.1002/oby.21538] [Medline: 27136388]
34. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
35. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015 Dec;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
36. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mHealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
37. National Health Survey: First Results, 2017-18. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2021-07-01]

38. Ri M, Aikou S, Seto Y. Obesity as a surgical risk factor. Ann Gastroenterol Surg 2018 Jan;2(1):13-21 [FREE Full text] [doi: 10.1002/ags3.12049] [Medline: 29863119]
39. Domi R, Laho H. Anesthetic challenges in the obese patient. J Anesth 2012 Oct;26(5):758-765. [doi: 10.1007/s00540-012-1408-4] [Medline: 22562644]
40. Tjeertes EK, Hoeks SE, Beks SB, Valentijn TM, Hoofwijk AG, Stolker RJ. Obesity--a risk factor for postoperative complications in general surgery? BMC Anesthesiol 2015 Jul 31;15:112 [FREE Full text] [doi: 10.1186/s12871-015-0096-7] [Medline: 26228844]
41. Bazurro S, Ball L, Pelosi P. Perioperative management of obese patient. Curr Opin Crit Care 2018 Dec;24(6):560-567. [doi: 10.1097/MCC.0000000000000555] [Medline: 30299311]
42. Admitted Patient Care 2017–18 - Australian Hospital Statistics. Canberra: Australian Institute of Health and Welfare; 2019:1-236.
43. Lemanu DP, Singh PP, MacCormick AD, Arroll B, Hill AG. Effect of preoperative exercise on cardiorespiratory function and recovery after surgery: a systematic review. World J Surg 2013 Apr;37(4):711-720. [doi: 10.1007/s00268-012-1886-4] [Medline: 23292047]
44. Tew G, Ayyash R, Durrand J, Danjoux G. Clinical guideline and recommendations on pre-operative exercise training in patients awaiting major non-cardiac surgery. Anaesthesia 2018 Jun;73(6):750-768 [FREE Full text] [doi: 10.1111/anae.14177] [Medline: 29330843]
45. Obesity and overweight. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-07-01]
46. Castelnuovo G, Pietrabissa G, Manzoni GM, Cattivelli R, Rossi A, Novelli M, et al. Cognitive behavioral therapy to aid weight loss in obese patients: current perspectives. Psychol Res Behav Manag 2017;10:165-173 [FREE Full text] [doi: 10.2147/PRBM.S113278] [Medline: 28652832]
47. Cui HW, Turney BW, Griffiths J. The preoperative assessment and optimization of patients undergoing major urological surgery. Curr Urol Rep 2017 Jul;18(7):54 [FREE Full text] [doi: 10.1007/s11934-017-0701-z] [Medline: 28589402]
48. Gao K, Yu P, Su J, He C, Liu L, Zhou Y, et al. Cardiopulmonary exercise testing screening and pre-operative pulmonary rehabilitation reduce postoperative complications and improve fast-track recovery after lung cancer surgery: a study for 342 cases. Thorac Cancer 2015 Jul;6(4):443-449 [FREE Full text] [doi: 10.1111/1759-7714.12199] [Medline: 26273399]
49. Ghanta RK, LaPar DJ, Zhang Q, Devarkonda V, Isbell JM, Yarboro LT, et al. Obesity increases risk-adjusted morbidity, mortality, and cost following cardiac surgery. J Am Heart Assoc 2017 Mar 08;6(3):e003831 [FREE Full text] [doi: 10.1161/JAHA.116.003831] [Medline: 28275064]
50. Teixeira PJ, Silva MN, Mata J, Palmeira AL, Markland D. Motivation, self-determination, and long-term weight control. Int J Behav Nutr Phys Act 2012;9:22 [FREE Full text] [doi: 10.1186/1479-5868-9-22] [Medline: 22385818]
51. Heymsfield SB, Wadden TA. Mechanisms, pathophysiology, and management of obesity. N Engl J Med 2017 Jan 19;376(3):254-266. [doi: 10.1056/NEJMra1514009] [Medline: 28099824]
52. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
53. Salwen-Deremer JK, Khan AS, Martin SS, Holloway BM, Coughlin JW. Incorporating health behavior theory into mhealth: an examination of weight loss, dietary, and physical activity interventions. J Technol Behav Sci 2019 Nov 04;5(1):51-60. [doi: 10.1007/s41347-019-00118-6]
54. Elvin-Walsh L, Ferguson M, Collins PF. Nutritional monitoring of patients post-bariatric surgery: implications for smartphone applications. J Hum Nutr Diet 2018 Feb;31(1):141-148. [doi: 10.1111/jhn.12492] [Medline: 28691740]
55. Connor K, Brady RR, Tulloh B, de BA. Smartphone applications (apps) for bariatric surgery. Obes Surg 2013 Oct;23(10):1669-1672. [doi: 10.1007/s11695-013-0950-y] [Medline: 23609330]
56. EBCD: Experience-based co-design toolkit. The Point of Care Foundation. URL: https://www.pointofcarefoundation.org.uk/resource/experience-based-co-design-ebcd-toolkit/step-by-step-guide/1-experience-based-co-design/ [accessed 2021-07-01]
57. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
58. Lawson H, Caringi J, Pyles L, Jurkowski J, Bozlak C. Participatory Action Research - Pocket Guide to Social Work Research Methods. Oxford, United Kingdom: Oxford University Press; 2015:1-224.
59. Mulvale A, Miatello A, Hackett C, Mulvale G. Applying experience-based co-design with vulnerable populations: lessons from a systematic review of methods to involve patients, families and service providers in child and youth mental health service improvement. Patient Exp J 2016 Apr 28;3(1):117-129. [doi: 10.35680/2372-0247.1104]
60. Fernandez ME, Ruiter RA, Markham CM, Kok G. Intervention Mapping: Theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]

61. National Health Survey: First Results, 2014-15. Australian Bureau of Statistics. 2016. URL: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CDA852A349B4CEE6CA257F150009FC53/$File/national%20health%20survey%20first%20results,%202014-15.pdf [accessed 2021-07-01]
62. Freckelton L, Lambert K, Smith NA, Westley-Wise V, Lago L, Mullan J. Impact of body mass index on utilization of selected hospital resources for four common surgical procedures. ANZ J Surg 2019 Jul;89(7-8):842-847. [doi: 10.1111/ans.15085] [Medline: 30974502]
63. Kaushik V, Gupta K, Gupta D. React native application development. Int J Adv Stud Scient Res 2019;4(1):1-7 [FREE Full text]
64. 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]
65. Velapati SR, Shah M, Kuchkuntla AR, Abu-Dayyeh B, Grothe K, Hurt RT, et al. Weight regain after bariatric surgery: prevalence, etiology, and treatment. Curr Nutr Rep 2018 Dec;7(4):329-334. [doi: 10.1007/s13668-018-0243-0] [Medline: 30168043]
66. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
67. Song T, Yu P. mHealth. In: Gu D, Dupre ME, editors. Encyclopedia of Gerontology and Population Aging. Switzerland: Springer International Publishing; 2020:1-7.
68. Spring B, Champion KE, Acabchuk R, Hennessy EA. Self-regulatory behaviour change techniques in interventions to promote healthy eating, physical activity, or weight loss: a meta-review. Health Psychol Rev 2020 Feb 17:1-32. [doi: 10.1080/17437199.2020.1721310] [Medline: 31973666]
69. Flocke SA, Clark E, Antognoli E, Mason MJ, Lawson PJ, Smith S, et al. Teachable moments for health behavior change and intermediate patient outcomes. Patient Educ Couns 2014 Jul;96(1):43-49 [FREE Full text] [doi: 10.1016/j.pec.2014.03.014] [Medline: 24856449]
70. Pétré B, Margat A, Servotte JC, Guillaume M, Gagnayre R, Ghuysen A. Patient education in the emergency department: take advantage of the teachable moment. Adv Health Sci Educ Theory Pract 2020 May;25(2):511-517. [doi: 10.1007/s10459-019-09893-6] [Medline: 31028515]
71. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. Am J Health Promot 1996;11(2):87-98. [doi: 10.4278/0890-1171-11.2.87] [Medline: 10163601]
72. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
73. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
74. Bracht N, Kingsbury L. Community organization principles in health promotion: a five-stage model. In: Bracht N, editor. Health Promotion at the Community Level. Thousand Oaks, California, United States: Sage Publications Inc; 1990:66-88.
75. Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009 Sep 1;19(5):503-518 [FREE Full text] [doi: 10.1177/1049731509335569] [Medline: 20976022]
76. Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, New Jersey, United States: John Wiley & Sons; 2008:1-592.
77. DiClemente CC. Addiction and Change: How Addictions Develop and Addicted People Recover. New York, United States: Guilford Press; 2018:1-364.
78. Lohrmann DK. A complementary ecological model of the coordinated school health program. J Sch Health 2010 Jan;80(1):1-9. [doi: 10.1111/j.1746-1561.2009.00460.x] [Medline: 20051084]
79. Brownie S, Muggleston H, Oliver C. The 2013 Australian dietary guidelines and recommendations for older Australians. Aust Fam Physician 2015 May;44(5):311-315 [FREE Full text] [Medline: 26042404]
80. Physical activity and exercise guidelines for all Australians. Australian Government - Department of Health. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines [accessed 2021-07-01]
81. Kulinski K, Smith NA. Surgical prehabilitation using mobile health coaching in patients with obesity: a pilot study. Anaesth Intensive Care 2020 Sep;48(5):373-380. [doi: 10.1177/0310057X20947731] [Medline: 33104444]
82. National Heart Foundation of Australia. URL: https://www.heartfoundation.org.au/ [accessed 2021-07-08]
83. Get Healthy. NSW Health Service. URL: https://www.gethealthynsw.com.au/ [accessed 2021-07-08]
84. Ryder JR, Kaizer AM, Jenkins TM, Kelly AS, Inge TH, Shaibi GQ. Heterogeneity in response to treatment of adolescents with severe obesity: the need for precision obesity medicine. Obesity (Silver Spring) 2019 Feb;27(2):288-294 [FREE Full text] [doi: 10.1002/oby.22369] [Medline: 30677258]
85. Goldstein SP, Thomas JG, Foster GD, Turner-McGrievy G, Butryn ML, Herbert JD, et al. Refining an algorithm-powered just-in-time adaptive weight control intervention: a randomized controlled trial evaluating model performance and behavioral outcomes. Health Informatics J 2020 Dec;26(4):2315-2331 [FREE Full text] [doi: 10.1177/1460458220902330] [Medline: 32026745]

86. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A mobile gaming intervention to increase adherence to antiretroviral treatment for youth living with HIV: development guided by the information, motivation, and behavioral skills model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e96 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
87. Huygens MW, Vermeulen J, Swinkels IC, Friele RD, van Schayck OC, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 08;16:232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
88. Boh B, Lemmens LH, Jansen A, Nederkoorn C, Kerkhofs V, Spanakis G, et al. An Ecological Momentary Intervention for weight loss and healthy eating via smartphone and internet: study protocol for a randomised controlled trial. Trials 2016 Mar 22;17:154 [FREE Full text] [doi: 10.1186/s13063-016-1280-x] [Medline: 27000058]
89. Beran D, Lazo-Porras M, Cardenas MK, Chappuis F, Damasceno A, Jha N, et al. Moving from formative research to co-creation of interventions: insights from a community health system project in Mozambique, Nepal and Peru. BMJ Glob Health 2018;3(6):e001183 [FREE Full text] [doi: 10.1136/bmjgh-2018-001183] [Medline: 30498592]
90. White BK, Martin A, White JA, Burns SK, Maycock BR, Giglia RC, et al. Theory-based design and development of a socially connected, gamified mobile app for men about breastfeeding (Milk Man). JMIR Mhealth Uhealth 2016 Jun 27;4(2):e81 [FREE Full text] [doi: 10.2196/mhealth.5652] [Medline: 27349756]
91. Rahimi B, Nadri H, Afshar HL, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Dec;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
92. Keane R, Cormican K, Flynn P. User acceptance of mandated technology: an analysis of the role of moderating factors. Procedia Comput Sci 2019;164:72-78. [doi: 10.1016/j.procs.2019.12.156]

Abbreviations

HBM: Health Belief Model
mHealth: mobile health
SCT: Social Cognitive Theory

Edited by L Buis; submitted 27.05.20; peer-reviewed by R McKenna, J Chen, N Deng, MDG Pimentel; comments to author 25.09.20; revised version received 09.11.20; accepted 17.05.21; published 20.07.21.

Please cite as:
Song T, Yu P, Bliokas V, Probst Y, Peoples GE, Qian S, Houston L, Perez P, Amirghasemi M, Cui T, Hitige NPR, Smith NA
A Clinician-Led, Experience-Based Co-Design Approach for Developing mHealth Services to Support the Patient Self-management of Chronic Conditions: Development Study and Design Case
JMIR Mhealth Uhealth 2021;9(7):e20650
URL: https://mhealth.jmir.org/2021/7/e20650
doi:10.2196/20650
PMID:34283030

Original Paper

Evaluating Chinese Mobile Health Apps for Ankylosing Spondylitis Management: Systematic App Search

Yuqing Song[1], MSN; Hong Chen[1], PhD

West China School of Nursing/West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Hong Chen, PhD
West China School of Nursing/West China Hospital
Sichuan University
No 37, Guoxuexiang
Wuhou District
Chengdu, 610041
China
Phone: 86 18980601733
Email: 1366109878@qq.com

Abstract

Background: Patients with ankylosing spondylitis (AS), a chronic systematic inflammatory disease, require long-term treatment and management. Mobile health (mHealth) apps can deliver health services through mobile devices, facilitate long-term disease management, support patient–health care provider communication, and enable patients to engage in disease management. There are some apps targeted at patients with AS, but the feature and quality of these apps have not been systematically examined.

Objective: The aim of this study was to identify existing, publicly available Chinese mHealth apps for AS management and to evaluate their features and quality.

Methods: We systematically searched potential apps for AS management on the Apple and Huawei App Stores, using 4 search terms: *ankylosing spondylitis*, *spondyloarthritis*, *rheumatic disease*, and *arthritis*. Apps were included if they were in the Chinese language, targeted at patients with AS, could be downloaded and run on Android and/or iOS operating systems, and incorporated elements of disease management and/or patient education. We excluded apps that were not for patient use, not relevant to AS, or had not been updated since 2018. Apps that met the inclusion criteria were downloaded for final analysis. We formulated a list of app quality measures from and consistent with international guidelines for mHealth apps and AS management to evaluate the features and quality of the included app. The user version of the Mobile App Rating Scale (uMARS) was also used to rate the apps' quality.

Results: Of the 354 apps screened, 5 met the inclusion criteria and were included in our analysis. All apps were free, and most apps (4/5, 80%) had a privacy policy. Of the 5 apps, 1 (20%) involved medical professionals in the development process, 2 (40%) were developed by companies, and 2 (40%) were developed by medical institutions. All apps provided educational information about AS. Around half of the apps had functions like a basic information record (ie, users can input gender, age, disease history, etc) (n=3, 60%), patient–health care provider (and patient-patient) communication (n=2, 40%), symptom tracking (n=2, 40%), and information sharing (n=3, 60%). Only 1 (20%) app provided comprehensive functions that adhered to international guidelines for AS management and mHealth apps. The overall uMARS scores ranged from 2.7 to 4.2; only 1 app, with an overall uMARS score of 4.2, was considered as a high-quality app.

Conclusions: Most apps lacked comprehensive functions for AS management. One high-quality app provided comprehensive functions to help patients manage their conditions. This study assessed and summarized the features and quality of the apps but did not evaluate their efficacy. Future studies should evaluate the feasibility and efficacy of these apps. International guidelines and regulations for the design, development, validation, and implementation of mHealth apps are needed in the future. Meanwhile, health care providers, patients with AS, and app developers should collaborate to develop high-quality, evidence-based apps that take into account patients' needs and health care professionals' perspectives.

(JMIR Mhealth Uhealth 2021;9(7):e27234) doi:10.2196/27234

KEYWORDS

ankylosing spondylitis; app; eHealth; mHealth; smartphone; mobile phone

Introduction

Ankylosing spondylitis (AS) is a chronic systematic inflammatory disease that often causes structural impairment, functional disability, and impaired quality of life [1,2]. AS often affects young people at their most productive age (approximately 20-35 years) and has a lifelong impact on their life [1]. AS can interfere with work and schooling, and can impose substantial physical and social burdens on patients [3]. The guidelines of the European League Against Rheumatism (EULAR) recommend that AS requires long-term management, including pharmacological and nonpharmacological treatment, to control inflammation, prevent structural damage, and optimize function and quality of life [4]. The management of long-term conditions requires timely health care and treatment [5].

The traditional approach to AS management requires face-to-face rheumatology clinic appointments and treatment [6,7]. Patients with chronic diseases may have difficulty in attending regular rheumatology clinic appointments and obtaining timely treatment due to transportation difficulties, physical limitations, time constraints, and geographical barriers [8,9]. Moreover, China has a large population, with a broad geographical distribution of patients with AS; around 90% rheumatologists work at tertiary hospitals [10-12]. The barriers to receiving timely treatment and health care are much more prominent in China [11,13].

Mobile health (mHealth) has the potential to facilitate long-term disease management and deliver timely health care to hard-to-reach populations [14,15]. mHealth is defined as medical and public health practice supported by mobile devices, such as mobile phones, personal digital assistants, and other wireless devices [16,17]. mHealth apps can overcome barriers of time and geography to deliver health services through mobile devices and increase access to health care service [18]. Mobile apps offer a solution for patients with chronic diseases to monitor chronic conditions, support patient–health care provider communication, provide health advice, and enable patients to engage in disease management [10]. Many mHealth apps have been developed to help individuals with chronic diseases, such as asthma, rheumatoid arthritis, and inflammatory bowel disease, to manage their conditions [16,19-21]. Prior work has supported the usefulness of mHealth apps for enhancing disease management and improving clinical outcomes among patients with chronic diseases [15,22].

In China, 932 million people used smartphones to access the internet, and 3.59 million mobile apps were available as of 2020 [23]. Thus, mHealth interventions are increasingly accessible for Chinese people [10]. In the area of rheumatology, many studies revealed that mHealth apps for patients with rheumatoid arthritis have the potential to monitor their disease and support high-quality medical care [14,16,20,24]. Patients with AS may be one of the best target users of mHealth apps because AS affects young people (approximately 20-35 years) [10]. Young people are more likely to be acceptable to receiving an mHealth intervention [25,26].

Limited published works in the literature have focused on mHealth apps for AS management in China [7,27]. Although the Smart System of Disease Management (SSDM) was used for AS management, the characteristics and functions of the SSDM have not been clearly described in the literature [7,27]. Our search of app stores found that there are some publicly available apps for AS management despite limited published works. However, there is little information to users on which apps provide evidence-based tools [20]. A previous review revealed that many apps are not supported by research, do not provide evidence-based therapies, and do not follow clinical guidelines [28]. Kwan et al [29] reviewed English-language apps for monitoring disease activity in patients with spondyloarthritis and found only 2 high-quality apps. However, the characteristics and functionalities of Chinese mHealth apps for AS management have not been systematically evaluated (eg, in terms of star rating, privacy policy, evidence-based content, and functionality).

The purpose of this study was to identify existing, publicly available apps targeted at patients with AS for disease management, formulate a list of app assessment measures from and consistent with evidence-based guidelines to evaluate the features and quality of these apps, and rate app quality using the user version of the Mobile App Rating Scale (uMARS). This will help patients choose a high-quality app for managing their disease and identify gaps in the current apps available for AS management.

Methods

App Selection

We conducted a systematic search of all potential apps targeted at patients with AS in November 2020. Android and iOS are the two most popular smartphone operating systems among Chinese smartphone users [30]. The Huawei App Store is one of the biggest Android app stores in China [31]. Thus, we searched for iOS apps in the Apple App Store and Android apps in the Huawei App Store. Preliminary test searches were conducted to determine the search terms before the final search. To ensure all potential apps were screened, our final search terms included *ankylosing spondylitis* or *spondyloarthritis* or *rheumatic disease* or *arthritis* in both the Huawei and Apple app stores. The final search was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews. Two reviewers independently downloaded potential apps to test devices for screening according to the inclusion and exclusion criteria. The inclusion criteria for the apps were (1) targeted at patients with AS; (2) in the Chinese language; (3) available for downloading from the Huawei and/or Apple App Stores; (4) able to run on Android and/or iOS operating systems; (4) incorporated at least one of the following elements of disease management and/or patient education: educational information, medication management, exercise management, psychological strategy, symptom management, etc. Apps were excluded if they were not for patient use or not relevant to AS (eg, apps for other chronic conditions or other use). We also excluded apps that had not been updated since 2018. Android apps were downloaded and tested using the Huawei ATH-AL00 with

Android version 5.1.1, and iOS apps were downloaded and tested using the iPhone 11 with iOS 13.3.1 installed.

App Quality Measures

We systematically searched guidelines and recommendations for mHealth apps and AS management to identify app quality measures [3,4,32-35]. We formulated a list of quality assessment measures from and consistent with guidelines from the EULAR [4,32-34], the American College of Rheumatology [3], and the French Society for Rheumatology [35]. Then, rheumatologists and rheumatology nurses modified the app quality assessment measures. Finally, the list of app quality measures included two domains: basic characteristics and functionalities. Basic characteristics related to the following features extracted: app name, developer, operating system (iOS or Android), version, provider involvement, star rating (out of 5), number of reviews, number of downloads (only available for Android apps), cost, and privacy policy. Functionalities of the included apps were analyzed as follows: basic information record, educational information, communication, symptom tracking, general or psychological health tracking, medication management, visuals or analysis, exercise management, reminder feature, and information sharing. The app quality measures are shown in Table 1.

Table 1. App quality measures of apps for patients with ankylosing spondylitis.

Assessment measure	Description and definition
Basic characteristics	
App name	App name as shown in the Huawei and/or Apple app stores
Developer	Name of the developer (ie, who developed and uploaded the app)
Operating system	iOS and/or Android operating systems
Version	Latest update version and date of update
Provider involvement	Involvement of relevant health care providers in the design, development, and validation of the app
Star rating	Star rating score (out of 5) that users left on the Huawei and/or Apple app stores
Number of reviews	Number of reviews that users left on the Huawei and/or Apple app stores
Number of downloads	Number of downloads since the app released
Cost	Free apps or the cost of apps
Privacy policy	Information on how user data are stored and shared
Functionalities	
Basic information record	Enables users to record their basic information (eg, gender, age, disease history)
Educational information	The information content is up to date, scientifically justifiable, acceptable to users, and evidence-based. The content includes disease overview, pathogenesis, treatment goal and options, exercise advice, medication, joint protection, and health advice for daily life
Communication	Facilitates patient–health care provider communication and patient-patient communication
Symptom tracking	Prompts users to assess their general symptoms as follows: disease activity, pain, fatigue, morning stiffness, and functional ability
General or psychological health tracking	Allows users to record information about their general or psychological health, such as sleep quality, depression, anxiety, quality of life, and general health
Medication management	Allows users to record medication name, dosing, time, and frequency
Visuals or analysis	Displays recorded information as graphs or tables
Exercise management	Allows users to record information pertaining to exercise (eg, frequency, time, type)
Reminders	Allows users to set reminders for appointments or when to take their medication
Information sharing	Allows users to share educational information and/or their disease data with health care providers or others

App Rating Using the uMARS

We also used the uMARS to evaluate the quality of apps [36], which is a simplified version of the MARS that has been used to assess the quality of mHealth apps [36,37]. The uMARS is a 20-item measure that comprises 4 objective quality subscales (engagement, functionality, aesthetics, information quality) and 1 subjective quality subscale [36]. We did not include the latter subscale in this study since our aim was to assess the objective quality of apps. All items were rated on a 5-point Likert scale from 1 (inadequate) to 5 (excellent). Mean scores were calculated for each subscale and a mean total score was calculated across all 4 subscales [36]. Apps that scored ≥3 out of 5 on the uMARS were considered to be of acceptable quality, and scores higher than 4 were rated as high quality [38]. Two trained reviewers independently rated the apps using the

uMARS. Subsequently, they discussed inconsistencies and doubts about the apps and reached a consensus on the final uMARS scores.

Data Abstraction

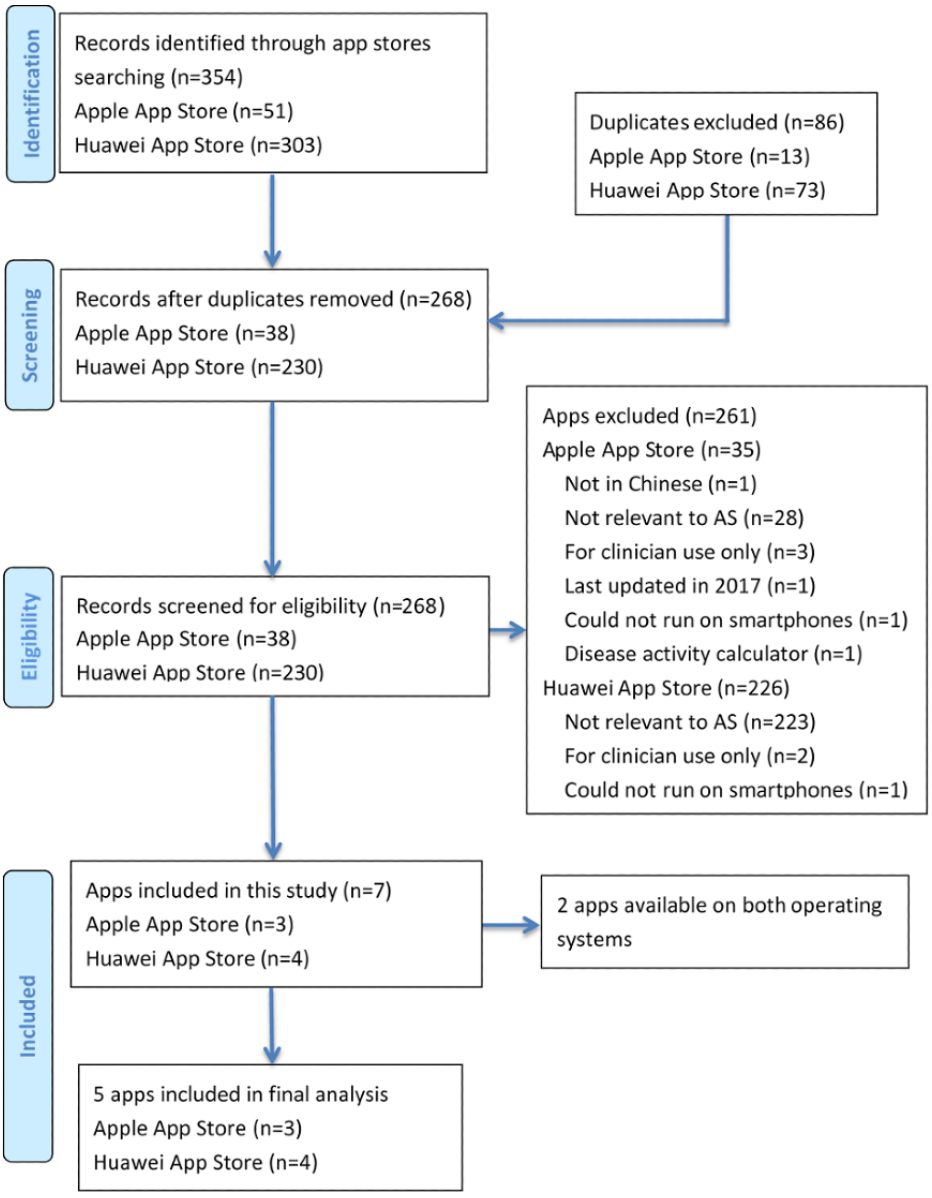
Two reviewers independently downloaded the apps that met the inclusion criteria and evaluated them using our list of quality assessment measures. In addition, they rated the app quality using the uMARS. The main features and quality of all apps were recorded, including basic characteristics, functional features, and uMARS rating.

Results

App Selection

Our searches from the Apple App Store retrieved 51 apps. We excluded apps that were duplicates (n=13), not in Chinese (n=1), not relevant to AS (eg, many apps focused on other chronic conditions or provided general health information) (n=28), only for clinician use (n=3), disease activity calculators (n=1), or last updated in 2017 (n=1). Of the remaining 4 apps, 1 app was not able to run on the iOS operating system and was excluded from the study, leaving 3 apps for the final analysis. A total of 303 Android apps were retrieved from the Huawei App Store. Of these, 299 apps were excluded because they were duplicates (n=73), not relevant to AS (eg, many apps focused on other chronic conditions, or were intended for other use including online shopping, work, and games) (n=223), only for clinician use (n=2), or could not run on the Android operating system (n=1). Four Android apps were included for analysis. A total of 5 apps were included in the final analysis since 2 apps were available on both the Android and iOS operating systems (Figure 1). Since the 2 apps had the same functionalities on both the iOS and Android operating systems, they were described once. The 5 apps were iRheuma, Wen Wen Feng Shi, Feng Shi Mian Yi Kou Dai Shu, Lei Feng Shi Hu Zhu, and Jian Feng Yuan. iRheuma is an app by the SSDM, which is part of a series of doctor-patient interactive apps developed for the self-management of patients with chronic diseases.

Figure 1. Flow diagram for the systematic search and selection of apps from the Apple and Huawei app stores. AS: ankylosing spondylitis.



App Characteristics

The basic characteristics of the app, based on the app quality measures, are summarized in Tables 2 and 3. All apps were specific to people with AS, but some apps were also relevant to other rheumatic diseases, such as rheumatoid arthritis and gout. Two apps (40%) were developed by companies, and 2 apps (40%) by medical institutions. Four apps (80%) were last updated in 2020, and 1 app was updated in 2018. Only 1 app, Feng Shi Mian Yi Kou Dai Shu, described medical professionals involvement in the development of this app. However, we did not know if patients with AS and/or health care providers were involved in the development of the other apps since these apps did not provide information about their design and development. Two app (40%) received no review, and there was only 1 app (20%) with more than 10 app store reviews. Of the 4 Android apps with available download data, 3 (60%) were downloaded more than 10,000 times, and 1 (20%) was downloaded 450,000 times. All apps were free for users, and most apps (n=4, 80%) had a privacy policy.

Table 2. Operating system, developer, version, and health care provider involvement of the included apps.

App name	Operating system	Developer	Date of latest update	iOS version	Android version	Provider involvement
iRheuma	iOS, Android	Shanghai Gete Internet Technology Co, Ltd	Dec 2020	3.12.4	3.12.4	N/A[a]
Wen Wen Feng Shi	iOS, Android	Shanghai Wenyun Biotechnology Co, Ltd	Dec 2020	3.2.4	3.2.3	N/A
Feng Shi Mian Yi Kou Dai Shu	iOS	St. Bear Inc	May 2020	3.1	—[b]	Yes
Lei Feng Shi Hu Zhu	Android	Hu Ze Min Rheumatoid Hospital of TCM	Jul 2020	—	3.8.7	N/A
Jian Feng Yuan	Android	Not stated	Dec 2018	—	2.1	N/A

[a]N/A: not applicable.

[b]Not available.

Table 3. Star rating, number of reviews and downloads, cost, and privacy policy of the included apps.

App name	Star rating	Reviews, n	Downloads, n	Cost	Privacy policy
iRheuma	4.2 (iOS); 4.3 (Android)	54 (iOS);10 (Android)	450,000	Free	Yes
Wen Wen Feng Shi	4.4 (iOS); 5 (Android)	28 (iOS); 1 (Android)	90,000	Free	Yes
Feng Shi Mian Yi Kou Dai Shu	N/A[a]	0	N/A	Free	Yes
Lei Feng Shi Hu Zhu	3.7	2	20,000	Free	Yes
Jian Feng Yuan	N/A	0	<10,000	Free	No

[a]N/A: not applicable.

App Functionalities

The functionalities of the included apps, based on the app quality measures, are shown in Table 4. All apps provided educational information about AS, including disease overview, pathogenesis and cause, symptoms, pharmacological treatment, exercise, and health advice, but none were to tailored individual needs and preferences. Four apps (80%) updated their educational information. Of the 5 included apps, only Wen Wen Feng Shi provided educational materials, including the pathogenesis, cause and treatment of AS, health advice on exercise and lifestyle, and updated medical news.

Table 4. Functionalities of the included apps.

Functionality	iRheuma	Wen Wen Feng Shi	Feng Shi Mian Yi Kou Dai Shu	Lei Feng Shi Hu Zhu	Jian Feng Yuan
Educational information	✓[a]	✓	✓	✓	✓
Basic information record	✓		✓	✓	
Communication	✓				✓
Symptom tracking	✓				✓
General or psychological health tracking	✓				
Medication management	✓				
Laboratory result tracking	✓				
Visuals or analysis	✓				
Exercise management					
Reminders	✓				
Information sharing	✓	✓		✓	

[a]Check mark indicates presence of feature.

Three apps (60%) allowed users to record their basic information, such as gender, age, health history, and medical record. Two apps (40%) had a communication feature. iRheuma enabled users to directly consult doctors through text message, telephone call, and video call, but the communication function did not run well because many doctors did not provide online consultation services. Jian Feng Yuan only supported patient-patient communication via a message board.

Symptom tracking was available for 2 apps (40%): iRheuma and Jian Feng Yuan. iRheuma assessed users' disease activity and functional ability using validated instrument (ie, Ankylosing Spondylitis Disease Activity Score [ASDAS]; Bath Ankylosing Spondylitis Disease Activity Index [BASDAI]; Bath Ankylosing Spondylitis Functional Index [BASFI]). If patients inputted their data, iRheuma could give patients feedback on disease activity and function, and provide individual health advice. Jian Feng Yuan used only the ASDAS to assess users' disease activity and calculated an ASDAS score.

iRheuma also provided general or psychological health tracking, medication management, laboratory result tracking, visuals or analysis, and reminders. iRheuma used validated instruments (Hospital Anxiety and Depression Scale, Pittsburgh Sleep Quality Index, Medical Outcomes Study Short Form 36-item Health Survey) to assess patients' health and to give them feedback and tailored health advice. iRheuma enabled users to record their medication and laboratory results, and to set reminders to take their medication. This app could create graphs from user-reported data to track users' symptoms and laboratory results.

No app had an exercise management feature. Three apps (60%) supported an information-sharing function, but they only allowed users to share health information with other people via WeChat, Tencent QQ, and email. Some functions of the included apps did not run well due to technical issues. Among these apps, iRheuma was downloaded 450,000 times more than other apps.

Additional Functionalities

Two apps (40%) supported online shopping. Users could purchase medications, medical devices, and medical books on these apps. One app provided questions about the Modified New York Classification Criteria for AS [39] to screen for AS. Two apps (40%) enabled users to make appointments with doctors. Most apps (4/5, 80%) linked to a WeChat public account with content and functions similar that in the app.

App Rating Based on the uMARS

Table 5 shows the uMARS ratings for all included apps. The overall uMARS scores for the apps ranged from 2.7 to 4.2. Three (60%) apps scored more than 3 out of 5; only iRheuma (4.2) app scored than 4 out of 5. Information quality scores (2.8-4.5) showed the greatest variability. The engagement scores (2.2-3.8) were the lowest of the 4 subscales.

Table 5. The user version of the Mobile App Rating Scale (uMARS) scores of the included apps.

App name	uMARS subscale score				Overall score
	Engagement	Functionality	Aesthetics	Information quality	
iRheuma					
iOS	3.8	4.3	4.3	4.5	4.2
Android	3.8	4.3	4.3	4.5	4.2
Wen Wen Feng Shi					
iOS	2.6	3.5	3.7	3.5	3.3
Android	2.6	3.5	3.7	3.5	3.3
Feng Shi Mian Yi Kou Dai Shu					
iOS	2.4	3.3	3.0	2.8	2.9
Lei Feng Shi Hu Zhu					
Android	2.2	3.0	2.7	2.8	2.7
Jian Feng Yuan					
Android	2.8	3.0	3.7	3.3	3.2

Discussion

Principal Findings

This study utilized a systematic approach to identify and evaluate 5 mHealth apps for AS management in China. We found only 1 app (with an overall uMARS score of 4.2) that provided comprehensive functions that adhered to evidence-based guidelines for AS management. This result was line with previous reviews of apps on gout and rheumatoid arthritis [20,40]. Most apps in the study did not provide further information regarding the development process of the apps. Similarly, Najm et al [18] also found that the development process of most apps were not sufficient or were not described in the existing literature, which might raise questions about apps' credibility. Thus, international guidelines and regulations for the design, development, and validation of mHealth apps are needed in the future. Additionally, our results were in line with previous evidence that most apps designed for patient use did not involve health care providers or patients in the development stage [18]. Prior evidence suggested that the development and validation of self-management apps should involve target users and health care providers [32,41].

The EULAR has recommended that mHealth apps should be relevant and tailored to the individual needs of people with rheumatic and musculoskeletal disease [32,34]. In our study, educational information in most of the apps was not tailored to patients' needs and preferences. It may be that patients and health care professionals were not involved in the development of the apps [18]. Moreover, the educational information in some apps was not divided into different disease modules, which may make it difficult for users to find information on AS. We did not systematically evaluate patient acceptability of these apps due to lack of a quantitative assessment. Future mobile apps should provide evidence-based educational information and increase apps' usability based on patients' needs and health care professionals' perspectives.

Many symptom-tracking apps for rheumatic disease did not use validated instruments [16,29]; only 2 apps in this study used validated instruments to track patients' symptoms. iRheuma used the ASDAS, BASDAI, and BASFI to track users' disease activity and functional ability. Moreover, iRheuma was also able to provide feedback on users' conditions and encourage users to consult with doctors, which may enable users to better understand their conditions and monitor their disease. Another app, Jian Feng Yuan, could calculate users' disease activity (using the ASDAS). Both apps lacked the ability to directly transmit data to health care providers. Although evidence has suggested that mobile apps that use a validated instrument and have a tracking function may be useful for patients with arthritis for symptom monitoring [16,29], most of these apps, as well as the ones from our study, were not assessed in clinical trials. Thus, future studies should explore the effects of these apps on health and economic outcomes.

In this study, 1 app had a reminder feature. Prior work has suggested that apps with such a function may be effective in improving medication adherence in nonadherent patients [42]. It is important to integrate this function into clinical practice. However, the feasibility and effects of reminder apps have not been well studied. Future studies should explore the efficacy of reminder apps in large populations.

Of the included 5 apps, only iRheuma provided comprehensive functions, including direct communication, health tracking, medication management, and laboratory results tracking. These functions may help patients better manage their conditions. Additionally, we found that iRheuma was the most frequently downloaded app of the 5 apps, indicating that multifunction apps may be attractive for patients with AS. Luo et al [20] revealed that only 25% of reviewed apps provided symptom tracking and education for patients about management strategies [20]. Although international guidelines have revealed the importance of exercise for AS management [3,43,44], none of the apps provided this function. Future studies should develop comprehensive, evidence-based apps for patients with AS.

The overall uMARS scores ranged from 2.7 to 4.2, indicating inconsistent quality among the included apps. This result is similar to previous studies reviewing mHealth apps [29,45]. Only 1 app (iRheuma) provided comprehensive functions for AS management and was considered high quality based on the uMARS. The 2 highest scoring apps, iRheuma (4.2) and Wen Wen Feng Shi (3.3), were most frequently downloaded, indicating that the high-quality apps may be useful for targeted users. Engagement scores were lower compared to the other 3 subscales, and nearly half (2/5, 40%) of the apps were considered to be of poor quality. Thus, future studies and practice should improve apps' quality for AS management, especially engagement quality.

Strengths and Limitations

Strengths of this study included the assessment of a broad range of app characteristics, including basic characteristics, functionalities, and app ratings using uMARS scores. This study also had several limitations. One limitation is that we only focused on Chinese-language apps available on the two most popular operating systems (Android and iOS), and thus missed apps available in other languages or on other operating systems. Although WeChat public accounts have been increasingly used as mHealth tools in China [46], we did not include WeChat public accounts for AS management, such as the Smart-phone SpondyloArthritis Management System [11]. We searched for apps in two of the most popular app stores (the Huawei and Apple App Stores) for Android and iOS operating systems in China, but apps exclusively in other app stores (eg, Microsoft) were not included in this study. This study also only focused on publicly available apps. Apps available in the published literature were not included since limited published works have focused on apps for AS management. The uMARS does not focus on AS management apps. Thus, we developed a list of app quality assessment measures based on evidence-based guidelines and recommendations for mHealth apps and AS management. The app quality measures have not been validated, which may limit the findings of this study.

Conclusions

This study found a lack in high-quality apps available to assist in the management of AS in China. Only 1 out of the 5 apps was of high quality and provided comprehensive functions to help patients manage their conditions. Most apps lacked key features for disease management, such as symptom tracking, medication management, and reminders. This study only assessed the app quality and did not evaluate the usability and efficacy of the included apps. Future studies and clinical practice should explore the efficacy and feasibility of mHealth apps. International guidelines and regulations for the design, development, validation, and implementation of mHealth apps are also needed in the future. In the meantime, health care providers, patients with AS, and app developers should collaborate to develop high-quality, evidence-based apps that consider patients' needs and health care professionals' perspectives.

Acknowledgments

The authors thank Ms Yao Yin for her valuable input, as well as the academic editor, peer reviewers, and copyeditor, who supported this study. The authors also thank their research assistants for their assistance.

Conflicts of Interest

None declared.

References

1. Braun J, Sieper J. Ankylosing spondylitis. Lancet 2007 Apr 21;369(9570):1379-1390. [doi: 10.1016/S0140-6736(07)60635-7] [Medline: 17448825]
2. Kotsis K, Voulgari PV, Drosos AA, Carvalho AF, Hyphantis T. Health-related quality of life in patients with ankylosing spondylitis: a comprehensive review. Expert Rev Pharmacoecon Outcomes Res 2014 Dec;14(6):857-872. [doi: 10.1586/14737167.2014.957679] [Medline: 25193010]
3. Ward MM, Deodhar A, Gensler LS, Dubreuil M, Yu D, Khan MA, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Recommendations for the Treatment of Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis. Arthritis Rheumatol 2019 Oct;71(10):1599-1613. [doi: 10.1002/art.41042] [Medline: 31436036]
4. van der Heijde D, Ramiro S, Landewé R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017 Jun;76(6):978-991. [doi: 10.1136/annrheumdis-2016-210770] [Medline: 28087505]
5. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
6. Liu Y, Yang H. The value of chronic disease management on rheumatic disease. Medical Journal of West China 2019;31(11):1641-1642, 1645. [doi: 10.3969/J.issn.1672-3511.2019.11.001]
7. Zhou Y, Liu M, Pan X, Xie C, Li Z. Clinical Significance of Mobile Medical Model in Diagnosis and Treatment of Ankylosing Spondylitis. Rheumatism and Arthritis 2019 Mar;8(3):28-32. [doi: 10.3969/j.issn.2095-4174.2019.03.005]

8. Muellmann S, Bragina I, Voelcker-Rehage C, Rost E, Lippke S, Meyer J, et al. Development and evaluation of two web-based interventions for the promotion of physical activity in older adults: study protocol for a community-based controlled intervention trial. BMC Public Health 2017 May 25;17(1):512 [FREE Full text] [doi: 10.1186/s12889-017-4446-x] [Medline: 28545506]
9. Srikesavan C, Bryer C, Ali U, Williamson E. Web-based rehabilitation interventions for people with rheumatoid arthritis: A systematic review. J Telemed Telecare 2019 Jun;25(5):263-275. [doi: 10.1177/1357633x18768400] [Medline: 29669470]
10. Ji X, Hu L, Wang Y, Luo Y, Zhu J, Zhang J, et al. "Mobile Health" for the Management of Spondyloarthritis and Its Application in China. Curr Rheumatol Rep 2019 Nov 19;21(11):61. [doi: 10.1007/s11926-019-0860-7] [Medline: 31741083]
11. Ji X, Wang Y, Ma Y, Hu Z, Man S, Zhang Y, et al. Improvement of Disease Management and Cost Effectiveness in Chinese Patients with Ankylosing Spondylitis Using a Smart-Phone Management System: A Prospective Cohort Study. Biomed Res Int 2019 Feb 25;2019:2171475. [doi: 10.1155/2019/2171475] [Medline: 30931322]
12. Zhang F. Second Survey of Chinese Rheumatologists. Chin J Allergy Clin Immunol 2014;8(3):165-169. [doi: 10.3969/j.issn.1673-8705.2014.03.001]
13. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using Mobile Apps for Health Management: A New Health Care Mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
14. Azevedo AR, de Sousa HM, Monteiro JA, Lima AR. Future perspectives of Smartphone applications for rheumatic diseases self-management. Rheumatol Int 2014 Aug 29;35(3):419-431. [doi: 10.1007/s00296-014-3117-9] [Medline: 25168866]
15. 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]
16. 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]
17. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: Second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2020-12-01]
18. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile Health Apps for Self-Management of Rheumatic and Musculoskeletal Diseases: Systematic Literature Review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730 [FREE Full text] [doi: 10.2196/14730] [Medline: 31769758]
19. 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]
20. 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]
21. Con D, De Cruz P. Mobile Phone Apps for Inflammatory Bowel Disease Self-Management: A Systematic Assessment of Content and Tools. JMIR Mhealth Uhealth 2016 Feb 01;4(1):e13 [FREE Full text] [doi: 10.2196/mhealth.4874] [Medline: 26831935]
22. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
23. The 46th China statistical report on internet development. China Internet Network Information Center. 2020 Sep 29. URL: http://www.cnnic.cn/gywm/xwzx/rdxw/202009/W020200929343125745019.pdf [accessed 2020-12-02]
24. Mollard E, Michaud K. Mobile Apps for Rheumatoid Arthritis: Opportunities and Challenges. Rheum Dis Clin North Am 2019 May;45(2):197-209. [doi: 10.1016/j.rdc.2019.01.011] [Medline: 30952393]
25. Cooksey R, Brophy S, Husain MJ, Irvine E, Davies H, Siebert S. The information needs of people living with ankylosing spondylitis: a questionnaire survey. BMC Musculoskelet Disord 2012 Dec 10;13:243. [doi: 10.1186/1471-2474-13-243] [Medline: 23227937]
26. Haglund E, Bremander A, Bergman S, Larsson I. Educational needs in patients with spondyloarthritis in Sweden – a mixed-methods study. BMC Musculoskelet Disord 2017 Aug 2;18(1):335. [doi: 10.1186/s12891-017-1689-8] [Medline: 28768510]
27. Xiao F, Liu X, Wei H, Zhou H, Wang H, Jiang Y, et al. AB1001 Online Consultation for Chinese Patients with Rheumatic Diseases Based on Smart System of Disease Management (SSDM) Mobile Tools: A Study of Medical Economics. Ann Rheum Dis 2016 Jul 15;75(Suppl 2):1243-1244. [doi: 10.1136/annrheumdis-2016-eular.3933]
28. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
29. Kwan YH, Ong WJ, Xiong M, Leung YY, Phang JK, Wang CTM, et al. Evaluation of Mobile Apps Targeted at Patients With Spondyloarthritis for Disease Monitoring: Systematic App Search. JMIR Mhealth Uhealth 2019 Oct 28;7(10):e14753 [FREE Full text] [doi: 10.2196/14753] [Medline: 31661080]

30. Statista Research Department. Market share of mobile operating systems in China from January 2013 to May 2020. Statista. 2020. URL: https://www.statista.com/statistics/262176/market-share-held-by-mobile-operating-systems-in-china/ [accessed 2021-03-01]
31. 2018-2019 China Mobile App Store Market Annual Monitoring Report. iiMedia. 2018. URL: https://www.iimedia.cn/c400/63240.html [accessed 2021-02-28]
32. Najm A, Nikiphorou E, Kostine M, Richez C, Pauling JD, Finckh A, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self-management in people living with rheumatic and musculoskeletal diseases. RMD Open 2019 Sep 13;5(2):e001014. [doi: 10.1136/rmdopen-2019-001014] [Medline: 31565245]
33. Harvard S, Gossec L, Pham T, Richette P, Dougados M, Anis A, et al. Measurable definitions of ankylosing spondylitis management recommendations are needed for use in observational studies. Joint Bone Spine 2016 Jan;83(1):101-103. [doi: 10.1016/j.jbspin.2015.01.019] [Medline: 25776447]
34. Zangi HA, Ndosi M, Adams J, Andersen L, Bode C, Boström C, European League Against Rheumatism (EULAR). EULAR recommendations for patient education for people with inflammatory arthritis. Ann Rheum Dis 2015 Jun;74(6):954-962. [doi: 10.1136/annrheumdis-2014-206807] [Medline: 25735643]
35. Wendling D, Lukas C, Prati C, Claudepierre P, Gossec L, Goupille P, et al. 2018 update of French Society for Rheumatology (SFR) recommendations about the everyday management of patients with spondyloarthritis. Joint Bone Spine 2018 May;85(3):275-284. [doi: 10.1016/j.jbspin.2018.01.006] [Medline: 29407043]
36. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
37. 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]
38. Stec MA, Arbour MW, Hines HF. Client‐Centered Mobile Health Care Applications: Using the Mobile Application Rating Scale Instrument for Evidence‐Based Evaluation. J Midwifery Womens Health 2019 May;64(3):324-329. [doi: 10.1111/jmwh.12941] [Medline: 30887711]
39. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984 Apr;27(4):361-368. [doi: 10.1002/art.1780270401] [Medline: 6231933]
40. Nguyen AD, Baysari MT, Kannangara DR, Tariq A, Lau AY, Westbrook JI, et al. Mobile applications to enhance self-management of gout. Int J Med Inform 2016 Oct;94:67-74. [doi: 10.1016/j.ijmedinf.2016.06.021] [Medline: 27573313]
41. Llorens-Vernet P, Miró J. Standards for Mobile Health-Related Apps: Systematic Review and Development of a Guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
42. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: Potential benefits to patients and providers. J Am Pharm Assoc (2003) 2013;53(2):172-181. [doi: 10.1331/JAPhA.2013.12202] [Medline: 23571625]
43. Ozgocmen S, Akgul O, Altay Z, Altindag O, Baysal O, Calis M, Anatolian Group for the Assessment in Rheumatic Diseases. Expert opinion and key recommendations for the physical therapy and rehabilitation of patients with ankylosing spondylitis. Int J Rheum Dis 2012 Jun;15(3):229-238. [doi: 10.1111/j.1756-185x.2011.01700.x] [Medline: 22709485]
44. Millner JR, Barron JS, Beinke KM, Butterworth RH, Chasle BE, Dutton LJ, et al. Exercise for ankylosing spondylitis: An evidence-based consensus statement. Semin Arthritis Rheum 2016 Feb;45(4):411-427. [doi: 10.1016/j.semarthrit.2015.08.003] [Medline: 26493464]
45. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-Related Mobile Apps in the China App Store: Assessment of Functionality and Quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261 [FREE Full text] [doi: 10.2196/13261] [Medline: 31364606]
46. Wang F, Wang Z, Sun W, Yang X, Bian Z, Shen L, et al. Evaluating the Quality of Health-Related WeChat Public Accounts: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 May 08;8(5):e14826 [FREE Full text] [doi: 10.2196/14826] [Medline: 32383684]

Abbreviations

AS: ankylosing spondylitis
ASDAS: Ankylosing Spondylitis Disease Activity Score
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
EULAR: European League Against Rheumatism
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SSDM: Smart System of Disease Management

uMARS: Mobile App Rating Scale, user version

Edited by L Buis; submitted 18.01.21; peer-reviewed by F Huang, KL Mauco, B Nievas Soriano; comments to author 22.02.21; revised version received 01.04.21; accepted 27.05.21; published 14.07.21.

Please cite as:
Song Y, Chen H
Evaluating Chinese Mobile Health Apps for Ankylosing Spondylitis Management: Systematic App Search
JMIR Mhealth Uhealth 2021;9(7):e27234
URL: https://mhealth.jmir.org/2021/7/e27234
doi:10.2196/27234
PMID:34259644

Original Paper

Popular Evidence-Based Commercial Mental Health Apps: Analysis of Engagement, Functionality, Aesthetics, and Information Quality

Nancy Lau[1,2,3], PhD; Alison O'Daffer[1], BA; Joyce P Yi-Frazier[1], PhD; Abby R Rosenberg[1,3,4], MD, MSc, MA

[1]Palliative Care and Resilience Lab, Center for Clinical and Translational Research, Seattle Children's Research Institute, Seattle, WA, United States
[2]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States
[3]Cambia Palliative Care Center of Excellence, University of Washington, Seattle, WA, United States
[4]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States

Corresponding Author:
Nancy Lau, PhD
Palliative Care and Resilience Lab
Center for Clinical and Translational Research
Seattle Children's Research Institute
1920 Terry Ave
Seattle, WA, 98101
United States
Phone: 1 2068840569
Email: nancy.lau@seattlechildrens.org

Abstract

Background: There is a robust market for mobile health (mHealth) apps focused on self-guided interventions to address a high prevalence of mental health disorders and behavioral health needs in the general population. Disseminating mental health interventions via mHealth technologies may help overcome barriers in access to care and has broad consumer appeal. However, development and testing of mental health apps in formal research settings are limited and far outpaced by everyday consumer use. In addition to prioritizing efficacy and effectiveness testing, researchers should examine and test app design elements that impact the user experience, increase engagement, and lead to sustained use over time.

Objective: The aim of this study was to evaluate the objective and subjective quality of apps that are successful across both research and consumer sectors, and the relationships between objective app quality, subjective user ratings, and evidence-based behavior change techniques. This will help inform user-centered design considerations for mHealth researchers to maximize design elements and features associated with consumer appeal, engagement, and sustainability.

Methods: We conducted a user-centered design analysis of popular consumer apps with scientific backing utilizing the well-validated Mobile Application Rating Scale (MARS). Popular consumer apps with research support were identified via a systematic search of the App Store iOS (Apple Inc) and Google Play (Google LLC) and literature review. We evaluated the quality metrics of 19 mental health apps along 4 MARS subscales, namely, Engagement, Functionality, Aesthetics, and Information Quality. MARS total and subscale scores range from 1 to 5, with higher scores representing better quality. We then extracted user ratings from app download platforms and coded apps for evidence-based treatment components. We calculated Pearson correlation coefficients to identify associations between MARS scores, App Store iOS/Google Play consumer ratings, and number of evidence-based treatment components.

Results: The mean MARS score was 3.52 (SD 0.71), consumer rating was 4.22 (SD 0.54), and number of evidence-based treatment components was 2.32 (SD 1.42). Consumer ratings were significantly correlated with the MARS Functionality subscale (r=0.74, P<.001), Aesthetics subscale (r=0.70, P<.01), and total score (r=0.58, P=.01). Number of evidence-based intervention components was not associated with MARS scores (r=0.085, P=.73) or consumer ratings (r=–0.329, P=.16).

Conclusions: In our analysis of popular research-supported consumer apps, objective app quality and subjective consumer ratings were generally high. App functionality and aesthetics were highly consistent with consumer appeal, whereas evidence-based components were not. In addition to designing treatments that work, we recommend that researchers prioritize aspects of app design that impact the user experience for engagement and sustainability (eg, ease of use, navigation, visual appeal). This will

help translate evidence-based interventions to the competitive consumer app market, thus bridging the gap between research development and real-world implementation.

(JMIR Mhealth Uhealth 2021;9(7):e29689) doi:10.2196/29689

KEYWORDS

mobile health; mental health; behavioral health; user-centered design; evidence-based health management; smartphones; mobile phones

Introduction

In the Digital Age, smartphones have permeated all aspects of personal and professional life. There is a robust market for mobile health (mHealth) apps focused on self-help for mental health and behavioral health needs [1,2]. Despite the widespread appeal of mHealth for mental health, we raise 2 important considerations for its adoption. First, development and testing of mental health apps in formal research settings are limited and far outpaced by everyday consumer use [2-5]. Second, app design elements such as engagement and functionality impact whether users continue to use a mobile app for sustained behavior change over time beyond initial download [6].

The aim of this study was to conduct a user-centered design analysis of the usability, engagement, and quality of popular evidence-based apps for mental health self-management utilizing the well-established Mobile Application Rating Scale (MARS) [7]. Previous publications have utilized the MARS to evaluate apps on an eclectic array of health-related topics including blood pressure, mindfulness, nutrition, diet and physical activity, deafness and hard-of-hearing, and drug–drug interactions [8-13]. This study evaluated the quality of mental health apps that are successful across both research and commercial sectors. We evaluated the relationships between objective app quality, subjective user ratings, and evidence-based behavior change techniques. This will help inform design considerations for mHealth researchers to maximize consumer appeal, engagement, and sustainability.

Methods

Overview

In a recent review of consumer apps, we identified 21 mental health self-management apps that were publicly available and research supported [5]. Two have since been removed by the developers, consistent with previous findings that consumer apps are retired at a rapid rate [5]. For this pool of 19 apps, we conducted the following data collection and analyses (February to April 2021) to address the current study objectives: MARS evaluations by 2 independent coders, extraction of consumer ratings from the App Store iOS (Apple Inc) and Google Play (Google LLC), coding of evidence-based treatment components, and correlation analyses.

We utilized the MARS, a validated objective measure for assessing the quality of mHealth apps [7]. The 23-item MARS provides a total score and Engagement, Functionality, Aesthetics, and Information Quality subscale scores. MARS total score and subscale scores range from 1 to 5, with higher scores representing better quality. Independent raters (NL and AO) downloaded and evaluated each app; individual item scores were averaged between raters according to accepted standards [11]. App consumer ratings were extracted and averaged across App Store iOS and Google Play. We coded app content as evidence-based behavior change techniques (ie, based in behavior change theory or psychological interventions shown to be efficacious or effective) or not evidence based (ie, digital content/modules such as daily inspirational quotes that are not a component of traditional evidence-based mental health interventions).

Statistical Analysis

Interrater reliability was assessed using the intraclass correlation coefficient (ICC) according to established guidelines; we ran a 2-way mixed effects, average measures model with a consistency of agreement definition [14,15]. We calculated Cronbach α to assess the internal consistency of the MARS [16,17]. Descriptive statistics were used to summarize MARS scores, consumer ratings, and number of evidence-based treatment components. Pearson correlation coefficients were calculated to compare (1) the MARS overall score with consumer ratings, (2) the MARS subscale scores with consumer ratings, and (3) number of evidence-based treatment components with MARS and consumer ratings. All analyses were conducted in IBM SPSS Statistics version 27.

Results

The MARS demonstrated high interrater reliability (ICC 0.97, 95% CI 0.97-0.98), and the total score had high internal consistency (Cronbach α=.99). Table 1 shows the MARS total score and subscale scores, mean consumer ratings obtained from app download platforms, and number of evidence-based treatment components for each of the 19 apps. The MARS total mean score for all apps was 3.52 (SD 0.71; range 2.22-4.32). The MARS subscale mean scores were as follows: Engagement, 3.98 (SD 0.82; range 2.30-4.80); Functionality, 3.42 (SD 0.80; range 2.00-4.63); Aesthetics, 3.23 (SD 0.90; range 1.67-4.67); and Information Quality, 3.47 (0.69; range 2.00-4.29).

Average consumer ratings across the App Store iOS and Google Play were 4.22 (SD 0.54; range 3.24-4.86). Average number of evidence-based treatment components was 2.32 (SD 1.42; range 1-5). Notably, 8/19 (42%) of apps consisted of a singular approach to treatment (ie, had only 1 evidence-based treatment component). Headspace had both the highest MARS total score and highest consumer rating, and is a unimodal intervention that teaches mindfulness meditation for emotion regulation and health behavior change.

Each of the MARS subscales was significantly correlated with each other (r=0.63-0.88, *P*<.01) and the MARS total score

(r=0.84-0.91, *P*<.001). Consumer ratings were correlated with the MARS Functionality subscale (r=0.74, *P*<.001; Figure 1), Aesthetics subscale (r=0.70, *P*<.01; Figure 2), and total score (r=0.58, *P*=.01; Figure 3). Number of evidence-based treatment components was not significantly associated with MARS scores (r=0.09, *P*=.73) or average consumer ratings (r=–.33, *P*=.17).

Table 1. MARS scores, consumer ratings, and number of evidence-based treatment components.

App name	MARS scores					Consumer ratings		Total number of evidence-based components
	Engagement	Functionality	Aesthetics	Information quality	Total score	Average user ratings	Total number of ratings	
10% Happier	3.80	4.50	4.33	3.43	4.02	4.80	96,707	3
AEON Mindfulness	2.80	3.00	2.17	2.71	2.67	4.00	45	1
Calm	4.60	4.13	4.67	3.43	4.21	4.66	1,431,242	1
DeStressify	4.10	3.00	2.33	3.36	3.20	4.15	17	3
Habitica	4.20	3.00	3.00	2.57	3.19	4.28	19,354	1
Happify	4.50	3.75	3.67	3.57	3.87	4.22	5615	4
Headspace	4.40	4.25	4.33	4.29	4.32	4.86	861,451	1
MindSurf	3.00	2.00	1.67	2.21	2.22	3.24	18	4
MoodMission	4.50	2.25	2.67	3.86	3.32	3.31	203	5
One Moment Meditation	2.30	2.38	2.33	2.00	2.25	4.81	1281	1
Pacifica/Sanvello	4.80	4.00	4.17	4.00	4.24	4.62	32,749	5
Provider Resilience	3.40	3.25	3.00	4.00	3.41	3.43	43	2
PTSD Coach	4.60	3.75	4.00	4.07	4.11	4.66	1781	3
Smiling Mind	3.60	3.50	3.50	4.07	3.67	3.86	3600	2
Stop, Breathe and Think/MyLife Meditation	4.60	4.13	3.83	3.79	4.09	4.68	39,329	1
SuperBetter	4.60	4.00	3.67	4.07	4.00	4.50	12,737	1
T2 Mood Tracker	2.40	2.25	2.00	2.71	2.34	3.60	1875	1
Virtual Hope Box	4.80	3.25	2.50	4.00	3.64	3.86	1133	3
Woebot	4.60	4.63	3.50	3.71	4.11	4.73	13,041	2

Figure 1. MARS Functionality subscale score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.

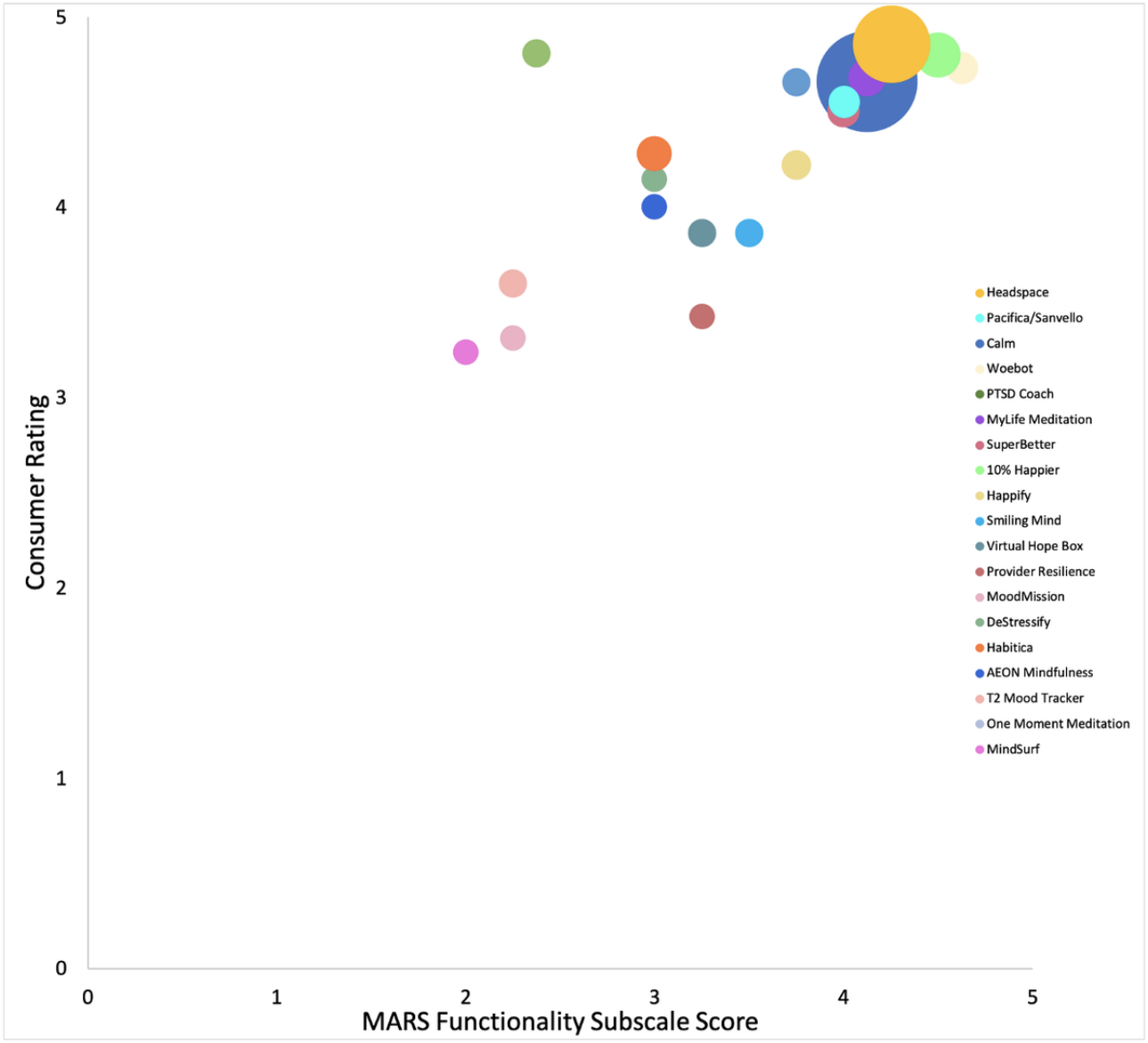


Figure 2. MARS Aesthetics subscale score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.

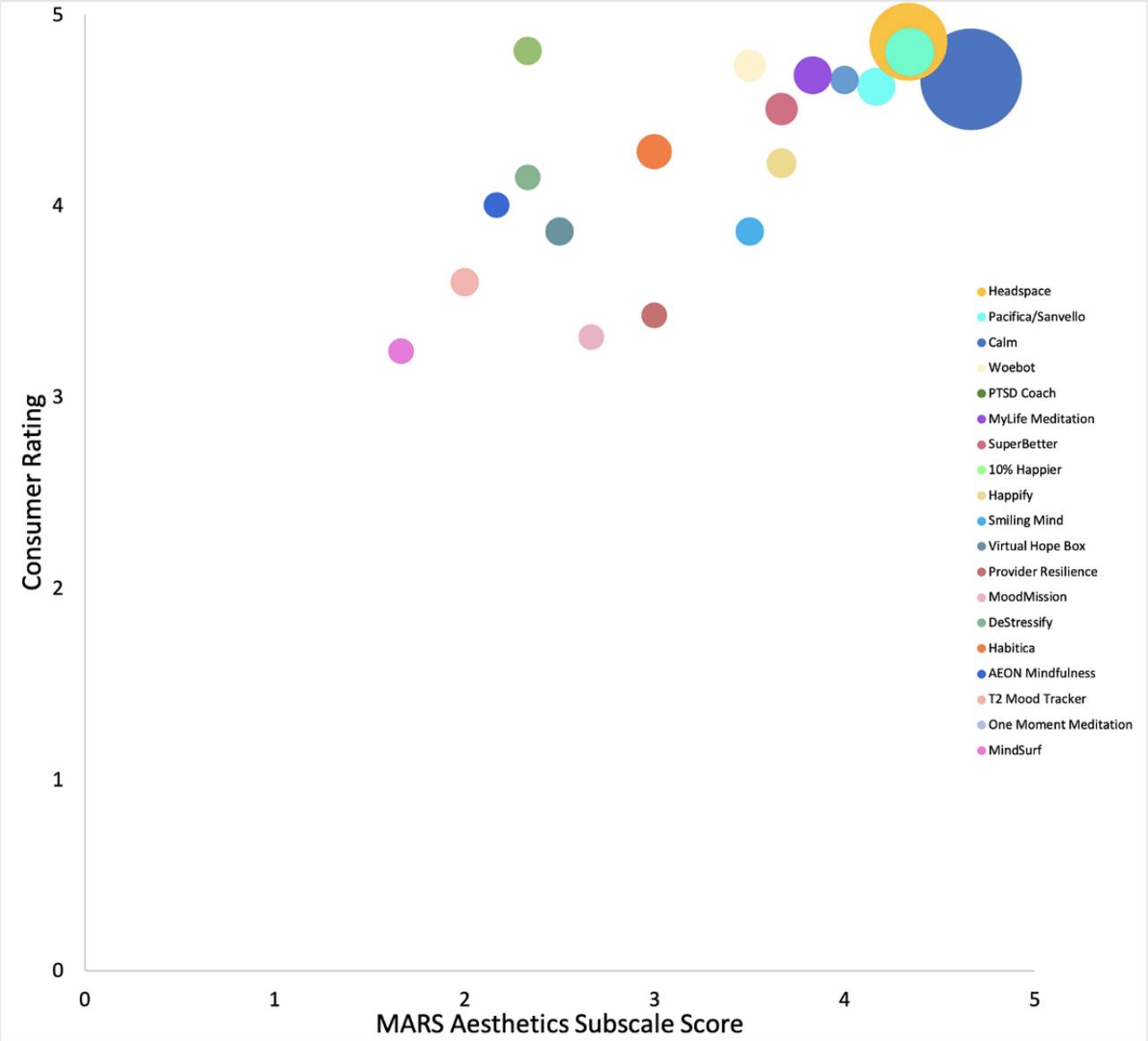
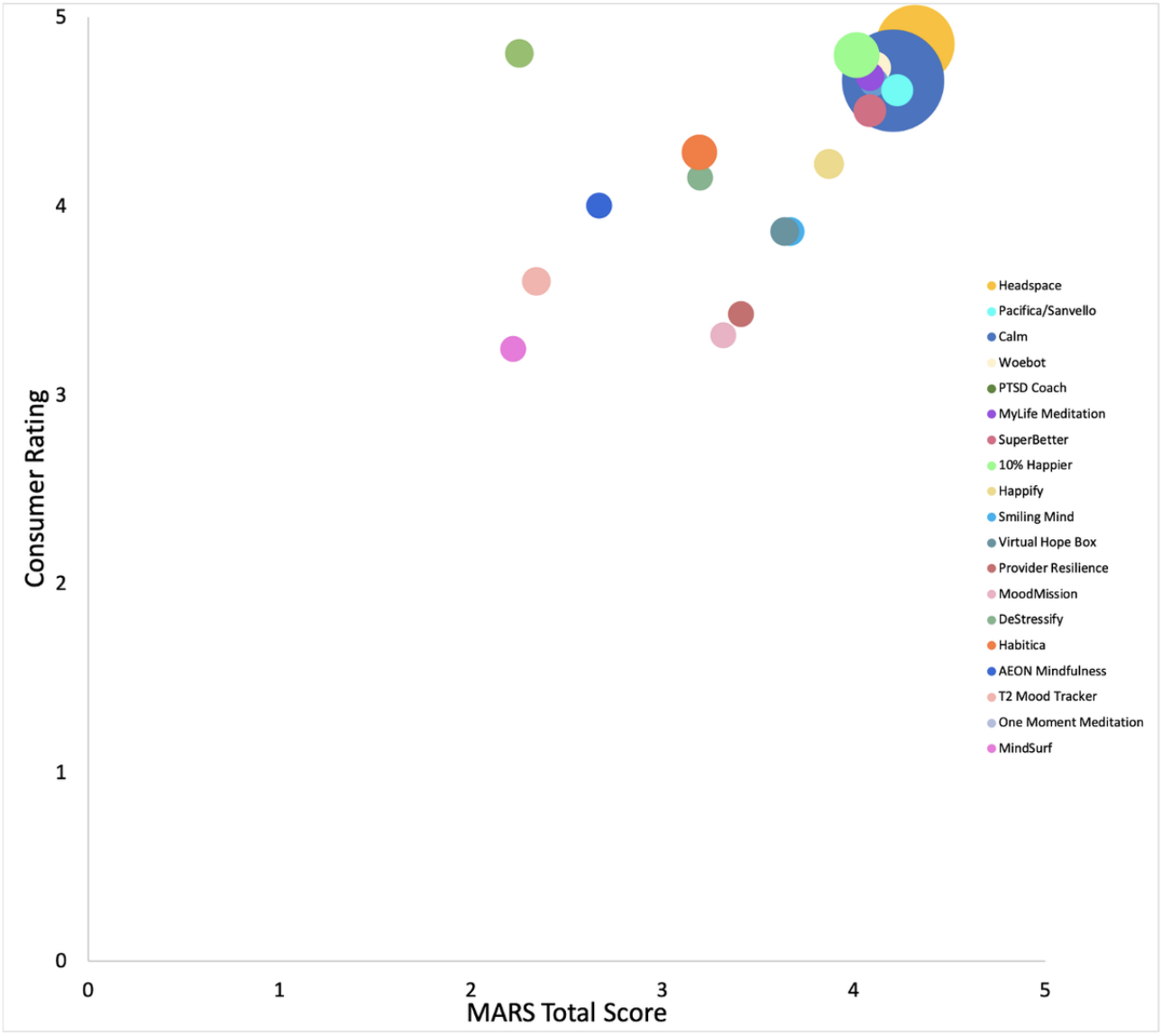


Figure 3. MARS total score × consumer rating scatterplot of 19 apps with variations in bubble size proportionate to app's total number of consumer ratings. MARS: Mobile Application Rating Scale.



Discussion

Principal Findings

In this paper, we described an analysis of the usability and quality of popular research-supported consumer mental health apps using the MARS which provides a bite-sized synthesis of app usability and quality that are easily accessible to consumers and researchers alike [1,18]. The mental health apps we evaluated were of good quality overall. We observed overall MARS scores that were comparable to other reviews of self-management apps [12,19,20]. Previous research has compared MARS scores with consumer ratings for various health-related apps; results varied in whether MARS scores were correlated with consumer ratings [8,9,12].

With regard to popular research-supported mental health apps, we draw the following conclusions: First, we found that consumer ratings were related to objective quality of the app overall. This suggests alignment between subjective assessment of quality by app users and objective assessment of quality by researchers. Second, consumer ratings were related to functionality and aesthetics. This suggests that design elements such as ease of use, navigation, graphics, and visual appeal may be more likely to impact the positivity of the user experience. Third, evidence-based treatment components were not associated with app quality or consumer ratings, and almost half of the apps had a singular skill focus. This suggests that quantity of evidence-based behavior change techniques designed and tested in traditional face-to-face mental health interventions is not what appeals to app consumers. Perhaps unimodal rather than multimodal intervention approaches lend themselves better to a self-guided format and decreases user burden.

Limitations

This study did not evaluate all mental health apps available for public download, nor would it be feasible to do so. Rather, our analysis focused on a small targeted subset of apps identified in a prior publication that we utilized to address a new research question focused on objective and subjective app quality and user-centered design considerations for engagement and sustainability. Thus, our analysis may not represent the whole of all available resources. Apps were limited to English language

that were available in the United States, downloadable on major app platforms, and with peer-reviewed publications.

Conclusions and Implications

We found that objective app quality—and functionality and aesthetics in particular—was highly consistent with consumer appeal. Quantity of evidence-based—and presumably effective—behavior change techniques was not associated with app quality or consumer appeal. To translate evidence-based interventions to the competitive consumer app space, researchers should prioritize aspects of design that impact the user experience such as ease of use, navigation, graphics, and visual appeal. This work may be informed by user-centered design approaches in which iterative development of apps prioritize end user's needs in the contexts in which the intervention will be implemented [21]. In addition, the complexity of evidence-based multimodal interventions may hinder chances of mHealth adoption. Adapting and optimizing design features to the individuals and settings that are unique to the digital space will help engage and retain users over time.

Acknowledgments

NL is funded as an Implementation Science Scholar through the National Heart, Lung, and Blood Institute of the National Institutes of Health (Grant number: 5K12 HL137940-02). The opinions herein represent those of the authors and not necessarily the funders.

Conflicts of Interest

None declared.

References

1. Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
2. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
3. Chib A, Lin SH. Theoretical Advancements in mHealth: A Systematic Review of Mobile Apps. J Health Commun 2018;23(10-11):909-955 [FREE Full text] [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
4. Bose J, Hedden S, Lipari R, Park-Lee E. Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18-5068, NSDUH Series H-53). Rockville, MD: Center for Behavioral Health Statistics and Quality/Substance Abuse and Mental Health Services Administration; 2017. URL: https://www.samhsa.gov/data/report/2017-nsduh-annual-national-report [accessed 2021-06-27]
5. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
6. Chang TR, Kaasinen E, Kaipainen K. What influences users' decisions to take apps into use? A framework for evaluating persuasive and engaging design in mobile apps for well-being. New York, NY: ACM; 2012 Presented at: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia; December 2012; Ulm, Germany. [doi: 10.1145/2406367.2406370]
7. 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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
8. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
9. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
10. 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 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]
11. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van LW, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
12. Flaherty S, McCarthy M, Collins A, McAuliffe F. Can existing mobile apps support healthier food purchasing behaviour? Content analysis of nutrition content, behaviour change theory and user quality integration. Public Health Nutr 2018 Feb;21(2):288-298. [doi: 10.1017/S1368980017002889] [Medline: 29081322]
13. 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]

14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
15. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
16. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
17. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
18. Schueller SM, Washburn JJ, Price M. Exploring Mental Health Providers' Interest in Using Web and Mobile-Based Tools in their Practices. Internet Interv 2016 May;4(2):145-151. [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
19. Weekly T, Walker N, Beck J, Akers S, Weaver M. A Review of Apps for Calming, Relaxation, and Mindfulness Interventions for Pediatric Palliative Care Patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]
20. Bardus M, van BSB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
21. Lyon AR, Koerner K. User-Centered Design for Psychosocial Intervention Development and Implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale

Edited by L Buis; submitted 16.04.21; peer-reviewed by R Cochran, B Chaudhry, S Acquilano; comments to author 07.05.21; revised version received 07.05.21; accepted 11.06.21; published 14.07.21.

Please cite as:
Lau N, O'Daffer A, Yi-Frazier JP, Rosenberg AR
Popular Evidence-Based Commercial Mental Health Apps: Analysis of Engagement, Functionality, Aesthetics, and Information Quality
JMIR Mhealth Uhealth 2021;9(7):e29689
URL: https://mhealth.jmir.org/2021/7/e29689
doi:10.2196/29689
PMID:34259639

Original Paper

Evaluating Quality, Usability, Evidence-Based Content, and Gamification Features in Mobile Learning Apps Designed to Teach Children Basic Life Support: Systematic Search in App Stores and Content Analysis

Nino Fijačko[1*], MSc; Ruth Masterson Creber[2*], RN, MSc, PhD; Lucija Gosak[1*], MSc; Gregor Štiglic[1,3,4*], PhD; Dominic Egan[5*], RN; Brian Chaka[6*], MSc; Nika Debeljak[1*], BA; Matej Strnad[7*], MD, PhD; Pavel Skok[7*], MD, PhD

[1]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia

[2]Healthcare Policy and Research, Division of Health Informatics, Weill Cornell Medicine, New York, NY, United States

[3]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[4]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[5]School of Nursing and Healthcare Leadership, University of Bradford, Bradford, United Kingdom

[6]School of Allied Health Professions and Midwifery, University of Bradford, Bradford, United Kingdom

[7]Faculty of Medicine, University of Maribor, Maribor, Slovenia

[*]all authors contributed equally

Corresponding Author:
Nino Fijačko, MSc
Faculty of Health Sciences
University of Maribor
Žitna 15
Maribor, 2000
Slovenia
Phone: 386 23004764
Fax: 386 23004747
Email: nino.fijacko@um.si

Abstract

Background: Globally, 3.7 million people die of sudden cardiac death annually. Following the World Health Organization endorsement of the *Kids Save Lives* statements, initiatives to train school-age children in basic life support (BLS) have been widespread. Mobile phone apps, combined with gamification, represent an opportunity for including mobile learning (m-learning) in teaching schoolchildren BLS as an additional teaching method; however, the quality of these apps is questionable.

Objective: This study aims to systematically evaluate the quality, usability, evidence-based content, and gamification features (GFs) of commercially available m-learning apps for teaching guideline-directed BLS knowledge and skills to school-aged children.

Methods: We searched the Google Play Store and Apple iOS App Store using multiple terms (eg, *cardiopulmonary resuscitation* [*CPR*] or *BLS*). Apps meeting the inclusion criteria were evaluated by 15 emergency health care professionals using the user version of the Mobile Application Rating Scale and System Usability Scale. We modified a *five-finger* mnemonic for teaching schoolchildren BLS and reviewed the apps' BLS content using standardized criteria based on three CPR guidelines. GFs in the apps were evaluated using a gamification taxonomy.

Results: Of the 1207 potentially relevant apps, only 6 (0.49%) met the inclusion criteria. Most apps were excluded because the content was not related to teaching schoolchildren BLS. The mean total scores for the user version of the Mobile Application Rating Scale and System Usability Scale score were 3.2/5 points (95% CI 3.0-3.4) and 47.1/100 points (95% CI 42.1-52.1), respectively. Half of the apps taught hands-only CPR, whereas the other half also included ventilation. All the apps indicated when to start chest compressions, and only 1 app taught BLS using an automated external defibrillator. Gamification was well integrated into the m-learning apps for teaching schoolchildren BLS, whereas the *personal and fictional*, *educational*, *and performance* gamification groups represented most GFs.

Conclusions: Improving the quality and usability of BLS content in apps and combining them with GFs can offer educators novel m-learning tools to teach schoolchildren BLS skills.

(JMIR Mhealth Uhealth 2021;9(7):e25437) doi:10.2196/25437

KEYWORDS

cardiopulmonary resuscitation; basic life support; mobile learning; mobile phone; gamification; schoolchildren

Introduction

Background

Sudden cardiac arrest is a leading cause of mortality, responsible for 3.7 million deaths per year [1-5]. Most deaths occur in the community and can be prevented with basic life support (BLS) [6], specifically cardiopulmonary resuscitation (CPR), which doubles the chances of survival [6-8]. The European Resuscitation Council (ERC) [9] and American Heart Association (AHA) [10] guidelines recommend that lay persons respond immediately after a patient collapses and before the arrival of emergency paramedic personnel; however, CPR bystander response rates are <50%, primarily because of low self-efficacy and knowledge of safely conducting CPR by the lay public [6,11,12].

Following the World Health Organization endorsement of the *Kids Save Lives* statement [13], initiatives to include BLS training in primary and secondary schools have been implemented in the hope of increasing rates of bystander CPR [14-16]. Early findings demonstrate that when schoolchildren were educated in BLS, bystander rates of CPR have doubled [17]. Proponents of educating schoolchildren in BLS can do so using interactive digital technologies, including mobile learning (m-learning) [18,19] with gamification features (GFs) [20,21]. Gamification, popularly defined as “the use of game design elements in non-game contexts,” [20] has emerged as a means of harnessing competitiveness by integrating gamification elements such as leaderboards, rewards, badges, avatars, and competitions to engage and motivate consumers [21,22]. The most common tools for teaching BLS to schoolchildren include self-made games [23-25], posters [26], songs [27], and manikins [23,28,29].

Objectives

The ERC guidelines for resuscitation recommend that the use and development of technology and social media should be encouraged and the impact, assessed [30]. Reviews of apps offering real-time instructions for adult learning and bystander CPR have been published [31,32]; however, they exclude school-aged children. This study aims to systematically evaluate the quality, usability, evidence-based content, and gamification features of commercially available m-learning apps for teaching guideline-directed BLS knowledge and skills to school-aged children.

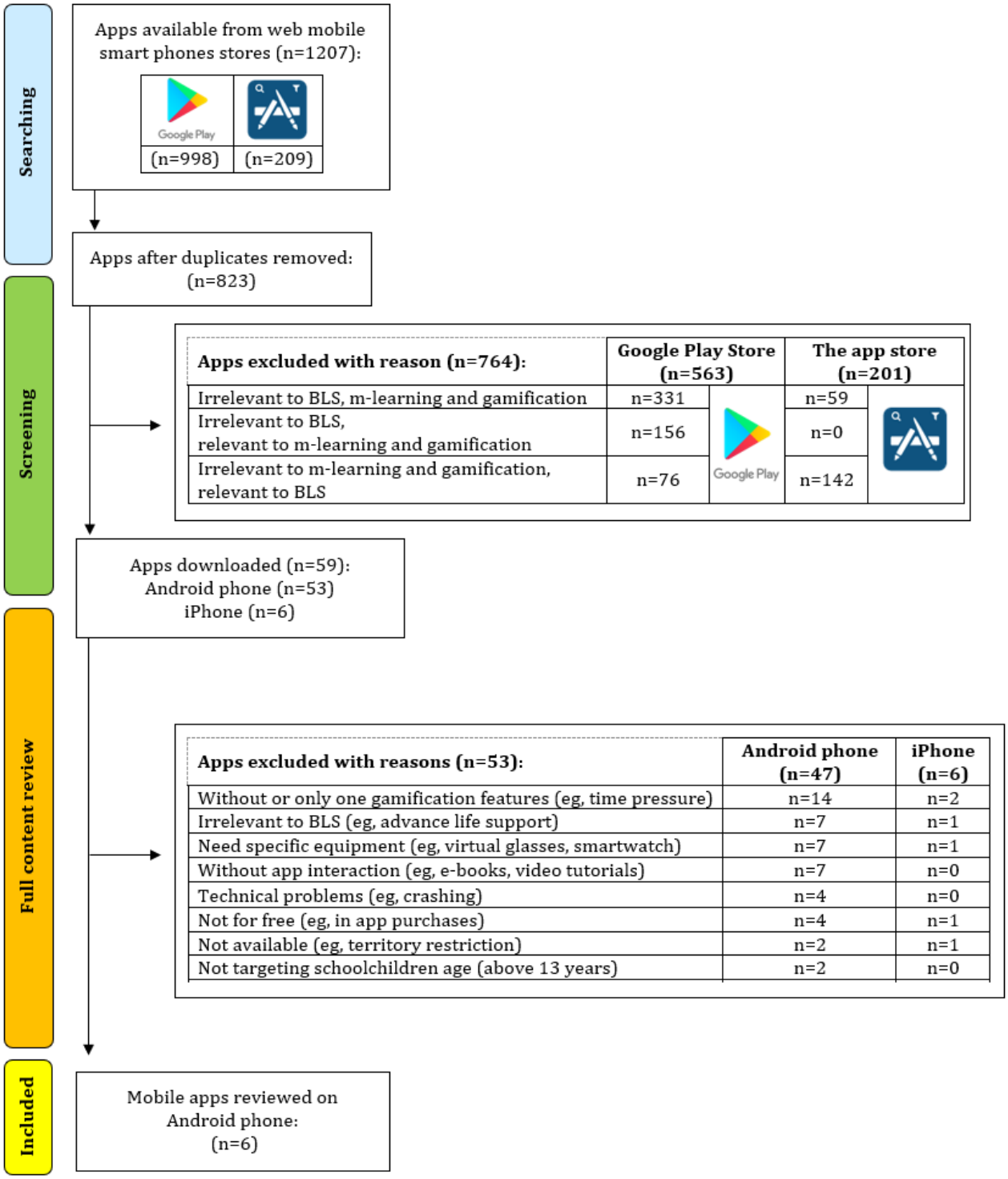
Methods

Searching, Screening, and Reviewing of Commercially Available Apps

We conducted a systematic search of commercially available apps using a rigorous methodology that has been previously published [33,34]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [35] is available in Multimedia Appendix 1. We searched the Google Play Store [36] (for Android apps) and the App Store (for Apple iOS apps) [37] in May 2020. We created inclusion criteria including the *population* (schoolchildren aged 6-13 years), *intervention* (free apps without in-app purchases that contain GFs for schoolchildren to learn CPR), and *outcomes* (app contents for teaching CPR by emergency health care professionals) [38].

Our search strategy was conducted in three rounds. First, the apps were searched using search strings (*cardiopulmonary resuscitation*, *CPR BLS*, *CPR BLS kids game*, *CPR BLS game*, and *CPR BLS kids*). After removing duplicates, the apps were screened based on their title, icons, screenshots, photography, pictures, videos, and descriptions by 2 independent investigators. During this round, apps were excluded based on three criteria: (1) irrelevant to BLS, m-learning, and gamification; (2) irrelevant to BLS, relevant to m-learning and gamification; and (3) irrelevant to m-learning and gamification, relevant to BLS. In the third round, the apps were downloaded onto the Samsung Galaxy S8 (Android 9.0 *Pie*) and iPhone 7 (iOS 12.3.1 Apple Inc) mobile phones for Android and iOS apps, respectively, and content was fully reviewed. During this round, apps were excluded based on nine criteria: (1) without or only one GF, (2) irrelevant to BLS, (3) need of specific equipment, (4) without app interaction, (5) technical problems, (6) not for free, (7) not available, (8) not targeting schoolchildren age, and (9) not in English. To ensure consistency, when discrepancies arose, a consensus was reached through discussion by the researchers. The PRISMA flow diagram [35] was used to represent the selection of the included and excluded apps (Figure 1). If the apps were found in both the Google Play Store [36] and Apple App Store [37], they were reviewed in the Google Play Store [39].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of app selection. BLS: basic life support.



Evaluator's Recruitment

The inclusion criteria for the selection of emergency health care professionals were as follows: above 18 years, CPR training certification by an established medical association, and more than 5 years of experience teaching emergency medicine. A total of 15 emergency health care professionals rated each app independently in a laboratory environment using two validated rating tools, the user version of the Mobile Application Rating Scale (uMARS) [40] and System Usability Scale (SUS) [41]. The duration of each app review was recorded, and 3 additional investigators with expertise in consumer health informatics and emergency medicine reviewed each app and rated them for BLS content.

Ethics Approval

The study was conducted in a central European country (Slovenia). All emergency health care professionals signed informed consent to participate, and ethics approval was obtained from the two Health Care Centre in the north-eastern part of Slovenia.

Measures and Rating Tools

Rating Tools for App Quality and Usability

uMARS provides a multidimensional measure of performance indicators, functionality, esthetics, and quality of information. Apps' subjective quality was not assessed. All items were rated on a 5-point Likert scale ranging from 1 (inadequate) to 5 (excellent). The SUS includes 10 statements on a 5-point Likert scale with both positive and negative statements about usability. The total score for the SUS is 100 points, which is divided into six usability categories including worst imaginable (0-25 SUS score), poor (25.1-51.6 SUS score), ok (51.7-62.6 SUS score), good (62.7-72.5 SUS score), excellent (72.6-84.0 SUS score), and best imaginable (84.1-100 SUS score) [42-44]. The uMARS and SUS have been used in similar studies [31,32,34,45-47].

Rating Tool for BLS-Related Content Based on CPR Evidence

In total, 3 investigators evaluated evidence-based BLS in each app using the ERC [9,48], AHA [10,49], and Australian Resuscitation Council [50] guidelines for teaching BLS. On the basis of these guidelines, the team identified 17 discrete BLS contents and divided them into five groups based on the Slovenian Resuscitation Council *five-finger* BLS teaching mnemonic for teaching first responders BLS [51]: (1) *safety* (1 item, BLS1), (2) *consciousness* (2 items, BLS2-3), (3) *breathing and call* (4 items, BLS4-7), (4) *CPR* (9 items, BLS8-16), and (5) *defibrillation* (1 item, BLS17), as shown in Multimedia Appendix 2 [9,10,48,49,51]. The *five-finger* BLS teaching mnemonic is based on two well-known ways of remembering CPR: (1) *DRSABCD* (pronounced *drs A-B-C-D*; danger, responses, send, airway, breathing, CPR, defibrillation) [52] action plan and (2) *chain of survival* [53]. The scoring system included one point if the BLS content was correctly implemented based on the BLS guidelines. We used a digital metronome to compare the frequency of chest compressions in apps.

Rating Tool for GFs in Apps

We modified gamification taxonomy [54,55] into five gamification groups, where each group represented different GFs: (1) *ecological* (4 GFs, GF1-4), (2) *social* (3 GFs, GF5-7), (3) *personal and fictional* (7 GFs, GF8-14), (4) *performance* (8 GFs, GF15-22), and (5) *educational* (4 GFs, GF23-26). The purpose of this gamification taxonomy is to evaluate m-learning environments such as apps. Each gamification group has a different relationship with the environment and learners in the form of implementation (ecological group), interaction (social group), usage (personal and fictional group), response (performance group), and knowledge (educational group). A total of 26 GFs were included in the final gamification taxonomy, as presented in Multimedia Appendix 3 [54,55]. The gamification rating was classified using a dummy coding [56] by 2 investigators, one point for inclusion of gamification taxonomy, no points for no gamification taxonomy, and 0.5 points for partial implementation of gamification taxonomy.

Data Analysis

We used Microsoft Office Professional 2016, R (version 3.6.0, R Foundation for Statistical Computing), SPSS Statistics for Windows (version 27.0, IBM Corp), and Inkscape 1.0 (Inkscape Developers, GNU General Public License) to analyze and visualize the results. The interrater reliability for uMARS and SUS was calculated using the intraclass correlation coefficient ($ICC_{2,\ k}$; intraclass correlation coefficient, two-way random, average measures, and absolute agreement) [57,58].

Results

Overview

We identified 1207 apps. The PRISMA flow diagram presents the process of selecting and scanning apps using the inclusion and exclusion criteria. After removing duplicates from multiple search strings from web-based mobile smartphone stores, 63.29% (764/1207) apps remained; 4.88% (59/1207) apps were relevant to BLS, m-learning, and gamification. After applying all the inclusion and exclusion criteria, 0.49% (6/1207) apps were included in the final evaluation (Figure 1).

All apps were classified into the *educational* category by the Google Play Store. Half of the apps had a disclaimer that the app was made for educational purposes only and was not a substitute for accredited BLS training. Only 2 apps required registration. Apps were developed across multiple countries, including Australia, Italy, Finland, China, and the United States. All apps were developed in collaboration with one or more health care organizations (Table 1 and Multimedia Appendix 4).

Table 1. Description of the included apps.

Full App Name	Health care organization collaborator	Country, BLS[a] guideline organization	Description of BLS scenarios
First Aid Action Hero [59]	St John Ambulance Australia (Victoria)	Australia (St John Ambulance Australia)	One scenario in which the user performed BLS, first on a conscious and second on an unconscious animated cartoon figure
CPR APP [60]	Emergency Medicine Unit, Li Ka Shing Faculty of Medicine, The University of Hong Kong	United States (American Heart Association)	No scenario. In the simulation environment, the user performed BLS on an animated human figure
Everyday Lifesaver [61]	Life Saving Victoria Limited	Australia (St John Ambulance Australia)	Three scenarios in which the user performed BLS on a drowned adult, a drowned child, and an unconscious animated cartoon figure
A Breathtaking Picnic [62]	The Italian Resuscitation Council	Italy (The Italian Resuscitation Council)	Two scenarios in which the user performed BLS on an animated animal that was choking and one that experienced cardiac arrest
ReLIVe Responder [63]	The University of Pittsburgh, Department of Emergency Medicine	United States (American Heart Association)	Two scenarios in which the user performed BLS on an unconscious and a conscious animated human figure
Responder Rescuebusters: Fire and First-Aid [64]	Emergency Response Centre Agency Finland, Finnish Recovery Council, Finnish Fire Officers' Association's	Finland (Finnish Recovery Council)	One scenario in which the user performed BLS on an animated human figure that experienced cardiac arrest

[a]BLS: basic life support.

Most apps targeted children aged above 4 years of age, and 1 app—*A Breathtaking Picnic* [62]—targeted schoolchildren aged between 6 and 8 years. All the apps had a Pan European Game Information 3 certificate [65]. According to game genres [65], 3 apps were developed as animated tutorials, 2 apps were developed as simulations, and 1 app was developed as a virtual world; 2 apps represented the first responder as virtual characters (eg, animated boy or animal), and an animated victim was included in each app (Table 1).

uMARS Quality and SUS Usability Rating

A total of 15 emergency health care professionals participated (3 females and 12 males) in evaluating apps using uMARS and SUS. In total, 40% (6/15) of the participants were nurses, 27% (4/15) were nurses with a master's degree, and 33% (5/15) were physicians. Overall, the mean age of emergency health care professionals was 36 years. All emergency health care professionals had an Advanced Life Support (ALS) certificate provided by the ERC, and their mean professional experience was 13 years. All emergency health care professionals own and were proficient daily users of mobile smartphones.

The mean total uMARS rating of apps was 3.2/5 (95% CI 3.0-3.4), and the details across the four domains are shown in Multimedia Appendix 5. The mean testing app time was 9.5 minutes. The most time-consuming app was the *Everyday Lifesaver* [61] app (mean 24 minutes) because the app included multiple features for evaluation. Interrater reliability between emergency health care professionals was good for the overall uMARS score ($ICC_{2,k}$ 0.8, 95% CI 0.8-0.9; Tables 2 and 3; Multimedia Appendix 6) but poor for overall SUS score ($ICC_{2,k}$ 0.3, 95% CI 0.03-0.5). The *A Breathtaking Picnic* [62] app had the highest mean SUS score (54.8 points). The mean SUS score of all assessed apps was 47.1/100 (95% CI 42.1-52.1) points. The usability of the apps was rated from *poor* to *ok*. The mean SUS score of the apps is indicated by a red dashed line. The six bands in the Figure 2 indicate the six levels of SUS categories of usability.

Table 2. User version of the Mobile Application Rating Scale results and time spent on each app.

Full app name	uMARS[a] section					Time[b]
	Engagement	Functionality	Aesthetics	Information	Overall app quality	
First Aid Action Hero [59]	3.7	4.0	4.0	3.9	3.9	12
CPR APP [60]	3.1	3.7	3.2	3.7	3.4	7
Everyday Lifesaver [61]	3.5	3.0	3.7	3.5	3.4	24
A Breathtaking Picnic [62]	2.9	3.5	4.1	3.7	3.5	5
ReLIVe Responder [63]	2.8	3.6	3.2	3.5	3.3	5
Responder Rescuebusters: Fire and First-Aid [64]	3.4	3.4	3.6	3.2	3.4	4

[a]uMARS: user version of the Mobile Application Rating Scale.

[b]Mean time for testing apps (in minutes).

Table 3. Mean scores and intraclass correlation coefficients of the user version of the Mobile Application Rating Scale and time spent on each app.

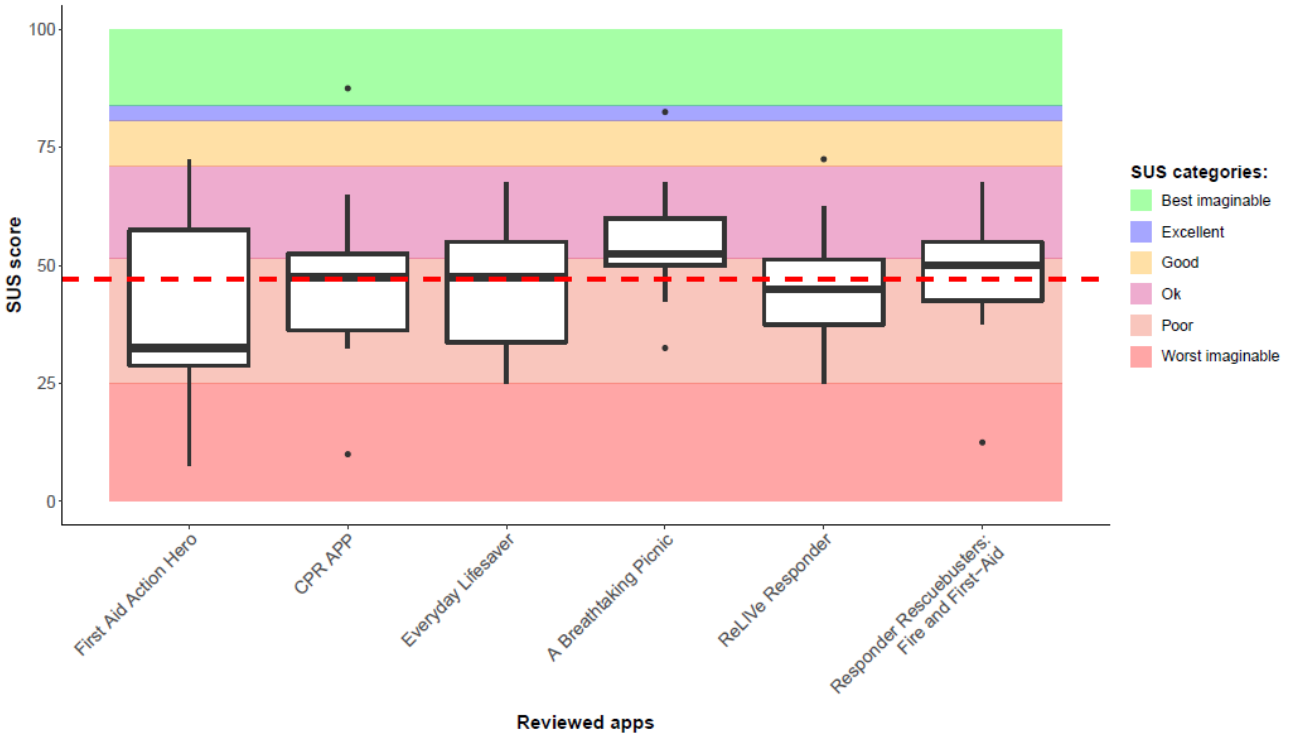
Variable	Score, mean (95% CI)	$ICC_{2,k}$[a] (95% CI)
uMARS[b] section		
Engagement	3.2 (3.0-3.4)	0.9 (0.8-0.9)
Functionality	3.5 (3.4-3.7)	0.7 (0.5-0.8)
Aesthetics	3.6 (3.4-3.8)	0.8 (0.6-0.9)
Information	3.6 (3.4-3.8)	0.8 (0.6-0.9)
Overall app quality	3.2 (3.0-3.4)	0.9 (0.8-0.9)
Time for testing apps (min)	9.2 (7.7-10.7)	N/A[c]

[a]$ICC_{2,k}$: intraclass correlation coefficient; two-way random, average measures, absolute agreement.

[b]uMARS: user version of the Mobile Application Rating Scale.

[c]N/A: not applicable.

Figure 2. System Usability Scale results in the form of a box plot for each review app. SUS: System Usability Scale.



Evidence-Based BLS Content

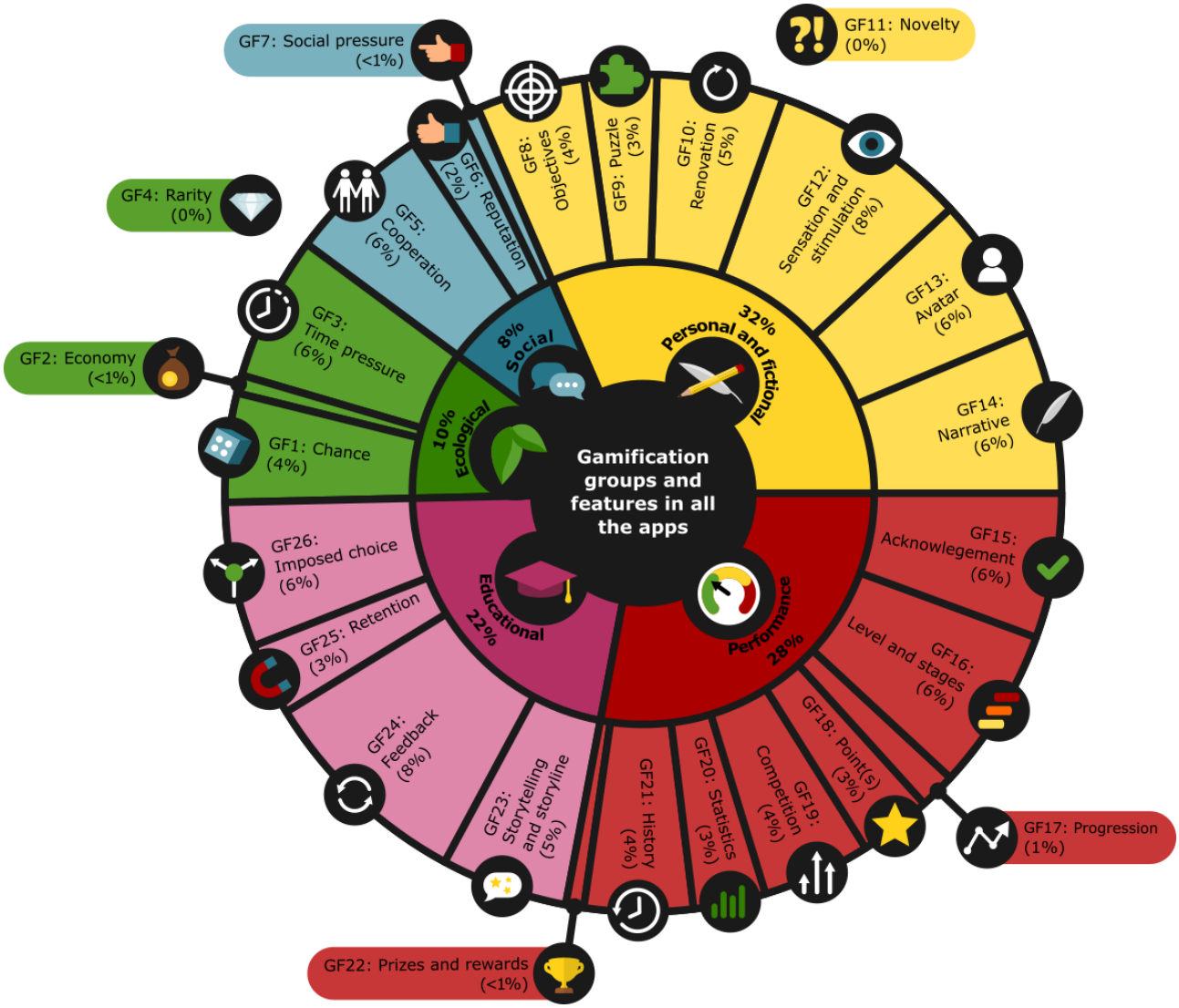
The overall evaluation of evidence-based BLS content in the apps was *poor* to *average* based on expert consensus. Within each of the five categories, there were inconsistencies regarding mapping to the ERC [9,48], AHA [10,49], and Australian Resuscitation Council [50] guidelines for teaching BLS (Figure 3 and Multimedia Appendix 2). In the *safety* category, 50% (3/6) of the apps included checking for safety. For *consciousness*, most apps (4/6, 67%) included checking for responsiveness. Within the *breathing and call* category, the vast majority (5/6, 83%) included calling the emergency number or asking someone to call them; however, none of the apps included placing the patient in the right recovery position. In the *CPR* category, half of the apps taught hands-only CPR, whereas the other half also included ventilation, which is inconsistent with the most recent BLS guidelines internationally. Only 1 app included teaching BLS using an automated external defibrillator (AED). In the 4 apps, the chest compression frequency was set to 100 beats per minute.

Figure 3. Basic life support groups and percentages of basic life support contents in the apps. AED: automated external defibrillator; BLS: basic life support; CPR: cardiopulmonary resuscitation.

Gamification Features

The most common GFs in apps were in *personal and fictional* (32%), *performance* (28%), and *educational* (22%; Figure 4 and Multimedia Appendix 3) gamification groups. One of the most integrated GFs in the apps for teaching schoolchildren BLS was feedback and sensation or stimulation (both 8%).

Figure 4. Gamification groups and features in all the apps. GF: gamification feature.



Discussion

Principal Findings

We conducted a systematic evaluation of the quality, usability, evidence-based content, and GFs of commercially available m-learning apps for teaching guideline-directed BLS knowledge and skills to school-aged children. Overall, the quality of the apps was *average* based on the uMARS, the usability was *poor* to *ok* based on the SUS, and the quality of the content was *poor* to *average* in terms of alignment with international BLS guidelines, and GFs were well represented across the gamification taxonomy.

Quality and Usability of Apps

Many of the apps analyzed in this review were not high-quality apps according to the uMARS tool. Overall, the lowest mean uMARS score was represented in the engagement section, evaluating entertainment, interest, customization, interactivity, and target group. Future apps could learn from this review by ensuring that user engagement is prioritized during the development phase. From a customization perspective, some personal app options (eg, selecting gender or adding names of a player) or BLS options such as changing BLS victims or scenarios, including or excluding ventilation as a part of CPR, and varying chest compression frequency should be added. Nevertheless, the information section represented the highest uMARS score, and app developers should consider adding more relevant evidence-based BLS contents to BLS m-learning apps.

Relevant to functionality and esthetics, schoolchildren prefer visually attractive apps [22] with high levels of interactivity [66-68] and a relatable storyline (eg, bicycle accident) that evokes empathy for the victim [26]. However, in this study, the quality of visual esthetics was inversely proportional to the app's learnability and usability. Of the 6 apps examined with the SUS tool, the app *A Breathtaking Picnic* [62] had the highest mean usability score. Educators must consider a few critical aspects of choosing the most usable m-learning app for teaching BLS to schoolchildren. Currently, there is no single app

available that is appropriate for teaching BLS to all children aged 8-18 years. On the basis of this review, we recommend using the *A Breathtaking Picnic* [62] app (54.8 points) for teaching schoolchildren aged 6-8 years, the *First Aid Action Hero* [59] app (40.1 points) for children aged 8-10 years, and the *Everyday Lifesaver* [61] app (45.7 points) for children aged 11-18 years. Apps with SUS scores higher than 82 have a considerable chance of being recommended to a colleague [32]; in this review, none of the apps achieved this score.

Evaluation of Evidence-Based BLS Content

This review represented the poor-to-average quality of BLS content regarding international BLS guidelines [9,10,48-50].

Five-Finger BLS Teaching Mnemonic

We classified BLS training for schoolchildren using a modified version of the *five-finger* BLS teaching mnemonic [26,49], including (1) *safety*, (2) *consciousness*, (3) *breathing and call*, (4) *CPR*, and (5) *defibrillation* (Figure 3).

Safety

In terms of safety, only half of the apps were designed to check whether the area was safe before approaching the victim. All BLS guidelines reinforce the importance of ensuring safety for first responders, victims, and bystanders [9,10,48-50].

Consciousness

Consciousness, in most apps, was assessed by checking the victim's responsiveness to the question, "Are you all right?" and gently shaking the victim. For example, in the *Everyday Lifesaver* [61] app, responsiveness is taught using the acronym COWS [50,69] ("Can you hear me?", "Open your eyes," "What's your name?", and "Squeeze my hand"). Studies have shown that schoolchildren do not have problems in correctly assessing consciousness [24,26,70,71].

Breathing and Call

As recommended in the CPR guidelines [9,10,48-50], for the breathing and call category, the head tilt–chin lift maneuver is generally well taught in the apps, except when the jaw has to be lifted upward to bring the chin forward and the teeth almost to occlusion. Most apps included *calling the emergency number or asking somebody to call them*; however, none of the apps correctly showed the process of moving the victim into the right recovery position or turning away from the rescuer (Figure 3). A problem with most apps is teaching the *look*, *listen*, and *feel* method for signs of breathing discretely. Importantly, in the context of the COVID-19 pandemic, the head tilt–chin lift maneuver and *look*, *listen*, and *feel* method is no longer recommended in the 2020 guidelines [72]. Most apps do not emphasize abnormal types of breath; only the *ReLIVe Responder* [63] app and *Everyday Lifesaver* [61] app provide information about agonal breathing or gasping or gurgling, as was shown in a study conducted in 2018 [32].

Most apps correctly demonstrated how to make an emergency service call using the speakerphone function. For example, in the *Everyday Lifesaver* [61] app game scenario, the mock operator's questions are based on the 5 Ps (*place, phone number, problem, people, and progress*). Most schoolchildren can correctly remember the information needed to make emergency calls [73-75].

CPR

Chest Compression

The most common BLS content included in apps was chest compressions. According to the modified CPR guidelines [72], hands-only CPR is recommended to decrease the risk of COVID-19 infection. Chest compressions were typically indicated by a circle on the chest for each compression. The major problem is that the area for pressing frequently does not correspond with the correct anatomical location, and this could provide users with misleading visual information, especially schoolchildren. As such, the biggest limitation of the apps overall is the gestural design of chest compressions, especially because there is no universal gestural design. In addition, GFs focus on users' attention to pressing a specific circle rather than focusing on the victim. Similarly, in a real scenario, there are no clear indicators of where to compress, and not knowing appropriate anatomical landmarks could create confusion. This is further complicated by the fact that most of the victims were cartoon animals, so specific locations for chest compressions are unclear when translated to humans.

An alternative approach to indicating the point for compression using a circle is to interact with the phone by holding the smartphone in the palm, facing up, and moving it up and down with the chest compressions. This method facilitates immediate visual feedback on how chest compressions should be performed. A limitation of the apps was that the compression site was inconsistently labeled, leading to inaccurate hand and arm positions. According to a study conducted in 2019, consistency with hand and arm positions is critical for chest compression accuracy [76]. According to the CPR guidelines [9,10,48-50], chest compressions should be at least 5-6 cm in depth, at a rate of 100-120 compressions per minute (2 per second). In general, the apps provided appropriate BLS information regarding the depth and rate of chest compression. On average, the frequency of chest compressions was set to 100 beats per minute and could not be changed to higher frequency rates. Only a few apps emphasize chest recoil and fraction. One of the challenges for schoolchildren is having the strength to perform chest compressions [77]. Even if a child is not physically able to perform chest compressions, they can still learn the fundamentals of BLS and are capable of learning comprehensive BLS content and selecting skills [29,78].

Ventilation

Our results indicate that half of the apps do not include steps for ventilation when teaching CPR. Those that do include ventilation provide accurate BLS contents about how to perform mouth-to-mouth ventilation. However, ventilation volume and verification of chest rising are poorly integrated into the ventilation part of CPR. A study from 2019 indicated that teaching schoolchildren ventilation requires more teaching time, and it is harder to establish good quality BLS results [76]. Apps that teach ventilation as a part of CPR are more time-consuming; however, they also adhere more closely to the BLS guidelines. In addition, according to the modified CPR guidelines,

mouth-to-mouth ventilation is not recommended to decrease the risk of COVID-19 infection [72].

Defibrillation

Overall, the use of AEDs was poorly represented. Only the *Everyday Lifesaver* [61] app included AEDs for teaching schoolchildren BLS training. A study reported that using an app that provides AEDs may be beneficial in terms of performance and security but at the cost of delivering a shock [79]. However, a small proportion of the schoolchildren without previous training could use an AED correctly in less than 3 minutes following the device's acoustic and visual instructions [80].

GFs in Apps

Recently, gamified m-learning has become increasingly popular in various medical and educational contexts, including BLS training [81-83]. Through gamification, not only can apps create a mindset that encourages schoolchildren to try new things without being afraid of failing [84,85] but it also enables schoolchildren to engage in the learning experience.

Personal and fictional, *educational*, and *performance* GFs were the most represented in apps for teaching BLS to schoolchildren. Gamification audio or visual BLS feedback features, levels and stages in the way of BLS steps, and sensation in a sort of stimulation were well integrated into apps for teaching schoolchildren BLS. Knowledge retention plays an important role in teaching BLS [22,86-88]; only two apps (eg, the Everyday Lifesaver [61] app) included knowledge retention in the form of repeated BLS content. It is recommended that retention content be integrated into apps because of the rapid deterioration of BLS skills after training.

Schoolchildren today have high smartphone literacy, but less is known about educators. To use m-learning to teach BLS to schoolchildren, educators must feel confident about the platform. Some resuscitation councils, such as the Italian Resuscitation Council [89] or recently ERC [90], have already recognized that m-learning is a new trend in education and are starting to emphasize m-learning in the future resuscitation teaching guidelines. The most recent ERC draft guidelines [91] recommend that schoolchildren need supervision for learning BLS and that schoolteachers are more appropriately positioned to teach schoolchildren than health care professionals [92,93].

Using gamified learning features, educators can expect changes in psychomotor, cognitive, and affective learning outcomes [82,83,94]. A systematic review [65] demonstrated that knowledge acquisition and retention of content, productive learning experience, and motor skills are all improved when GFs are incorporated into m-learning. Therefore, learning BLS should include both knowledge transfer and the motivation to perform BLS. We conclude that m-learning has the potential to be used to enhance BLS education for schoolchildren by improving the retention of BLS knowledge.

Limitations

We deliberately selected emergency health care professionals and not schoolteachers to review the apps because we were focused on adherence to evidence-based guidelines for educational purposes. There were also a small number of final apps, probably as a result of our prespecified criteria, including focusing on a target population below 13 years of age and free apps. In addition, a few apps were excluded because of limitations in language design (eg, Held: Reanimatie Game, in Dutch) [95] and the area where the study was conducted (eg, the First Aid Skills app is available only in Australia) [96]. Finally, all of the apps were developed before the COVID-19 pandemic; therefore, COVID-19–specific modifications to BLS content were not included.

Future Research

The results of this study provide opportunities for developing an app for teaching BLS to schoolchildren. The *First Aid Action Hero* [59] app and *A Breathtaking Picnic* [62] app have potential to be part of a randomized controlled study in which the effects of m-learning on knowledge retention, motor skills, and motivation to perform BLS can be evaluated.

Conclusions

Our study represents an opportunity to include m-learning apps for teaching BLS to schoolchildren. Using an adapted *five-finger* BLS teaching mnemonic and m-learning with GFs, there is tremendous potential for teaching BLS to schoolchildren to improve survival rates of cardiac arrest.

Acknowledgments

This study was partially supported by the *Digital toolbox for innovation in nursing education (I-BOX)*, *Improving Health Care Students' Competences for Behaviour Change to Effectively Support Self-care in Chronic Diseases (Train4Health)*, which is cofunded by the European Union Erasmus+ program and the Slovenian Research Agency (grant N2-0101 and P2-0057). In addition, RMC is supported by the National Institute of Nursing Research (R00NR016275). The authors would like to thank all the emergency health care professionals (Vrečar V, Borovnik Lesjak V, Vec M, Kramberger A, Vihar D, Moravs D, Koželj A, Lešnik D, Lešnik B, Mažič M, Nežmah J, Zabukovšek D, Vitka V, and Petrovčič R) for participating in the app evaluation.

Authors' Contributions

This study was conducted with collaboration among all authors. NF developed the study design and supervised the study. NF, LG, and RMC drafted the manuscript. NF, LG, ND, and GS collected and analyzed the data. NF, MS, and PS interpreted the results from a BLS point of view. RMC, DE, and BC conducted a comprehensive review of the content. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 35 KB - mhealth_v9i7e25437_app1.docx]

Multimedia Appendix 2
Basic life support groups and contents.
[DOCX File , 27 KB - mhealth_v9i7e25437_app2.docx]

Multimedia Appendix 3
Gamification groups and features.
[DOCX File , 26 KB - mhealth_v9i7e25437_app3.docx]

Multimedia Appendix 4
Description of the included apps.
[DOCX File , 14 KB - mhealth_v9i7e25437_app4.docx]

Multimedia Appendix 5
User version of the Mobile Application Rating Scale scores and sections.
[DOCX File , 20 KB - mhealth_v9i7e25437_app5.docx]

Multimedia Appendix 6
User version of the Mobile Application Rating Scale and time spent on each app results.
[DOCX File , 15 KB - mhealth_v9i7e25437_app6.docx]

References

1. Beck B, Bray J, Cameron P, Smith K, Walker T, Grantham H, Aus-ROC Steering Committee. Regional variation in the characteristics, incidence and outcomes of out-of-hospital cardiac arrest in Australia and New Zealand: results from the Aus-ROC Epistry. Resuscitation 2018 May;126:49-57 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.02.029] [Medline: 29499230]
2. Kudenchuk PJ, Sandroni C, Drinhaus HR, Böttiger BW, Cariou A, Sunde K, et al. Breakthrough in cardiac arrest: reports from the 4th Paris International Conference. Ann Intensive Care 2015 Dec 17;5(1):22 [FREE Full text] [doi: 10.1186/s13613-015-0064-x] [Medline: 26380990]
3. Wong CX, Brown A, Lau DH, Chugh SS, Albert CM, Kalman JM, et al. Epidemiology of sudden cardiac death: global and regional perspectives. Heart Lung Circ 2019 Jan;28(1):6-14. [doi: 10.1016/j.hlc.2018.08.026] [Medline: 30482683]
4. Abhilash S, Namboodiri N. Sudden cardiac death - historical perspectives. Indian Heart J 2014 Jan;66 Suppl 1:4-9 [FREE Full text] [doi: 10.1016/j.ihj.2014.01.002] [Medline: 24568828]
5. Murugappan M, Boon H, Raj AN, Krishnan G, Palanisamy K. ECG morphological features based sudden cardiac arrest (SCA) prediction using nonlinear classifiers. In: Sengodan T, Murugappan M, Misra S, editors. Advances in Electrical and Computer Technologies. Singapore: Springer; 2020:789-801.
6. Hasselqvist-Ax I, Riva G, Herlitz J, Rosenqvist M, Hollenberg J, Nordberg P, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med 2015 Jun 11;372(24):2307-2315. [doi: 10.1056/nejmoa1405796]
7. Perkins GD, Lockey AS, de Belder MA, Moore F, Weissberg P, Gray H, Community Resuscitation Group. National initiatives to improve outcomes from out-of-hospital cardiac arrest in England. Emerg Med J 2016 Jul 23;33(7):448-451 [FREE Full text] [doi: 10.1136/emermed-2015-204847] [Medline: 26400865]
8. Sasson C, Rogers MA, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2010 Jan;3(1):63-81. [doi: 10.1161/circoutcomes.109.889576]
9. Monsieurs KG, Nolan JP, Bossaert LL, Greif R, Maconochie IK, Nikolaou NI, ERC Guidelines 2015 Writing Group. European Resuscitation Council Guidelines for Resuscitation 2015: Section 1. Executive summary. Resuscitation 2015 Oct;95:1-80. [doi: 10.1016/j.resuscitation.2015.07.038] [Medline: 26477410]
10. Neumar RW, Shuster M, Callaway CW, Gent LM, Atkins DL, Bhanji F, et al. Part 1: Executive Summary: 2015 American Heart Association Guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015 Oct 14;132(18 suppl 2):315-367. [doi: 10.1161/cir.0000000000000252]
11. Perkins GD, Brace-McDonnell SJ, OHCAO Project Group. The UK Out of Hospital Cardiac Arrest Outcome (OHCAO) project. BMJ Open 2015 Oct 01;5(10):e008736 [FREE Full text] [doi: 10.1136/bmjopen-2015-008736] [Medline: 26428332]

12. Meaney PA, Bobrow BJ, Mancini ME, Christenson J, de Caen AR, Bhanji F, et al. Cardiopulmonary Resuscitation Quality: improving cardiac resuscitation outcomes both inside and outside the hospital a consensus statement from the American Heart Association. Circulation 2013 Jul 23;128(4):417-435 [FREE Full text] [doi: 10.1161/CIR.0b013e31829d8654] [Medline: 23801105]
13. Böttiger BW, Van Aken H. Kids save lives - Training school children in cardiopulmonary resuscitation worldwide is now endorsed by the World Health Organization (WHO). Resuscitation 2015 Sep;94:5-7. [doi: 10.1016/j.resuscitation.2015.07.005] [Medline: 26209417]
14. Böttiger BW, Van Aken H. Training children in cardiopulmonary resuscitation worldwide. Lancet 2015 Jun;385(9985):2353. [doi: 10.1016/s0140-6736(15)61099-6]
15. Böttiger BW, Bossaert LL, Castrén M, Cimpoesu D, Georgiou M, Greif R, Board of European Resuscitation Council (ERC). Kids Save Lives - ERC position statement on school children education in CPR.: "Hands that help - Training children is training for life". Resuscitation 2016 Aug;105:1-3. [doi: 10.1016/j.resuscitation.2016.06.005] [Medline: 27339096]
16. Bohn A, Van Aken HK, Möllhoff T, Wienzek H, Kimmeyer P, Wild E, et al. Teaching resuscitation in schools: annual tuition by trained teachers is effective starting at age 10. A four-year prospective cohort study. Resuscitation 2012 May;83(5):619-625. [doi: 10.1016/j.resuscitation.2012.01.020] [Medline: 22286049]
17. Wissenberg M, Lippert FK, Folke F, Weeke P, Hansen CM, Christensen EF, et al. Association of national initiatives to improve cardiac arrest management with rates of bystander intervention and patient survival after out-of-hospital cardiac arrest. J Am Med Assoc 2013 Oct 02;310(13):1377-1384. [doi: 10.1001/jama.2013.278483] [Medline: 24084923]
18. Liu Y, Li H, Carlsson C. Factors driving the adoption of m-learning: an empirical study. Comput Edu 2010 Nov;55(3):1211-1219. [doi: 10.1016/j.compedu.2010.05.018]
19. Pedro LF, Barbosa CM, Santos CM. A critical review of mobile learning integration in formal educational contexts. Int J Educ Technol High Educ 2018 Mar 15;15(1):10. [doi: 10.1186/s41239-018-0091-4]
20. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; September, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
21. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
22. Mkpojiogu EO, Hussain A, Hassan F. A systematic review of usability quality attributes for the evaluation of mobile learning applications for children. AIP Conf Proc 2016 2018:020092 [FREE Full text] [doi: 10.1063/1.5055494]
23. Nagy R, Bánfai-Csonka H, Musch J, Betlehem J, Bánfai B. Comparison of effectiveness of video-based only, video-based with self instruction kit and conventional resuscitation training for primary school children. Resuscitation 2020 Oct;155:S9. [doi: 10.1016/j.resuscitation.2020.08.042]
24. Bánfai B, Deutsch K, Pandur A, Bánfai-Csonka H, Betlehem J. Preliminary results of teaching first aid to 5-6 year old children - a longitudinal study. Kontakt 2018 Jun 26;20(2):120-125. [doi: 10.1016/j.kontakt.2018.03.003]
25. Otero-Agra M, Varela-Casal C, Castillo-Pereiro N, Casillas-Cabana M, San Román-Mata S, Barcala-Furelos R, et al. Can we train the chain of survival while playing? Validation of the tool «Rescube». Anales de Pediatría (English Edition) 2021 Apr;94(4):213-222. [doi: 10.1016/j.anpede.2020.07.013]
26. Bollig G, Myklebust AG, Østringen K. Effects of first aid training in the kindergarten - a pilot study. Scand J Trauma Resusc Emerg Med 2011 Feb 28;19(1):13 [FREE Full text] [doi: 10.1186/1757-7241-19-13] [Medline: 21356047]
27. Del Pozo FJ, Alonso JV, Velis NB, Barahona MM, Siggers A, Lopera E. Basic life support knowledge of secondary school students in cardiopulmonary resuscitation training using a song. Int J Med Educ 2016 Jul 20;7:237-241 [FREE Full text] [doi: 10.5116/ijme.5780.a207] [Medline: 27442599]
28. Nakagawa NK, Oliveira KM, Lockey A, Semeraro F, Aikawa P, Macchione M, et al. Effectiveness of the 40-minute handmade manikin program to teach hands-on cardiopulmonary resuscitation at school communities. Am J Cardiol 2021 Jan 15;139:126-130. [doi: 10.1016/j.amjcard.2020.09.032] [Medline: 33007308]
29. Weidenauer D, Hamp T, Schriefl C, Holaubek C, Gattinger M, Krammel M, et al. The impact of cardiopulmonary resuscitation (CPR) manikin chest stiffness on motivation and CPR performance measures in children undergoing CPR training - a prospective, randomized, single-blind, controlled trial. PLoS One 2018 Aug 16;13(8):e0202430 [FREE Full text] [doi: 10.1371/journal.pone.0202430] [Medline: 30114240]
30. Greif R, Lockey AS, Conaghan P, Lippert A, De Vries W, Monsieurs KG, Education and implementation of resuscitation section Collaborators, Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 10. Education and implementation of resuscitation. Resuscitation 2015 Oct;95:288-301. [doi: 10.1016/j.resuscitation.2015.07.032] [Medline: 26477418]
31. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014 Mar 19;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]

32. Metelmann B, Metelmann C, Schuffert L, Hahnenkamp K, Brinkrolf P. Medical correctness and user friendliness of available apps for cardiopulmonary resuscitation: systematic search combined with guideline adherence and usability evaluation. JMIR Mhealth Uhealth 2018 Nov 06;6(11):e190 [FREE Full text] [doi: 10.2196/mhealth.9651] [Medline: 30401673]
33. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
34. Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, et al. The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16365 [FREE Full text] [doi: 10.2196/16365] [Medline: 32673235]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, The 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]
36. Google Play Store. URL: https://play.google.com/store/apps [accessed 2020-10-15]
37. Apple Store – Apple. URL: https://theappstore.org/ [accessed 2020-10-15]
38. Polit D, Beck C. Essentials of Nursing Research: Appraising Evidence for Nursing Practice. Philadelphia, Pennsylvania, United States: Lippincott Williams and Wilkins; 2018:1-512.
39. Zapata BC, Niñirola AH, Idri A, Fernández-Alemán JL, Toval A. Mobile PHRs compliance with Android and iOS usability guidelines. J Med Syst 2014 Aug 24;38(8):81. [doi: 10.1007/s10916-014-0081-6] [Medline: 24957397]
40. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
41. Brooke J. SUS: A quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
42. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usab Stud 2009;4(3):114-123 [FREE Full text]
43. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS): SUS in the wild. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2013 Sep 30 Presented at: Human Factors and Ergonomics Society Annual Meeting; September 30–October 4, 2013; San Diego, CA p. 192-196. [doi: 10.1177/1541931213571043]
44. Kortum P, Sorber M. Measuring the usability of mobile applications for phones and tablets. Int J Hum Comput Interact 2015 Jul 31;31(8):518-529. [doi: 10.1080/10447318.2015.1064658]
45. Ahn C, Cho Y, Oh J, Song Y, Lim TH, Kang H, et al. Evaluation of smartphone applications for cardiopulmonary resuscitation training in South Korea. Biomed Res Int 2016;2016:6418710-6418718 [FREE Full text] [doi: 10.1155/2016/6418710] [Medline: 27668257]
46. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr 8;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
47. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the mobile application rating scale. JMIR Mhealth Uhealth 2019 May 01;7(5):e13019 [FREE Full text] [doi: 10.2196/13019] [Medline: 31066712]
48. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult basic life supportautomated external defibrillation section Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015 Oct;95(10):81-99 [FREE Full text] [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
49. Kleinman ME, Brennan EE, Goldberger ZD, Swor RA, Terry M, Bobrow BJ, et al. Part 5: Adult basic life support and cardiopulmonary resuscitation quality. Circulation 2015 Oct 14;132(18 suppl 2):414-435. [doi: 10.1161/cir.0000000000000259]
50. Guidelines. New Zealand Resuscitation Council. URL: https://www.nzrc.org.nz/guidelines/#Sections%202-8%20Guidelines [accessed 2020-10-15]
51. Livk M, Mesec U, Mohor M. Pet prstov - nov način temeljnih postopkov oživljanja. In: Proceedings of the 25th International Symposium on Emergency Medicine. 2018 Presented at: 25th International Symposium on Emergency Medicine; June 14–6, 2018; Portorož, Slovenia p. 325 URL: https://www.szum.si/media/uploads/files/dokoncni-program-2018.pdf
52. Wilks J, Pendergast D. Skills for life: First aid and cardiopulmonary resuscitation in schools. Health Edu J 2017 Sep 08;76(8):1009-1023. [doi: 10.1177/0017896917728096]
53. Nolan J, Soar J, Eikeland H. The chain of survival. Resuscitation 2006 Dec;71(3):270-271. [doi: 10.1016/j.resuscitation.2006.09.001] [Medline: 17070646]
54. Toda AM, Klock AC, Oliveira W, Palomino PT, Rodrigues L, Shi L, et al. Analysing gamification elements in educational environments using an existing gamification taxonomy. Smart Learn Environ 2019 Dec 04;6(1):1-14. [doi: 10.1186/s40561-019-0106-1]

55. Nah F, Zeng Q, Telaprolu V, Ayyappa A, Eschenbrenner B. Gamification of education: a review of literature. In: International Conference on HCI in Business. Switzerland: Springer; 2014:401-409.
56. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
57. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420]
58. 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]
59. First aid action hero. App Store Preview. URL: https://apps.apple.com/au/app/first-aid-action-hero/id1423366252 [accessed 2020-10-15]
60. CPR APP. URL: https://apkpure.com/cpr-app/cocpr.teli.hku.hk [accessed 2020-10-15]
61. Everyday lifesaver. Google Play Store. URL: https://play.google.com/store/apps/details?id=air.com.lsv.everydaylifesaver [accessed 2020-10-15]
62. A breathtaking picnic. Google Play Store. URL: https://play.google.com/store/apps/details?id=com.elastico.picnic [accessed 2020-10-15]
63. ReLIVe responder. Google Play Store. URL: https://play.google.com/store/apps/details?id=com.simcoachgames.relive [accessed 2020-10-15]
64. Pan European Game Information. URL: https://pegi.info/ [accessed 2020-10-15]
65. Hainey T, Connolly TM, Boyle EA, Wilson A, Razak A. A systematic literature review of games-based learning empirical evidence in primary education. Comput Edu 2016 Nov;102:202-223. [doi: 10.1016/j.compedu.2016.09.001]
66. Mohammed W, Husni H. Reading apps for children: readability from the design perspective. AIP Conf Proc 2017;1891(1):020095. [doi: 10.1063/1.5005428]
67. Kwok K, Ghrear S, Li V, Haddock T, Coleman P, Birch SA. Children can learn new facts equally well from interactive media versus face to face instruction. Front Psychol 2016 Oct 25;7:1603 [FREE Full text] [doi: 10.3389/fpsyg.2016.01603] [Medline: 27826263]
68. Pitchford NJ, Kamchedzera E, Hubber PJ, Chigeda AL. Interactive apps promote learning of basic mathematics in children with special educational needs and disabilities. Front Psychol 2018 Mar 6;9:262 [FREE Full text] [doi: 10.3389/fpsyg.2018.00262] [Medline: 29559940]
69. The SLSA Public Safety and Aquatic Rescue Training Manual - 35th Edition (PSAR35). Surf Life Saving NSW. URL: https://static1.squarespace.com/static/57a457a5414fb54f51f98e49/t/5f911d2e43e34c13f599ac57/1603345737542/PSAR35+M6+RESUSCITATION.pdf [accessed 2021-06-01]
70. Lubrano R, Romero S, Scoppi P, Cocchi G, Baroncini S, Elli M, et al. How to become an under 11 rescuer: a practical method to teach first aid to primary schoolchildren. Resuscitation 2005 Mar;64(3):303-307. [doi: 10.1016/j.resuscitation.2004.09.004] [Medline: 15733758]
71. Bollig G, Wahl HA, Svendsen MV. Primary school children are able to perform basic life-saving first aid measures. Resuscitation 2009 Jun;80(6):689-692. [doi: 10.1016/j.resuscitation.2009.03.012] [Medline: 19386407]
72. European Resuscitation Council COVID-19 Guidelines. European Resuscitation Council. 2020. URL: https://erc.edu/sites/5714e77d5e615861f00f7d18/content_entry5ea884fa4c84867335e4d1ff/5ea885f34c84867335e4d20e/files/ERC_covid19_pages.pdf?1588257310 [accessed 2020-10-15]
73. Fleischhackl R, Nuernberger A, Sterz F, Schoenberg C, Urso T, Habart T, et al. School children sufficiently apply life supporting first aid: a prospective investigation. Crit Care 2009;13(4):R127 [FREE Full text] [doi: 10.1186/cc7984] [Medline: 19646229]
74. Ammirati C, Gagnayre R, Amsallem C, Némitz B, Gignon M. Are schoolteachers able to teach first aid to children younger than 6 years? A comparative study. BMJ Open 2014 Sep 19;4(9):e005848 [FREE Full text] [doi: 10.1136/bmjopen-2014-005848] [Medline: 25239292]
75. Nord A, Svensson L, Hult H, Kreitz-Sandberg S, Nilsson L. Effect of mobile application-based versus DVD-based CPR training on students' practical CPR skills and willingness to act: a cluster randomised study. BMJ Open 2016 Apr 29;6(4):e010717 [FREE Full text] [doi: 10.1136/bmjopen-2015-010717] [Medline: 27130166]
76. Abelsson A, Nygårdh A. To enhance the quality of CPR performed by youth layman. Int J Emerg Med 2019 Oct 07;12(1):30 [FREE Full text] [doi: 10.1186/s12245-019-0247-6] [Medline: 31590653]
77. Mpotos N, Iserbyt P. Children saving lives: Training towards CPR excellence levels in chest compression based on age and physical characteristics. Resuscitation 2017 Dec;121:135-140. [doi: 10.1016/j.resuscitation.2017.10.024] [Medline: 29107674]
78. Yeung J, Kovic I, Vidacic M, Skilton E, Higgins D, Melody T, et al. The school Lifesavers study - a randomised controlled trial comparing the impact of Lifesaver only, face-to-face training only, and Lifesaver with face-to-face training on CPR knowledge, skills and attitudes in UK school children. Resuscitation 2017 Nov;120:138-145. [doi: 10.1016/j.resuscitation.2017.08.010] [Medline: 28974344]

79. Carballo-Fazanes A, Jorge-Soto C, Abelairas-Gómez C, Bello-Rodríguez J, Fernández-Méndez F, Rodríguez-Núñez A. Could mobile apps improve laypeople AED use? Resuscitation 2019 Jul;140:159-160. [doi: 10.1016/j.resuscitation.2019.05.029] [Medline: 31153946]
80. Jorge-Soto C, Abelairas-Gómez C, Barcala-Furelos R, Garrido-Viñas A, Navarro-Patón R, Muiño-Piñeiro M, et al. Automated external defibrillation skills by naive schoolchildren. Resuscitation 2016 Sep;106:37-41. [doi: 10.1016/j.resuscitation.2016.06.007] [Medline: 27353288]
81. Márquez-Hernández VV, Gutiérrez-Puertas L, Garrido-Molina JM, García-Viola A, Granados-Gámez G, Aguilera-Manrique G. Using a mobile phone application versus telephone assistance during cardiopulmonary resuscitation: a randomized comparative study. J Emerg Nurs 2020 Jul;46(4):460-467. [doi: 10.1016/j.jen.2020.03.015] [Medline: 32444161]
82. Corkery G, Dawson-Howe K. A smartphone tool for evaluating cardiopulmonary resuscitation (CPR) delivery. Prague, Czech Republic: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP; 2019:489-496.
83. Leary M, McGovern SK, Chaudhary Z, Patel J, Abella BS, Blewer AL. Comparing bystander response to a sudden cardiac arrest using a virtual reality CPR training mobile app versus a standard CPR training mobile app. Resuscitation 2019 Jun;139:167-173. [doi: 10.1016/j.resuscitation.2019.04.017] [Medline: 31005588]
84. Lee J, Hammer J. Gamification in education: what, how, why bother? Acad Exch Q. 2011. URL: https://www.researchgate.net/publication/258697764_Gamification_in_Education_What_How_Why_Bother [accessed 2021-06-01]
85. Papadakis S, Kalogiannakis M. Using gamification for supporting an introductory programming course. The case of classcraft in a secondary education classroom. In: Brooks AL, Brooks E, Vidakis N, editors. Interactivity, Game Creation, Design, Learning, and Innovation. Switzerland: Springer; 2018:366-375.
86. Watanabe K, Lopez-Colon D, Shuster JJ, Philip J. Efficacy and retention of Basic Life Support education including Automated External Defibrillator usage during a physical education period. Prev Med Rep 2017 Mar;5:263-267 [FREE Full text] [doi: 10.1016/j.pmedr.2017.01.004] [Medline: 28138421]
87. Cutumisu M, Patel SD, Brown MR, Fray C, von Hauff P, Jeffery T, et al. RETAIN: A board game that improves neonatal resuscitation knowledge retention. Front Pediatr 2019 Jan 31;7:13 [FREE Full text] [doi: 10.3389/fped.2019.00013] [Medline: 30766862]
88. Roppolo LP, Pepe PE. Retention, retention, retention: targeting the young in CPR skills training!. Crit Care 2009;13(5):185 [FREE Full text] [doi: 10.1186/cc7997] [Medline: 19769783]
89. Semeraro F, Scapigliati A, Rota E, De Luca M, Tammaro G, Landini C, et al. A breathtaking picnic: a fairy tale to save lives. Resuscitation 2016 Sep;106:e47. [doi: 10.1016/j.resuscitation.2016.07.112]
90. Greif R, Lockey A, Conaghan P. Guidelines 2020: Chapter 12 Education. European Resuscitation Council. URL: https://cprguidelines.eu/ [accessed 2020-10-15]
91. Greif R, Lockey A, Breckwoldt J, Carmona F, Conaghan P, Kuzovlev A, et al. European Resuscitation Council Guidelines 2021: Education for resuscitation. Resuscitation 2021 Apr;161:388-407 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.02.016] [Medline: 33773831]
92. Lukas R, Van Aken H, Mölhoff T, Weber T, Rammert M, Wild E, et al. Kids save lives: a six-year longitudinal study of schoolchildren learning cardiopulmonary resuscitation: who should do the teaching and will the effects last? Resuscitation 2016 Apr;101:35-40 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.01.028] [Medline: 26868079]
93. Bohn A, Lukas RP, Breckwoldt J, Böttiger BW, Van Aken H. 'Kids save lives': why schoolchildren should train in cardiopulmonary resuscitation. Curr Opin Crit Care 2015;21(3):220-225. [doi: 10.1097/mcc.0000000000000204]
94. Sharda R, Romano Jr NC, Lucca JA, Weiser M, Scheets G, Chung J, et al. Foundation for the study of computer-supported collaborative learning requiring immersive presence. J Manag Info Syst 2014 Dec 08;20(4):31-64. [doi: 10.1080/07421222.2004.11045780]
95. Held: Reanimatie game. URL: https://play.google.com/store/apps/details?id=nl.fuego.held [accessed 2020-10-15]
96. First Aid Skills. Google Play Store. URL: https://play.google.com/store/apps/details?id=com.stjohnwa.firstaid&hl=en_US&gl=US [accessed 2020-10-15]

Abbreviations

AED: automated external defibrillator
AHA: American Heart Association
ALS: Advanced Life Support
BLS: basic life support
CPR: cardiopulmonary resuscitation
ERC: European Resuscitation Council
GF: gamification feature
ICC: intraclass correlation coefficient
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SUS: system usability scale

uMARS: user version of the Mobile Application Rating Scale

Edited by L Buis; submitted 02.11.20; peer-reviewed by T Baranowski, X Ding; comments to author 23.11.20; revised version received 12.12.20; accepted 07.05.21; published 20.07.21.

Please cite as:
Fijačko N, Masterson Creber R, Gosak L, Štiglic G, Egan D, Chaka B, Debeljak N, Strnad M, Skok P
Evaluating Quality, Usability, Evidence-Based Content, and Gamification Features in Mobile Learning Apps Designed to Teach Children Basic Life Support: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2021;9(7):e25437
URL: https://mhealth.jmir.org/2021/7/e25437
doi:10.2196/25437
PMID:34283034

Review

Telemedicine in Malignant and Nonmalignant Hematology: Systematic Review of Pediatric and Adult Studies

Aashaka C Shah[1], BS; Linda C O'Dwyer[2], MA, MSLIS; Sherif M Badawy[3,4], MBBCh, MS, MD

[1]Department of Medical Education, University of Illinois College of Medicine, Chicago, IL, United States

[2]Galter Health Sciences Library and Learning Center, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Division of Pediatric Hematology, Oncology, Neuro-Oncology & Stem Cell Transplantation, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[4]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Sherif M Badawy, MBBCh, MS, MD
Division of Pediatric Hematology, Oncology, Neuro-Oncology & Stem Cell Transplantation
Ann & Robert H Lurie Children's Hospital of Chicago
225 E Chicago Avenue, Box #30
Chicago, IL, 60611
United States
Phone: 1 312 227 4000 ext 74836
Fax: 1 312 227 9373
Email: sbadawy@luriechildrens.org

Abstract

Background: Telemedicine, including video-, web-, and telephone-based interventions, is used in adult and pediatric populations to deliver health care and communicate with patients. In the realm of hematology, telemedicine has recently been used to safely and efficiently monitor treatment side-effects, perform consultations, and broaden the reach of subspecialty care.

Objective: We aimed to synthesize and analyze information regarding the feasibility, acceptability, and potential benefits of telemedicine interventions in malignant and nonmalignant hematology, as well as assess the recognized limitations of these interventions.

Methods: Studies were identified through a comprehensive Medical Subject Headings (MeSH) search on the PubMed MEDLINE, Controlled Register of Clinical Trials (Cochrane CENTRAL from Wiley), Embase, and CINAHL (EBSCO) databases on February 7, 2018. A second search, utilizing the same search strategy, was performed on October 1, 2020. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in the reporting of included evidence. Included studies were original articles researching the feasibility, acceptability, and clinical outcomes of telemedicine or telehealth interventions in pediatric or adult populations with malignant or nonmalignant hematological conditions. Data items in the extraction form included first author name, publication year, country, malignant or nonmalignant hematological condition or disease focus of the study, participant age, participant age subgroup (pediatric or adult), study design and setting, telemedicine intervention type and description, study purpose, and main study outcomes.

Results: A total of 32 articles met the preset criteria and were included in this study. Most (25/32) studies were conducted in adults, and the remaining (7/32) were conducted in the pediatric population. Of the 32 studies, 12 studied malignant hematological conditions, 18 studied nonmalignant conditions, and two studied both malignant and nonmalignant conditions. Study types included pilot study (11/32), retrospective study (9/32), randomized controlled trial (6/32), cross-sectional study (2/32), case study (1/32), pre-post study (1/32), noncomparative prospective study (1/32), and prospective cohort study (1/32). The three main types of telemedicine interventions utilized across all studies were video-based (9/32), telephone-based (9/32), and web-based interventions (14/32). Study results showed comparable outcomes between telemedicine and traditional patient encounter groups across both pediatric and adult populations for malignant and nonmalignant hematological conditions.

Conclusions: Evidence from this review suggests that telemedicine use in nonmalignant and malignant hematology provides similar or improved health care compared to face-to-face encounters in both pediatric and adult populations. Telemedicine interventions utilized in the included studies were well received in both pediatric and adult settings. However, more research is needed to determine the efficacy of implementing more widespread use of telemedicine for hematological conditions.

(JMIR Mhealth Uhealth 2021;9(7):e29619) doi:10.2196/29619

KEYWORDS

telemedicine; telehealth; eHealth; Digital Health; Digital Medicine; mHealth; hematology; malignant; nonmalignant; intervention

Introduction

Globally, we are facing a growing shortage of specialist physicians, coupled with inequity in patient access to quality care [1-3]. In America alone, only 30 specialists practice for every 100,000 people living in rural communities compared to 263 specialists in urban environments [4]. This shortage is further exacerbated when examined by subspecialty, making it progressively more difficult for patients to reach their physicians and obtain necessary treatments [4]. Thus, health care officials have turned to the field of telemedicine to leverage technological tools with the goal of expanding and optimizing the delivery of medical care [3].

Telemedicine is defined as the use of teleconferencing interventions to provide and deliver health care to patients [5]. Although telemedicine is classified under the broader term, "telehealth," ambiguity surrounding which forms of technology each category encompasses still exists, and both terms are often used interchangeably [6]. Telemedicine is a subset of telehealth, which is defined as the use of technology in any aspect of health care. Telemedicine is specifically utilized for clinical patient care, not exclusively for research purposes.

Telemedicine has grown in parallel with society's growing desire for convenience, efficiency, and productivity, and today, frequently used modes of telemedicine include videoconferencing, email, wearable devices, cellular phones, and various mobile apps [3,6]. These new systems of telemedicine are promising for mitigating the current challenges in health care reform, since they offer medical professionals the novel opportunity to extend their presence to settings outside of their immediate reach [3]. In doing so, telemedicine presents health care industries with the potential to provide more cost-effective treatments, support patient self-management, respond to the growing demand for specialists, and uncover avenues for advancing the practice of medicine in underserved areas worldwide [3,6-9]. Furthermore, access to personal and mobile technologies is ubiquitous [10-14], which has provided an opportunity to optimize digital health care delivery approaches, including telemedicine. However, despite these benefits, there remain obstacles to implementing telemedicine in daily practice [15,16]. Barriers, including stable internet access, cost, and patient desire for in-person appointments, must be addressed for telemedicine to reach its full potential [15,16]. Nevertheless, there has been growing evidence to support the utility and clinical applications of various digital approaches for health care delivery, including telemedicine, across pediatric and adult populations with or without chronic medical conditions [17-39], although the cost-effectiveness remains unclear [40,41].

In the field of hematology, recent advances in telemedicine have been used to conduct patient visits, monitor treatment side-effects, and perform consultations [42-45]. During the current COVID-19 pandemic, telemedicine use in outpatient settings, such as hematology clinics, has increased significantly owing to stay-at-home orders and efforts to lessen exposure to ill patients [46,47]. Videoconferencing interventions have proved to be a safe and efficient way for health care providers and patients to continue managing and monitoring chronic health conditions, especially for sickle cell disease and other hematological conditions [48]. Additionally, both adult and pediatric populations in underserved areas have benefitted from telemedicine's ability to connect them to specialists in different cities [48].

In this systematic review, we evaluated the evidence available in the literature to analyze the feasibility, acceptability, and potential benefits of telemedicine interventions in malignant and nonmalignant hematology and assess the current limitations of utilizing these interventions. Owing to the heterogeneity of the included studies, the methods by which feasibility, accessibility, and clinical outcomes were assessed vary among the studies. In this review, we use these terms broadly to encompass the different interpretations of these objectives. We more specifically define each study's main objectives in the Results section.

Methods

Guideline

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting of evidence across the studies we reviewed [49].

Article Retrieval

A librarian, in collaboration with other review authors, collaboratively developed the highly sensitive Medical Subject Headings (MeSH) term–based search strategies and ran searches in the following databases: PubMed MEDLINE, Controlled Register of Clinical Trials (Cochrane CENTRAL from Wiley), Embase, and CINAHL (EBSCO), on February 7, 2018 (Multimedia Appendix 1). The bibliographies of hand-searched articles that had been previously identified were also included. The search strategy focused on articles that studied telemedicine and telehealth interventions for malignant and nonmalignant hematological conditions in both pediatric and adult populations. No date limits were applied to the search. On October 1, 2020, another literature search was conducted on PubMed to identify articles that were published since the last literature search and that met the inclusion criteria. The second literature search utilized the same search strategy as the first search.

Study Selection

The inclusion criteria were as follows: (1) studies involving pediatric or adult populations with malignant or nonmalignant hematological conditions, (2) studies involving telemedicine or telehealth interventions, (3) studies that included feasibility, acceptability, and clinical outcomes of the interventions as the primary or secondary outcomes, (4) original research articles,

and (5) studies designed as a randomized controlled trial (RCT), cohort study, pilot study, retrospective study, or cross-sectional study. "Telemedicine" and "telehealth" are often used interchangeably, so both terms were relevant to the search despite having different semantic meanings. The exclusion criteria were as follows: (1) studies not related to hematology, (2) studies without a telemedicine intervention, (3) nonclinical research studies, (4) abstracts only or nonoriginal research papers, and (5) studies not in English.

Data Extraction

A standardized format was used for data extraction. Data items in the extraction form included first author name, publication year, country, malignant or nonmalignant hematological condition or disease focus of the study, participant age, participant age subgroup (adult or pediatric), study design and setting, telemedicine intervention type and description (telephone, remote management, videoconferencing, etc), study purpose, and main outcomes. Two authors coded all included articles individually. Disagreements were resolved by discussion.

Quality Assessment and Evidence Strength

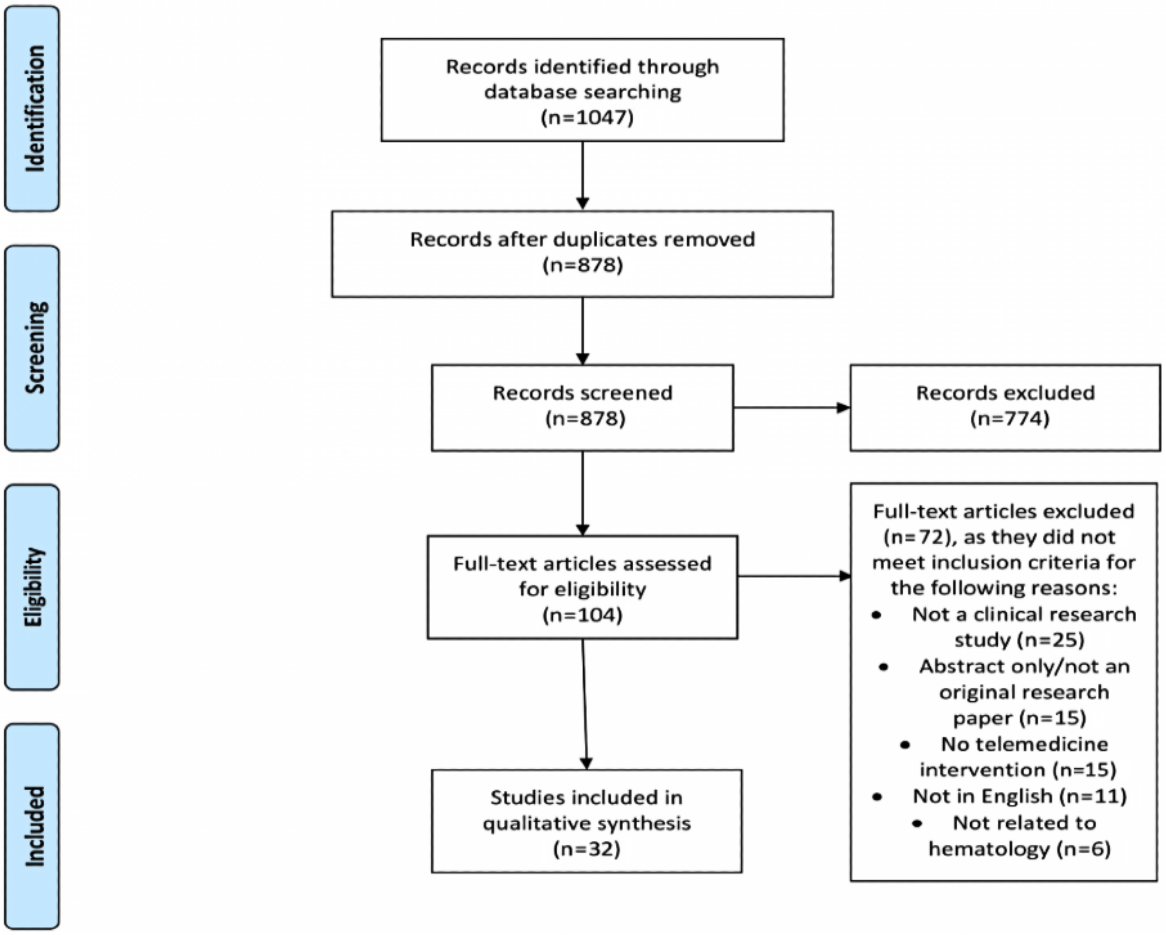
Studies described in each article were evaluated for the quality of evidence using the GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) approach [50]. This method evaluates four key domains, including consistency, directness, risk for bias, and precision of the evidence. Two authors graded all included articles individually. Disagreements were similarly resolved by discussion, if needed.

Results

Literature Search

Our literature search identified 1047 records. After removing duplicates, 878 articles remained. Two authors independently screened the titles and abstracts of 878 records, and 774 were excluded. Two authors independently screened the remaining 104 full-text articles, and 32 met our inclusion criteria for eligibility. The study flowchart and reasons for exclusion of full-text papers were documented in an adapted PRISMA study flowchart (Figure 1). We did not identify any non-English articles that met our inclusion criteria.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the included studies.



A total of 32 articles were included in this systematic review. Most (25/32, 78%) studies were conducted with adults [42,51-74], while the remaining (7/32, 22%) were conducted in the pediatric population [75-81]. The studies addressed whether a telemedical intervention for patients with malignant and/or nonmalignant hematological conditions is a safe, feasible, and efficacious method of health care provision.

Description of the Included Studies

Table 1 and Table 2 summarize study characteristics for the pediatric and adult populations, respectively. Out of the 32

studies, 16 (50%) were conducted in the United States [52,53,55-57,60,62,63,65,69,71-73,77,78,81], 6 (19%) in Canada [58,64,68,74,75,79], 2 (6%) each in Australia [42,61], India [51,76], and Italy [54,70], and 1 (3%) each in Brazil [80], France [59], Germany [66], Ireland [67], and Rwanda [81]. Although the search retrieval included results published since 1980, the earliest eligible study was from 1998, which corresponds to our knowledge of when telemedicine was first introduced [71]. The most recent study was published in 2018 [56].

The sample size ranged from 1 [74] to 217,014 [62], with a median of 157 and a mean of 7977 participants per study. Overall, 21 studies enrolled ≥100 participants [51,53-60,62,63,65-70,72,73,76,80] and 11 had <100 participants [42,52,61,64,71,74,75,77-79,81].

Of the 32 included studies, 12 (34%) studied malignant hematological conditions [42,53,58-61,64,67,69,77,80,81], 18 (59%) studied nonmalignant conditions with percentages [51,52,54-57,62,63,65,66,68,70-74,76,78], and 2 (6%) studied both malignant and nonmalignant conditions [75,79].

Table 1. Characteristics of pediatric malignant and nonmalignant hematology studies.

Source (country)	Condition	Sample size and age	Study design	Study setting	Grade
Adler 2015 [75] (Canada)	Leukemia, lymphoma, solid tumor, neuro-oncology, and nonmalignant hematology	N=54; age: 0-18 years	Pilot study	The Hospital for Sick Kids (Toronto, Canada) and six telemedicine facilities in Caribbean countries	Very low
Agarwal 2014 [76] (India)	Thalassemia	N=112	Pilot study	Participating centers in Italy, Pakistan, and India	Very low
Cox 2015 [77] (United States)	Survivors of childhood brain tumors or acute lymphoblastic leukemia	N=68; age: 8-16 years	Randomized controlled trial	St. Jude Children's Research Hospital	Moderate
Jacobson 2016 [78] (United States)	Severe hemophilia	N=12; age: 10-18 years, mean age: 10.25 years	Pilot study	Hemophilia Treatment Center	Very low
Johnston 2017 [79] (Canada)	Any hematology and oncology condition (majority anemia, hemoglobinopathy, and bleeding disorder)	N=85	Cross-sectional study	Champlain BASE e-Consult service (Ottawa, Canada) through a web portal	Low
Pedrosa 2017 [80] (Brazil)	Acute lymphoblastic leukemia	N=163; age: 1-15 years	Pilot study	Instituto Materno Infantil de Pernambuco (Recife, Brazil) and St. Jude Children's Research Hospital	Very low
Stulac 2016 [81] (United States, Rwanda)	Lymphoma, sarcoma, leukemia, and other malignancies	N=24	Retrospective study	Rural district hospitals in Rwanda (Rwinkwavu, Kirehe, and Butaro)	Very low

Table 2. Characteristics of adult malignant and nonmalignant hematology studies.

Source (country)	Condition	Sample size and age	Study design	Study setting	Grade
Agrawal 2014 [51] (India)	Blood transfusion donors	N=16,438 donations	Pilot study	Regional Blood Transfusion Centre (Dehradun, Uttarakhand)	Very low
Applebaum 2012 [52] (United States)	Posttraumatic stress disorder related to hematopoietic stem cell transplantation	N=46	Randomized controlled trial	Memorial Sloan-Kettering Cancer Center, Mount Sinai Medical Center, and Hackensack University Medical Center	Moderate
Bakitas 2015 [53] (United States)	Advanced-stage solid tumor or hematological malignancy with oncologist-determined prognosis of 6-24 months	N=207; early group mean age: 64.03 years, delayed group mean age: 64.6 years	Randomized controlled trial	National Cancer Institute Center and Veterans Affairs Medical Center	Moderate
Barcellona 2013 [54] (Italy)	Thromboembolic conditions	N=114; mean age: 61 years	Pre-post study	Thrombosis centers	Low
Blissit 2015 [55] (United States)	Thromboembolic conditions	N=200; face-to-face group (n=90) and telephone group (n=110)	Retrospective study	Veterans Affairs Medical Center and affiliated rural clinics in South Carolina	Very low
Breen 2015 [42] (Australia)	Hodgkin/non-Hodgkin lymphoma and chronic lymphocytic leukemia	N=18; mean age: 48.4 years	Pilot study	Chemotherapy day unit and hematology inpatient ward	Very low
Burwick 2018 [56] (United States)	Monoclonal gammopathy of undetermined significance	N=152; median age: 69 years	Retrospective study	Veterans Health Administration facilities	Very low
Cecchini 2016 [57] (United States)	Varying hematological conditions (majority anemia, thrombocytopenia, and venous thromboembolism)	N=909; e-consult (n=302), median age: 64 years; face-to-face consult before e-consult (n=305), median age: 69.3 years; face-to-face consult after e-consult (n=302), median age: 65.9 years	Retrospective study	Veterans Affairs Connecticut Healthcare System	Very low
Clarke 2011 [58] (Canada)	Cancer (majority gastrointestinal and lymphoma)	N=712	Retrospective study	48 Local Health Areas (British Columbia, Canada)	Very low
Compaci 2011 [59] (France)	Diffuse large B-cell lymphoma	N=100; median age: 57 years	Noncomparative prospective study	Toulouse University Hospital	Very low
Flannery 2009 [60] (United States)	Medical oncology and hematologic malignant diagnoses	N=5283; mean age 61.1 years	Descriptive retrospective study	University of Rochester Medical Center and James P. Wilmot Cancer Center	Very low
Hung 2014 [61] (Australia)	Hematological malignancies in patients who received peripheral blood stem cell transplants	N=37; usual care (n=19), age: 59.9 years; extended care (n=18), age: 57.5 years	Randomized controlled trial	The Hematology and Oncology Clinics of Australia, The Wesley Hospital (Brisbane, Australia)	Low
Kirsh 2015 [62] (United States)	Hematology and other specialties	N=217,014	Pilot study	Veterans Health Administration and Veterans Integrated Service Networks	Very low
Najafi 2017 [63] (United States)	Hematology and other specialties	N=313 (63 hematology)	Pilot study	University of California San Francisco Hospital in Mission Bay	Very low
Overend 2008 [64] (Canada)	Indolent and chronic hematological malignancies	N=53; median age: 66.5 years	Pilot study	British Columbia Cancer Agency's Centre for the Southern Interior in Kelowna	Very low
Philip 2015 [65] (United States)	Thromboembolic conditions	N=502	Pilot study	Harris Health System (Houston, Texas)	Very low
Prochaska 2017 [66] (Germany)	Thromboembolic conditions in patients receiving vitamin K antagonists	N=2221; median age: 73 years	Prospective cohort study	Center of Thrombosis and Hemostasis, University Medical Center Mainz	Very low

Source (country)	Condition	Sample size and age	Study design	Study setting	Grade
Reid 2011 [67] (Ireland)	Cancer	N=7498	Retrospective study	Regional Cancer Centre	Very low
Skeith 2017 [68] (Canada)	Thromboembolic conditions	N=162; mean age: 58.3 years	Cross-sectional study	Champlain Local Health Integration Network (Ontario, Canada)	Very low
Syrjala 2011 [69] (United States)	Cancer survivors treated with hemopoietic stem cell transplantation	N=775; mean age: 51.7 years	Randomized controlled trial	Fred Hutchinson Cancer Research Center (Seattle, Washington)	Moderate
Testa 2005 [70] (Italy)	Thromboembolic conditions	N=1393	Retrospective study	11 peripheral health care units and homes of 20 self-testing patients	Low
Woods 2000 [72] (United States)	Sickle cell disease	N=466; mean age: 27 years	Retrospective study	Telemedicine clinical sites of The Medical College of Georgia Sickle Cell Center (Augusta, Georgia)	Very low
Woods 1999 [73] (United States)	Sickle cell disease	N=120; standard (n=60), age: 33.32 years; telemedicine (n=60), age: 29.37 years	Randomized controlled trial	Telemedicine clinical sites of The Medical College of Georgia Sickle Cell Center (Augusta, Georgia)	Very low
Woods 1998 [71] (United States)	Sickle cell disease	N=28; mean age: 29.1 years	Pilot study	Remote telemedicine clinic site in Macon (affiliated with Medical Center of Central Georgia)	Very low
Wright 2007 [74] (Canada)	Allogenic blood and hemopoietic stem cell transplantation	N=1	Case study	Princess Margaret Hospital (Toronto, Ontario)	Very low

Methodological Quality of the Studies

Of the 32 studies, 11 (34%) were pilot studies [42,51,62-65,71,75,76,78,80], 9 (28%) were retrospective studies [55-58,60,67,70,72,81], 6 (19%) were RCTs [52,53,61,69,73,77], 2 (6%) were cross-sectional studies [68,79], 1 (3%) was a case study [74], 1 (3%) was a pre-post study [54], 1 (3%) was a noncomparative prospective study [59], and 1 (3%) was a prospective cohort study [66]. Of the 6 RCTs, 1 (17%) was a single-blinded study and the remaining 5 (83%) had no blinding. Follow-up activity after the telemedicine interventions was tracked in six of the studies in the forms of postintervention neuroimaging examination (n=1, 17%) [77], follow-up assessments (n=3, 50%) [52,53,64], number of patient visits (n=1, 17%) [62], and interview (n=1, 17%) [82]. Based on the GRADE criteria, four included studies were rated as "moderate" [52,53,69,77], four were rated as "low" [54,61,70,79], and 24 were rated as "very low" [42,51,55-60,62-68,71-76,78,80,81]. The "low" and "very low" ratings were mainly a result of the type of study and indirectness of evidence.

Description of Telemedicine Approaches

Table 3 and Table 4 provide detailed descriptions of the study purpose, telemedicine intervention used, and main findings of the included pediatric and adult studies, respectively. Additional study outcomes for pediatric and adult studies are also reported in Multimedia Appendix 2 and Multimedia Appendix 3, respectively. There were the following three main categories of telemedicine interventions in the included studies: video (n=9) [53,58,64,71-75,78], telephone (n=9) [52,55,59-61,65,67,76,81], and web (n=14) [42,51,54,56,57,62,63,66,68-70,77,79,80]. Video interventions used cameras and videoconferences to visually connect the patients to the providers. Telephone interventions involved nurse triages, counseling sessions over the phone, and telephone help lines available to patients 24/7. Web-based telemedicine interventions included online interfaces for patients to send and store data and to monitor their conditions from home while still maintaining intermittent contact with practitioners.

Table 3. Summary of the interventions and outcomes of the included studies in pediatric populations.

Source (country)	Purpose	Telemedicine intervention	Main findings
Adler 2015 [75] (Canada)	Improve outcomes and quality of life for children with cancer and blood disorders in the Caribbean by using telemedicine	Utilizing telemedicine facilities to conduct patient case consultations and discussions in real time	Satisfaction: Adequate to excellent overall attendee satisfaction rates Good to excellent patient care education satisfaction rates
Agarwal 2014 [76] (India)	Leverage an online collaborative information technology platform to improve outcomes of children with thalassemia receiving bone marrow transplants	Online open-access database with data storage, telemedicine, and knowledge exchange capabilities	Clinical outcomes: Comparable health outcomes with decreased cost Increased rates of family screening Targeted prenatal diagnoses for pregnant women
Cox 2015 [77] (United States)	Evaluate the feasibility and acceptability of a remote automated intervention to address late cognitive effects among childhood cancer survivors	Automated rotating exercises to train visual-spatial and verbal working memory over a 5- to 9-week period	Feasibility: Strong compliance with the intervention (88%) and pre- and postintervention imaging (91% and 93%, respectively) Satisfaction: Caregivers viewed the program as beneficial (70%) and would recommend it to others (93%)
Jacobson 2016 [78] (United States)	Evaluate the feasibility of using videoconferencing to assess breakthrough bleeds in children with severe hemophilia	Videoconferencing was utilized by health care providers to assess the patient's condition during breakthrough bleeding	Satisfaction: Comparable or improved satisfaction was reported among health care providers and caregivers with videoconferencing versus phone call
Johnston 2017 [79] (Canada)	Analyze the use of e-consult by primary care providers for pediatric oncology and hematology conditions	Web portal based e-consult service for pediatric oncology and hematology	Satisfaction: High satisfaction ratings of e-consult services from primary care providers and pediatric hematology/oncology specialists Feasibility: 40% deferral of consults that were originally supposed to be in person
Pedrosa 2017 [80] (Brazil)	Evaluate the efficacy of a telemedicine-based knowledge-sharing program to improve management of childhood acute lymphoblastic leukemia	Teleconferencing-based knowledge-sharing program between hospitals in the United States and Brazil	Clinical outcomes: Decreases in overall mortality (31.0% to 12.8%), early death (8.1% to 3.6%), and relapse (10.2% to 7.9%) after institution of the telemedicine program
Stulac 2016 [81] (United States/Rwanda)	Evaluate the impact of partnership-based treatment of pediatric cancer patients in Rwanda by physicians and nurses based in Rwanda and United States	Collaboration between clinicians based in Rwanda and United States via telephone, email, and online databases to manage pediatric cancer cases	Clinical outcomes: Mean overall survival of 31 months Mean disease-free survival of 18 months

Table 4. Summary of the interventions and outcomes of the included studies in adult populations.

Source (Country)	Purpose	Telemedicine intervention	Main findings
Agrawal 2014 [51] (India)	Improve blood donor recruitment, retention, and relationship management	Call center staffed with personnel to conduct phone calls and send SMS text messages to recruit and build relationships with blood donors	Feasibility: Telerecruitment contributed to 63% of in-house and 13% of total donations Clinical outcomes: Telerecruitment helped establish relationships with blood donors and the society in general
Applebaum 2012 [52] (United States)	Determine the relationship between therapeutic alliance through telephone delivered cognitive behavioral therapy and psychotherapy outcomes in survivors of hemopoietic stem cell transplantation with posttraumatic stress disorder	Telephone-administered cognitive behavioral therapy	Clinical outcomes: Therapeutic alliance by telephone cognitive behavior therapy predicted decreased depressive symptoms, decreased general distress, and lower likelihood for re-experiencing symptoms
Bakitas 2015 [53] (United States)	Investigate the effect of early versus delayed palliative care on the quality of life of advanced-stage cancer patients	Weekly telehealth nurse coaching sessions	Clinical outcomes: No statistically significant evidence to support improved patient-reported outcomes in early versus late palliative care groups Statistically significant improved 1-year survival rate in early versus late palliative care groups (P=.04)
Barcellona 2013 [54] (Italy)	Compare the effect of a point-of-care home monitoring testing device supplemented by telemedicine with conventional monitoring in patients chronically treated with anti-vitamin K antagonists	TAONet telemedicine platform allowed patients to send international normalized ratio (INR) results and other clinical information to the Thrombosis Centre, as well as communicate with providers, adjust medications, and screen for serious events	Clinical outcomes: Greater blood checks and fewer missed INR checks in the home monitoring group Significant increase in time spent within the therapeutic range in the unstable group with home monitoring compared to conventional monitoring No significant difference in time spent within the therapeutic range in the stable group with home monitoring compared to conventional monitoring
Blissit 2015 [55] (United States)	Compare the effect of telephone versus face-to-face care on time spent within the therapeutic range for patients on warfarin	Pharmacist-managed care via telephone for patients taking warfarin	Clinical outcomes: No significant difference in time spent within the therapeutic range, significant bleeding rates, death rates, and thromboembolic events between face-to-face and telephone groups
Breen 2015 [42] (Australia)	Evaluate a real-time remote telemedicine system to improve monitoring and management of side-effects in patients with blood cancers	Smart phone app collected ambulatory patient health data in real time and transmitted this information to the treatment hospital where alerts were generated for actioning based on imputed patient data	Satisfaction: Patients reported increased feelings of empowerment and health awareness and adherence with the use of the application Clinical outcomes: Patients were better able to recall side-effects when using the application.
Burwick 2018 [56] (United States)	Identify ways to improve care of patients with monoclonal gammopathy of undetermined significance through e-consult use	Review of electronic hematology consults for monoclonal gammopathy of undetermined significance through e-consult use	Accessibility: Short time (2 days) to completion of e-consult Majority of e-consults were low risk

Source (Country)	Purpose	Telemedicine intervention	Main findings
Cecchini 2016 [57] (United States)	Evaluate the efficacy of e-consults for the management of hematological disorders	Electronic consults with hematologists	Satisfaction: 65% of patients who responded said they preferred e-consults over face-to-face visits All providers who responded said they were either “satisfied” or “very satisfied” with e-consults Feasibility: 18% drop in face-to-face consults within 2 years of e-consult implementation
Clarke 2011 [58] (Canada)	Assess oncology telehealth usage in British Columbia	Oncology and genetic counselling appointments conducted with tele-conferencing units	Feasibility: Medical oncology teleconferences occurred more often than genetic counseling and medical genetics Clinical outcomes: A single medical oncologist conducted 58.7% of all telehealth encounters in 2009 Most common telehealth appointment types were gastrointestinal cancer and lymphoma
Compaci 2011 [59] (France)	Assess the feasibility and benefit of Ambulatory Medical Assistance phone calls in monitoring aggressive B-cell lymphoma treated with R-CHOP therapy	Standardized 10-minute telephone call twice a week by an oncology nurse to monitor vitals and side-effects during chemotherapy treatment	Clinical outcomes: Lower incidences of secondary hospitalization, delayed treatment, and reduced relative dose intensity, toxic death, and red blood cell transfusion compared to literature
Flannery 2009 [60] (United States)	Define telephone call volume, distribution, and reason in an ambulatory oncology practice	Telephone triage line managed by nurses to address patient symptoms	Feasibility: Seven calls were made or received for every 10 scheduled appointments Most calls were made on Monday mornings 30% of calls were made for more than one reason
Hung 2014 [61] (Australia)	Evaluate the impact of telephone counseling on nutrition and exercise after peripheral blood stem cell transplantation	Telephone counseling sessions conducted by a dietitian or exercise physiologist every 2 weeks for up to 100 days after transplantation	Clinical outcomes: Increased, but not statistically significant, protein intake (P=.17), cognitive functioning (P=.34), and social functioning (P=.17) in the extended care group versus usual care group Decreased, but not statistically significant, weight loss (P=.06) in the extended care group versus usual care group
Kirsh 2015 [62] (United States)	Analyze the impact of e-consults in improving specialty care access for veterans	E-consult service for various specialties	Feasibility: Hematology had the second highest (after cardiology) rate of e-consults Clinical outcomes: Within the first 3 months after an e-consult, there was a decreased likelihood of a subsequent face-to-face visit (P<.001) Within the first 3 months after an e-consult, there was an increased likelihood of a primary care visit (P<.001)

Source (Country)	Purpose	Telemedicine intervention	Main findings
Najafi 2017 [63] (United States)	Evaluate the feasibility and acceptability (by providers) of an inpatient e-consult service	E-consult service for subspecialist consults	Feasibility: Majority of hospitalists and consultants believe that e-consults are easy to use and efficient Hematology had the second highest volume of e-consults Satisfaction: All hospitalists who completed the survey were satisfied with quality 43% of consultants who completed the survey felt they were able to provide high-quality care
Overend 2008 [64] (Canada)	Determine the efficacy, safety, and patient satisfaction of a nurse-led teleclinic to manage patients with indolent and chronic hematological malignancies	Nurse-led teleclinic interviews for hematology/oncology patients in between oncologist visits	Satisfaction: 82% of patients felt strongly that they could talk easily and openly, and that the provider was able to understand their situation to provide satisfactory care 62% reported they would participate in a teleclinic again Accessibility: Majority of patients did not feel that they needed to see a physician in person
Philip 2015 [65] (United States)	Evaluate the efficacy of a telephone-based pharmacist anticoagulation service	Telephone-based anticoagulation service run by pharmacists	Clinical outcomes: Increase in clinical pharmacy patient volume at ambulatory care clinics after the intervention No significant difference in time in the therapeutic range, hospitalization from thrombotic events or bleeding, work hours, and project completion rates between groups
Prochaska 2017 [66] (Germany)	Compare the outcomes of oral anticoagulation patients managed by an eHealth-based coagulation service versus regular medical care	Electronic file was used to manage medication and data at remote locations and interface between patients and providers	Clinical outcomes: eHealth service participants had lower rates of bleeding, hospitalization, and all-cause mortality compared to participants who received regular medical care
Reid 2011 [67] (Ireland)	Investigate the usage and patient/caller profile of a nurse-led chemotherapy telephone helpline	Telephone service run by nurses experienced in oncology and chemotherapy to address patient questions and concerns	Feasibility: 7498 calls received by helpline service 35.2% of patients called with multiple symptoms or concerns Accessibility: 47.5% of face-to-face consultations were avoided 4.3% of e-consults resulted in a follow-up referral that was not already scheduled Clinical outcomes: 36.8% of calls led to direct medical assessment
Skeith 2017 [68] (Canada)	Analyze the use and impact of e-consults in managing thromboembolic conditions	E-consult service for thrombosis medicine	Feasibility: Most common referral topics were thrombophilia testing, superficial venous thrombosis, and venous thromboembolism anticoagulation Satisfaction: Positive responses by primary care providers regarding the e-consult service

Source (Country)	Purpose	Telemedicine intervention	Main findings
Syrjala 2011 [69] (United States)	Describe the development and feasibility of a new online survivorship care program for cancer patients treated with hemopoietic stem cell transplantation	Survivorship care delivered remotely through an online platform	Feasibility: 57% of participants required staff contact at least one time, usually for minor technical issues, or help in enrollment or the baseline assessment Most contacts were initiated by email instead of telephone
Testa 2005 [70] (Italy)	Implement a telemedicine system to deliver the same quality of care as traditional medicine in anticoagulation management, especially for those living far away from an anticoagulation center	Remote anticoagulation management via an electronic medical record system through which patients or health centers can send INR data to an anticoagulation clinic	Clinical outcomes: No difference in time in the therapeutic range between anticoagulation clinics (73%) and telemedicine use in general practitioner units (73.4%) Telemedicine use in nursing homes showed a lower percentage in the therapeutic range (66%) compared to anticoagulation clinics (73%) No difference in major complication rates between telemedicine use in peripheral units and anticoagulation clinics
Woods 2000 [72] (United States)	Evaluate the efficacy of the Georgia statewide telemedicine program in improving access to health services for patients with sick cell disease in remote areas	Telemedicine clinic for sickle cell patients with assistance of nurses at remote locations	Feasibility: Progressive increase in telemedicine clinic productivity over the course of the study Accessibility: Rural outreach increased from 19% to 29% of total clinic activity over the course of the study
Woods 1999 [73] (United States)	Compare patient satisfaction between telemedicine encounters and standard care for the management of sickle cell disease in adults	Telemedicine clinic for sickle cell patients with assistance of nurses at remote locations	Satisfaction: No significant difference in patient satisfaction between telemedicine and standard care groups Patients in the standard care group were more likely to offer positive open-ended comments than the telemedicine group Reasons for negative comments in the telemedicine group included confidentiality, technology, and access
Woods 1998 [71] (United States)	Evaluate the efficacy of a telemedicine clinic for adult sickle cell patients	Telemedicine clinic for sickle cell patients with assistance of nurses at remote locations	Clinical outcomes: No significant difference in clinic encounter time between telemedicine and standard care groups
Wright 2007 [74] (Canada)	Evaluate the efficacy and acceptability of an e-clinic for management of allogeneic blood and stem cell transplant patients	E-clinic visits at local health centers using videoconferencing technology	Clinical outcomes: High adherence rate for follow-up visits (87.5%) Improved symptom management Satisfaction: High satisfaction with e-clinic visits

Study Outcomes in Video Telemedicine Interventions

Pediatric Video Interventions

Adler et al reported greater than 80% case review round attendee satisfaction with telemedicine case consultations and patient care education for children with cancer and blood disorders [75]. Improved satisfaction and communication among health care providers and caregivers with the use of videoconferencing at home for children with hemophilia were also noted by Jacobson et al [78].

Adult Video Interventions

Many telehealth nursing sessions have shown positive outcomes for patients with malignant hematological conditions [53]. Likewise, patients with hematological malignancies felt that they could talk easily and openly with the nurse who led telemedicine follow-up visits in between visits with the oncologist [64]. Telemedicine reportedly improved productivity [72] and increased mean cognitive skills quotient scores for patients [73], but caused no significant decrease in clinic encounter times [71]. Wright et al reported high follow-up

adherence rates, good condition management, and high satisfaction for e-clinic visits in allogenic blood and stem cell transplant patients [74]. Clarke et al found that teleconsults in British Columbia were most often used with medical oncologists [58].

Study Outcomes in Telephone Telemedicine Interventions

Pediatric Telephone Interventions

Agarwal et al found comparable health outcomes at a lower cost through the implementation of an online information technology database platform for patients receiving bone marrow transplants [76]. Stulac et al found that 13 of 24 Rwandan patients with pediatric cancer were in remission at the end of data collection following collaboration with United States–based physicians and nurses through telemedicine interventions [81].

Adult Telephone Interventions

Telephone therapy was utilized by Applebaum et al, which resulted in decreased depressive symptoms in hematopoietic stem cell transplant survivors with concurrent posttraumatic stress disorder [52]. For patients taking anticoagulants, Blissit et al noticed that telephone interventions led to significantly lower rates of bleeding in patients compared to face-to-face visits [55]. However, Philip et al did not see a significant difference in time within the therapeutic range, hospitalization, or bleeding with a telephone-based anticoagulation service [65]. Reid et al and Flannery et al utilized phone services to address patient questions and symptoms [60,67]. Reid et al noted that 36.8% of calls led to a medical assessment [67]; however, Flannery et al reported no notable improvements with the telephone triage [60]. Biweekly calls for monitoring vitals, side-effects [59], and telephone counseling [61] have all shown benefits for patients with malignant hematological conditions.

Study Outcomes in Web-Based Telemedicine Interventions

Pediatric Web-Based Interventions

Cox et al found that patients, families, and caregivers had high rates of intervention participation and reported a positive experience with the introduction of a remote automated working memory intervention for childhood cancer survivors [77]. Telemedicine interventions also received high ratings from primary care providers and resulted in hematology consult deferral, according to Johnston et al [79]. Similarly, Pedrosa et al reported positive outcomes, with decreased rates of mortality, early death, and acute lymphoblastic leukemia relapse following implementation of a telemedicine program that allowed for knowledge sharing between high- and low-income countries [80].

Adult Web-Based Interventions

Based on a survey by Checchini et al, more than half of patients with nonmalignant hematological disorders preferred e-consults over face-to-face visits, and satisfaction was noted in all providers [57]. Breen et al found increased feelings of empowerment, health awareness, and medication adherence with the use of a telemedicine-based smartphone app that directly transmitted health data to the hospital for patients with blood cancers [42]. Najafi et al also noted that the majority of hospitalists and consultants found e-consult to be a feasible administration of subspecialty health services at a remotely-located hospital [63]. Primary care providers were satisfied with e-consult technology in a survey by Skeith et al, especially since 47.5% of in-person visits were avoided through the use of this telemedicine intervention [68]. Kirsh et al also reported a decreased number of face-to-face visits in patients using e-consult telemedicine interventions [62]. Additionally. Burwick et al saw that e-consults for hematology cases in patients with monoclonal gammopathy of undetermined significance led to a decrease in the time it took to complete the consultation [56]. In a study by Syrjala et al that included cancer patients treated with hematopoietic stem cell transplantation, over half of participating patients used an online platform and reported the need for staff contact one or more times per day, usually for minor assistance [69]. Barcellona et al reported an increased number of blood checks, fewer missed international normalized ratio checks in the home monitoring group, and a significant increase in the time within the therapeutic range for patients treated with vitamin K antagonists when using a smartphone app to monitor treatment [54]. Vitamin K antagonist users also had lower rates of bleeding, hospitalization, and all-cause mortality with the use of an online electronic file [66]. Electronic medical record implementation for remote anticoagulation management, however, showed no significant difference in the time within the therapeutic range or major complication rates in nursing homes [70]. Agrawal et al found that a call center allowed for increased recruitment of blood donors [51].

Discussion

Principal Findings

Overall, the vast majority of telemedicine interventions studied in this systematic review were shown to have a positive or neutral impact on patients, families, and health care providers. Telemedicine was found to be particularly useful for rural communities, patients in countries with less access to care, and patients with chronic conditions that require routine monitoring and communication with doctors [53,64,71-73,75,78,80,81]. Most studies showed improved or similar outcomes in groups that utilized telemedicine compared to those that did not. However, more research would be beneficial to determine telemedicine's role in hematology and future implications for its use in the clinical setting. This evidence could also potentially increase patient satisfaction and patient-reported health outcomes.

Telemedicine Interventions in Hematology: Expansion of Team-Based Medicine

In the realm of hematology, many new telemedicine interventions have emerged in recent years to expand the reach of specialty health care and improve patient outcomes [83,84]. Kulkarni et al reported that school-based telemedicine has provided access to multidisciplinary teams for the treatment of hemophilia and related bleeding disorders [84]. In patients who

have undergone hematopoietic stem cell transplantation, therapy compliance and surveillance are seen to be good areas to implement telemedicine [83].

Telemedicine Interventions in Hematology: Areas of Further Study

Future use of telemedicine in hematology and other subspecialties should focus on increasing the usage of videoconferencing with physicians from patient homes and making telemedicine more accessible to a wider range of patient populations [85]. This is especially the case for rural areas or regions where specialty health care and access to medical resources are limited [86]. The current COVID-19 pandemic has shown the value of maintaining and developing different ways of patient care, such as through the increased implementation of telemedicine in clinics and hospitals [87]. Various specialties should investigate ways in which they can best implement telemedicine interventions into their own field to improve access and care [86]. The next steps should include the wide use of triage systems that allow patients to be screened for telemedicine use eligibility [87].

Telemedicine Interventions in Other Medical Specialties

In addition to the use of telemedicine in hematological conditions, it is widely used in various other medical specialties [88-93]. Meta-analyses on telemental health care have reported that applying telemedicine to psychotherapy not only improves patient satisfaction and access to care by circumventing the stigma of seeking in-person mental health care, but also offers a more financially prudent and flexible method of treatment [88]. Telemedicine has also been reported to contribute to shorter lengths of hospital stay and lower hospital mortality [90]. Multiple studies have shown it to be a cost-effective and accessible provision of mental health services [88]. Positive user outcomes with telehealth implementation in emergency rooms have indicated another area of potential use [91]. Similarly, telemedicine has been seen to help in the secondary prevention of stroke and in transient ischemia attack patients, but more research is needed to bolster evidence [93]. Cancer survivors generally had positive attitudes toward self-management and eHealth, but there was variation in the care needs depending on the type of cancer [92]. Telemonitoring is rarely used in prenatal care, but has potential in monitoring pregnant women who are at risk for preterm delivery [89]. These studies emphasize the positive experiences patients and providers have had with telemedicine in various fields of medicine, which further suggests that telemedicine is feasible, accessible, and beneficial. Many of the telemedicine interventions used in these other medical specialties can also be applied to the setting of hematology.

Barriers to Telemedicine

Though studies indicate high reports of satisfaction from patients and providers in malignant and nonmalignant hematology [57,68,79,94], the incorporation of telemedicine in modern medicine still faces limitations and organizational barriers, including legal liability, cost and reimbursement, and confidentiality concerns [15]. Patient barriers, such as age, computer literacy, and education, also make telemedicine difficult to implement universally [15]. Additionally, technological barriers, including inadequate bandwidth, technically challenged staff, and licensing issues, can limit the scope of telemedicine [15]. There are also concerns regarding the availability of resources in languages other than English [82], lack of notable improvements with telephone triaging [60], efficiency of face-to-face visits [71], and quality of care delivered via e-consult [63]. Funding for both staff and equipment, privacy concerns, internet connection, and home recording all need to be addressed when considering the feasibility of telemedicine [15,95-97]. Since the cost-effectiveness of telemedicine varies across delivery settings [9], more research needs to be done to substantiate evidence of improvement in patient outcomes with the use of telemedicine. Furthermore, the ethics of telemedicine and digital health use in the clinical setting must be discussed. The ethical and legal concerns regarding the use of telemedicine have delayed the rapid widespread implementation of these technologies in multiple aspects of health care [98-100]. Specifically, data confidentiality, patient privacy, physician-patient relationships, and informed consent are all areas that need to be addressed [99,101-103]. Further work must be done to ensure that data management is conducted properly and the social implications or risks are communicated clearly to patients and physicians using telemedicine in practice [99,101-103]. However, despite its barriers, the ability of telemedicine to mediate many of the shortcomings associated with traditional face-to-face consultations still makes telemedicine particularly attractive to many fields, especially hematology.

Strengths

Our systematic review has a number of strengths. First, we conducted our review following the recommendations for rigorous systematic reviews [49]. Second, we used a highly sensitive and specific search strategy guided by a librarian information specialist. There was no restriction on the country of study in order to minimize publication bias by identifying as many relevant studies as possible. Additional resources were searched, including published systematic reviews, clinical trial registries, and multiple electronic databases. Third, we employed no date restrictions on our search; no included articles were published prior to 1998. Therefore, the possibility that we missed earlier studies is very small. Finally, two authors completed the review process independently at all stages.

Limitations

Some potential methodological limitations of our systematic review must be addressed. First, some relevant articles could have been missed in our literature search, despite our comprehensive search strategy in different databases. Second, articles included in the review were strictly peer-reviewed, which could cause publication bias from reporting only positive study results [104]. Third, the ranges of sample size and age, differences between malignant and nonmalignant conditions, and types of telemedicine interventions varied greatly among the studies. Lastly, the number of studies eligible for the review was relatively low; however, this may be a result of the

specificity of hematological conditions and the limited number of available publications about the topic.

Conclusions

Telemedicine is a recognized and cost-effective way of managing hematological conditions. Evidence from this systematic review suggests that telemedicine provides similar or improved health care compared to traditional face-to-face care. Videoconferencing, telephone-based services, and web-based services were also well received by patients, families, and health care providers in both pediatric and adult settings. However, due to the limited total number of articles and low quality of evidence of the included studies, further research must be done to determine the efficacy and plausibility of widespread implementation of telemedicine in hematology.

Acknowledgments

This project was supported by a grant (K23HL150232, PI: SMB) from the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the views of the NIH. We thank Mary Therese Forsyth, BS (UCD School of Medicine, University College Dublin, Dublin, Ireland) and Bryn Dougherty, BS (Ohio State University College of Medicine, Columbus, OH) for their help in different parts of the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategies across different databases.
[DOCX File , 20 KB - mhealth_v9i7e29619_app1.docx]

Multimedia Appendix 2
Additional results of the included studies in pediatric populations.
[DOCX File , 18 KB - mhealth_v9i7e29619_app2.docx]

Multimedia Appendix 3
Additional results of the included studies in adult populations.
[DOCX File , 25 KB - mhealth_v9i7e29619_app3.docx]

References

1. Culmer N, Smith T, Stager C, Meyer H, Quick S, Grimm K. Evaluation of the triple aim of medicine in prehospital telemedicine: A systematic literature review. J Telemed Telecare 2020 Dec;26(10):571-580. [doi: 10.1177/1357633X19853461] [Medline: 31238783]
2. Tomines A. Pediatric Telehealth: Approaches by Specialty and Implications for General Pediatric Care. Adv Pediatr 2019 Aug;66:55-85. [doi: 10.1016/j.yapd.2019.04.005] [Medline: 31230700]
3. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
4. Alonso W, Crouch E, Thorell N. National Rural Health Association Policy Paper: Telehealth in Rural America. National Rural Health Association. URL: https://www.ruralhealthweb.org/NRHA/media/Emerge_NRHA/Advocacy/Policy%20documents/2019-NRHA-Policy-Document-Telehealth-In-Rural-America.pdf [accessed 2021-02-01]
5. Bashshur RL. On the definition and evaluation of telemedicine. Telemed J 1995;1(1):19-30. [doi: 10.1089/tmj.1.1995.1.19] [Medline: 10165319]
6. Telemedicine: Opportunities and Developments in Member States. World Health Organization. 2010. URL: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2021-02-13]
7. AMA Telemedicine Policy. American Medical Association. 2017. URL: https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/arc-public/telemed-policy.pdf [accessed 2021-01-31]
8. Field MJ. Telemedicine: a guide to assessing telecommunications in healthcare. J Digit Imaging 1997 Aug;10(3 Suppl 1):28 [FREE Full text] [doi: 10.1007/BF03168648] [Medline: 9268830]
9. Wilson LS, Maeder AJ. Recent Directions in Telemedicine: Review of Trends in Research and Practice. Healthc Inform Res 2015 Oct;21(4):213-222 [FREE Full text] [doi: 10.4258/hir.2015.21.4.213] [Medline: 26618026]
10. Badawy SM, Thompson AA, Liem RI. Technology Access and Smartphone App Preferences for Medication Adherence in Adolescents and Young Adults With Sickle Cell Disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]

11. Heneghan MB, Hussain T, Barrera L, Cai SW, Haugen M, Morgan E, et al. Access to Technology and Preferences for an mHealth Intervention to Promote Medication Adherence in Pediatric Acute Lymphoblastic Leukemia: Approach Leveraging Behavior Change Techniques. J Med Internet Res 2021 Feb 18;23(2):e24893 [FREE Full text] [doi: 10.2196/24893] [Medline: 33599621]
12. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2014;38(2):99-103. [doi: 10.3109/03630269.2014.880716] [Medline: 24512633]
13. Perrin A. 10 facts about smartphones as the iPhone turns 10. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/ [accessed 2021-02-14]
14. Perrin A. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-02-15]
15. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
16. Burke BL, Hall RW, Section on telehealth care. Telemedicine: Pediatric Applications. Pediatrics 2015 Jul;136(1):e293-e308 [FREE Full text] [doi: 10.1542/peds.2015-1517] [Medline: 26122813]
17. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text Messaging and Mobile Phone Apps as Interventions to Improve Adherence in Adolescents With Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
18. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
19. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
20. Badawy SM, Morrone K, Thompson A, Palermo TM. Computer and mobile technology interventions to promote medication adherence and disease management in people with thalassemia. Cochrane Database Syst Rev 2019 Jun 28;6:CD012900 [FREE Full text] [doi: 10.1002/14651858.CD012900.pub2] [Medline: 31250923]
21. Badawy SM, Radovic A. Digital Approaches to Remote Pediatric Health Care Delivery During the COVID-19 Pandemic: Existing Evidence and a Call for Further Research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
22. Radovic A, Badawy SM. Technology Use for Adolescent Health and Wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194 [FREE Full text] [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
23. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
24. Alberts NM, Badawy SM, Hodges J, Estepp JH, Nwosu C, Khan H, et al. Development of the InCharge Health Mobile App to Improve Adherence to Hydroxyurea in Patients With Sickle Cell Disease: User-Centered Design Approach. JMIR Mhealth Uhealth 2020 May 08;8(5):e14884 [FREE Full text] [doi: 10.2196/14884] [Medline: 32383683]
25. Badawy SM, Shah R, Beg U, Heneghan MB. Habit Strength, Medication Adherence, and Habit-Based Mobile Health Interventions Across Chronic Medical Conditions: Systematic Review. J Med Internet Res 2020 Apr 28;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]
26. Shah AC, Badawy SM. Telemedicine in Pediatrics: Systematic Review of Randomized Controlled Trials. JMIR Pediatr Parent 2021 Feb 24;4(1):e22696 [FREE Full text] [doi: 10.2196/22696] [Medline: 33556030]
27. 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]
28. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013 Dec 05;12(12):CD007458 [FREE Full text] [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
29. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266 [FREE Full text] [doi: 10.1371/journal.pone.0104266] [Medline: 25100267]
30. Liu Q, Abba K, Alejandria MM, Sinclair D, Balanag VM, Lansang MAD. Reminder systems to improve patient adherence to tuberculosis clinic appointments for diagnosis and treatment. Cochrane Database Syst Rev 2014 Nov 18(11):CD006594 [FREE Full text] [doi: 10.1002/14651858.CD006594.pub3] [Medline: 25403701]
31. 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]

32. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
33. Badawy SM. Technology and Long-Term Health-Related Quality-of-Life Outcomes in Children with Nonmalignant Disorders after Reduced-Intensity Conditioning and Stem Cell Transplantation. Biol Blood Marrow Transplant 2016 Sep;22(9):1733 [FREE Full text] [doi: 10.1016/j.bbmt.2016.06.015] [Medline: 27328128]
34. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
35. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
36. Badawy SM, Thompson AA, Kuhns LM. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
37. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of Mobile Apps to Support the Care of Patients With Diabetes Mellitus: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
38. 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 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
39. Johnson A, Yang F, Gollarahalli S, Banerjee T, Abrams D, Jonassaint J, et al. Use of Mobile Health Apps and Wearable Technology to Assess Changes and Predict Pain During Treatment of Acute Pain in Sickle Cell Disease: Feasibility Study. JMIR Mhealth Uhealth 2019 Dec 02;7(12):e13671 [FREE Full text] [doi: 10.2196/13671] [Medline: 31789599]
40. Badawy SM, Kuhns LM. Economic Evaluation of Text-Messaging and Smartphone-Based Interventions to Improve Medication Adherence in Adolescents with Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
41. 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]
42. Breen S, Ritchie D, Schofield P, Hsueh Y, Gough K, Santamaria N, et al. The Patient Remote Intervention and Symptom Management System (PRISMS) - a Telehealth- mediated intervention enabling real-time monitoring of chemotherapy side-effects in patients with haematological malignancies: study protocol for a randomised controlled trial. Trials 2015 Oct 19;16:472 [FREE Full text] [doi: 10.1186/s13063-015-0970-0] [Medline: 26481873]
43. Fox P, Darley A, Furlong E, Miaskowski C, Patiraki E, Armes J, et al. The assessment and management of chemotherapy-related toxicities in patients with breast cancer, colorectal cancer, and Hodgkin's and non-Hodgkin's lymphomas: A scoping review. Eur J Oncol Nurs 2017 Feb;26:63-82. [doi: 10.1016/j.ejon.2016.12.008] [Medline: 28069154]
44. Fisher SI, Nandedkar MA, Williams BH, Abbondanzo SL. Telehematopathology in a clinical consultative practice. Hum Pathol 2001 Dec;32(12):1327-1333. [doi: 10.1053/hupa.2001.29654] [Medline: 11774165]
45. Doyle-Lindrud S. Telemedicine in Oncology. Clin J Oncol Nurs 2016 Feb;20(1):27-28. [doi: 10.1188/16.CJON.27-28] [Medline: 26800402]
46. Indini A, Aschele C, Cavanna L, Clerico M, Daniele B, Fiorentini G, et al. Reorganisation of medical oncology departments during the novel coronavirus disease-19 pandemic: a nationwide Italian survey. Eur J Cancer 2020 Jun;132:17-23. [doi: 10.1016/j.ejca.2020.03.024] [Medline: 32311643]
47. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
48. Jacob SA, Carroll AE, Bennett WE. A feasibility study of telemedicine for paediatric sickle cell patients living in a rural medically underserved area. J Telemed Telecare 2019 Nov 06:1357633X19883558. [doi: 10.1177/1357633X19883558] [Medline: 31694482]
49. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
50. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
51. Agrawal A, Tiwari A. Tele-recruitment for Donor Retention. Indian J Hematol Blood Transfus 2014 Mar;30(1):22-25 [FREE Full text] [doi: 10.1007/s12288-012-0197-0] [Medline: 24554816]

52. Applebaum AJ, DuHamel KN, Winkel G, Rini C, Greene PB, Mosher CE, et al. Therapeutic alliance in telephone-administered cognitive-behavioral therapy for hematopoietic stem cell transplant survivors. J Consult Clin Psychol 2012 Oct;80(5):811-816 [FREE Full text] [doi: 10.1037/a0027956] [Medline: 22468908]
53. Bakitas MA, Tosteson TD, Li Z, Lyons KD, Hull JG, Li Z, et al. Early Versus Delayed Initiation of Concurrent Palliative Oncology Care: Patient Outcomes in the ENABLE III Randomized Controlled Trial. J Clin Oncol 2015 May 01;33(13):1438-1445 [FREE Full text] [doi: 10.1200/JCO.2014.58.6362] [Medline: 25800768]
54. Barcellona D, Fenu L, Cornacchini S, Marongiu F. Telemedicine can improve the quality of oral anticoagulation using portable devices and self-testing at home. J Telemed Telecare 2013 Sep;19(6):298-301. [doi: 10.1177/1357633X13501764] [Medline: 24163291]
55. Blissit KT, Mullenix ML, Brittain KG. Evaluation of Time in Therapeutic Range on Warfarin Therapy Between Face-to-Face and Telephone Follow-Up in a VA Medical Center. Journal of Pharmacy Technology 2014 Oct 01;31(2):78-83. [doi: 10.1177/8755122514552497]
56. Burwick N, Stein J, Garcia DA, Broudy VC, Richard RE. Monoclonal gammopathies: Electronic subspecialty consultation. Eur J Haematol 2018 Apr;100(4):351-355. [doi: 10.1111/ejh.13019] [Medline: 29281132]
57. Cecchini M, Rose MG, Wong EY, Neparidze N. The implementation of electronic hematology consults at a VA hospital. Blood 2016 Mar 24;127(12):1610-1611 [FREE Full text] [doi: 10.1182/blood-2015-09-672113] [Medline: 26809506]
58. Clarke M, Barnett J. Teleoncology uptake in British Columbia. Stud Health Technol Inform 2011;164:399-404. [Medline: 21335743]
59. Compaci G, Ysebaert L, Obéric L, Derumeaux H, Laurent G. Effectiveness of telephone support during chemotherapy in patients with diffuse large B cell lymphoma: the Ambulatory Medical Assistance (AMA) experience. Int J Nurs Stud 2011 Aug;48(8):926-932. [doi: 10.1016/j.ijnurstu.2011.01.008] [Medline: 21349519]
60. Flannery M, Phillips SM, Lyons CA. Examining telephone calls in ambulatory oncology. J Oncol Pract 2009 Mar;5(2):57-60 [FREE Full text] [doi: 10.1200/JOP.0922002] [Medline: 20856719]
61. Hung Y, Bauer JD, Horsely P, Coll J, Bashford J, Isenring EA. Telephone-delivered nutrition and exercise counselling after auto-SCT: a pilot, randomised controlled trial. Bone Marrow Transplant 2014 Jun;49(6):786-792. [doi: 10.1038/bmt.2014.52] [Medline: 24710562]
62. Kirsh S, Carey E, Aron DC, Cardenas O, Graham G, Jain R, et al. Impact of a national specialty e-consultation implementation project on access. Am J Manag Care 2015 Dec 01;21(12):e648-e654 [FREE Full text] [Medline: 26760427]
63. Najafi N, Harrison JD, Duong J, Greenberg A, Cheng HQ. It All Just Clicks: Development of an Inpatient E-Consult Program. J Hosp Med 2017 May;12(5):332-334. [doi: 10.12788/jhm.2740] [Medline: 28459902]
64. Overend A, Khoo K, Delorme M, Krause V, Avanessian A, Saltman D. Evaluation of a nurse-led telephone follow-up clinic for patients with indolent and chronic hematological malignancies: a pilot study. Can Oncol Nurs J 2008;18(2):64-73. [doi: 10.5737/1181912x1826468] [Medline: 18649698]
65. Philip A, Green M, Hoffman T, Gautreaux S, Wallace D, Roux R, et al. Expansion of clinical pharmacy through increased use of outpatient pharmacists for anticoagulation services. Am J Health Syst Pharm 2015 Apr 01;72(7):568-572. [doi: 10.2146/ajhp140404] [Medline: 25788511]
66. Prochaska JH, Göbel S, Keller K, Coldewey M, Ullmann A, Lamparter H, et al. e-Health-based management of patients receiving oral anticoagulation therapy: results from the observational thrombEVAL study. J Thromb Haemost 2017 Jul;15(7):1375-1385 [FREE Full text] [doi: 10.1111/jth.13727] [Medline: 28457013]
67. Reid J, Porter S. Utility, caller, and patient profile of a novel Chemotherapy Telephone Helpline service within a regional cancer centre in Northern Ireland. Cancer Nurs 2011;34(3):E27-E32. [doi: 10.1097/NCC.0b013e318204c53c] [Medline: 21512342]
68. Skeith L, Mohamed M, Karovitch A, Liddy C, Afkham A, Archibald D, et al. The use of eConsults to improve access to specialty care in thrombosis medicine. Thromb Res 2017 Dec;160:105-108. [doi: 10.1016/j.thromres.2017.11.002] [Medline: 29145030]
69. Syrjala KL, Stover AC, Yi JC, Artherholt SB, Romano EM, Schoch G, et al. Development and implementation of an Internet-based survivorship care program for cancer survivors treated with hematopoietic stem cell transplantation. J Cancer Surviv 2011 Sep;5(3):292-304 [FREE Full text] [doi: 10.1007/s11764-011-0182-x] [Medline: 21544671]
70. Testa S, Paoletti O, Zimmermann A, Bassi L, Cancellieri E. Telemedicine For Managing Patients on Oral Anticoagulant Therapy. Telemedicine Techniques and Applications. 2011. URL: https://www.intechopen.com/books/telemedicine-techniques-and-applications/telemedicine-for-managing-patients-on-oral-anticoagulant-therapy [accessed 2021-07-05]
71. Woods K, Kutlar A, Grigsby RK, Adams L, Stachura ME. Primary-care delivery for sickle cell patients in rural Georgia using telemedicine. Telemed J 1998;4(4):353-361. [doi: 10.1089/tmj.1.1998.4.353] [Medline: 10220476]
72. Woods KF, Johnson JA, Kutlar A, Daitch L, Stachura ME. Sickle cell disease telemedicine network for rural outreach. J Telemed Telecare 2000;6(5):285-290. [doi: 10.1258/1357633001935923] [Medline: 11070590]
73. Woods KF, Kutlar A, Johnson JA, Waller JL, Grigsby RK, Stachura ME, et al. Sickle cell telemedicine and standard clinical encounters: a comparison of patient satisfaction. Telemed J 1999;5(4):349-356. [doi: 10.1089/107830299311916] [Medline: 10908450]

74. Wright J, Purdy B, McGonigle S. E-Clinic: an innovative approach to complex symptom management for allogeneic blood and stem cell transplant patients. Can Oncol Nurs J 2007;17(4):187-192. [Medline: 18286991]
75. Adler E, Alexis C, Ali Z, Allen U, Bartels U, Bick C, et al. Bridging the Distance in the Caribbean: Telemedicine as a means to build capacity for care in paediatric cancer and blood disorders. Stud Health Technol Inform 2015;209:1-8. [Medline: 25980698]
76. Agarwal RK, Sedai A, Dhimal S, Ankita K, Clemente L, Siddique S, et al. A prospective international cooperative information technology platform built using open-source tools for improving the access to and safety of bone marrow transplantation in low- and middle-income countries. J Am Med Inform Assoc 2014;21(6):1125-1128 [FREE Full text] [doi: 10.1136/amiajnl-2013-002594] [Medline: 24714444]
77. Cox LE, Ashford JM, Clark KN, Martin-Elbahesh K, Hardy KK, Merchant TE, et al. Feasibility and acceptability of a remotely administered computerized intervention to address cognitive late effects among childhood cancer survivors. Neurooncol Pract 2015 Jun;2(2):78-87 [FREE Full text] [doi: 10.1093/nop/npu036] [Medline: 27054041]
78. Jacobson K, Hooke MC. Telehealth Videoconferencing for Children With Hemophilia and Their Families: A Clinical Project. J Pediatr Oncol Nurs 2016 Jul;33(4):282-288. [doi: 10.1177/1043454215607340] [Medline: 26510644]
79. Johnston DL, Murto K, Kurzawa J, Liddy C, Keely E, Lai L. Use of Electronic Consultation System to Improve Access to Care in Pediatric Hematology/Oncology. J Pediatr Hematol Oncol 2017 Oct;39(7):e367-e369. [doi: 10.1097/MPH.0000000000000833] [Medline: 28437292]
80. Pedrosa F, Shaikh F, Rivera G, Ribeiro R, Qaddoumi I. The Impact of Prospective Telemedicine Implementation in the Management of Childhood Acute Lymphoblastic Leukemia in Recife, Brazil. Telemed J E Health 2017 Oct;23(10):863-867 [FREE Full text] [doi: 10.1089/tmj.2016.0273] [Medline: 28422613]
81. Stulac S, Mark Munyaneza RB, Chai J, Bigirimana JB, Nyishime M, Tapela N, et al. Initiating Childhood Cancer Treatment in Rural Rwanda: A Partnership-Based Approach. Pediatr Blood Cancer 2016 May;63(5):813-817. [doi: 10.1002/pbc.25903] [Medline: 26785111]
82. Breen S, Kofoed S, Ritchie D, Dryden T, Maguire R, Kearney N, et al. Remote real-time monitoring for chemotherapy side-effects in patients with blood cancers. Collegian 2017 Dec;24(6):541-549. [doi: 10.1016/j.colegn.2016.10.009]
83. Kharfan-Dabaja MA, Aljurf M. Hematopoietic cell transplantation: Training challenges and potential opportunities through networking and integration of modern technologies to the practice setting. Hematol Oncol Stem Cell Ther 2017 Dec;10(4):184-188 [FREE Full text] [doi: 10.1016/j.hemonc.2017.05.012] [Medline: 28633040]
84. Kulkarni R. Use of telehealth in the delivery of comprehensive care for patients with haemophilia and other inherited bleeding disorders. Haemophilia 2018 Jan;24(1):33-42. [doi: 10.1111/hae.13364] [Medline: 29205699]
85. Klasko SK. Healthcare Transformation: The Future of Telemedicine. Telemed J E Health 2016 May;22(5):337-341. [doi: 10.1089/tmj.2016.29007.skk] [Medline: 27128778]
86. Fortney JC, Pyne JM, Turner EE, Farris KM, Normoyle TM, Avery MD, et al. Telepsychiatry integration of mental health services into rural primary care settings. Int Rev Psychiatry 2015;27(6):525-539. [doi: 10.3109/09540261.2015.1085838] [Medline: 26634618]
87. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 Pandemic, Lessons for the Future. Telemed J E Health 2020 May 08;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
88. Langarizadeh M, Tabatabaei MS, Tavakol K, Naghipour M, Rostami A, Moghbeli F. Telemental Health Care, an Effective Alternative to Conventional Mental Care: a Systematic Review. Acta Inform Med 2017 Dec;25(4):240-246 [FREE Full text] [doi: 10.5455/aim.2017.25.240-246] [Medline: 29284913]
89. Lanssens D, Vandenberk T, Thijs IM, Grieten L, Gyselaers W. Effectiveness of Telemonitoring in Obstetrics: Scoping Review. J Med Internet Res 2017 Sep 27;19(9):e327 [FREE Full text] [doi: 10.2196/jmir.7266] [Medline: 28954715]
90. Lilly CM, Zubrow MT, Kempner KM, Reynolds HN, Subramanian S, Eriksson EA, Society of Critical Care Medicine Tele-ICU Committee. Critical care telemedicine: evolution and state of the art. Crit Care Med 2014 Nov;42(11):2429-2436. [doi: 10.1097/CCM.0000000000000539] [Medline: 25080052]
91. Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. Int J Med Inform 2015 Sep;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]
92. Jansen F, van Uden-Kraan CF, van Zwieten V, Witte BI, Verdonck-de Leeuw IM. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. Support Care Cancer 2015 Jun;23(6):1679-1688. [doi: 10.1007/s00520-014-2514-7] [Medline: 25424520]
93. Kraft P, Hillmann S, Rücker V, Heuschmann PU. Telemedical strategies for the improvement of secondary prevention in patients with cerebrovascular events-A systematic review and meta-analysis. Int J Stroke 2017 Aug;12(6):597-605. [doi: 10.1177/1747493017706188] [Medline: 28440116]
94. Kulkarni R, Mauser-Bunschoten EP, Stedman C, Street A. Medical co-morbidities and practice. Haemophilia 2014 May;20 Suppl 4:130-136. [doi: 10.1111/hae.12403] [Medline: 24762288]
95. Baker RI, Laurenson L, Winter M, Pritchard AM. The impact of information technology on haemophilia care. Haemophilia 2004 Oct;10 Suppl 4:41-46. [doi: 10.1111/j.1365-2516.2004.00995.x] [Medline: 15479371]
96. Binder AF, Gabrilove J. A path forward: Global health, telemedicine, and hematology. Am J Hematol 2016 Aug;91(8):E333-E334 [FREE Full text] [doi: 10.1002/ajh.24418] [Medline: 27178529]

97. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
98. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical Judicial Affairs American Medical Association. Ethical practice in Telehealth and Telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
99. Langarizadeh M, Moghbeli F, Aliabadi A. Application of Ethics for Providing Telemedicine Services and Information Technology. Med Arch 2017 Oct;71(5):351-355 [FREE Full text] [doi: 10.5455/medarh.2017.71.351-355] [Medline: 29284905]
100. Nittari G, Khuman R, Baldoni S, Pallotta G, Battineni G, Sirignano A, et al. Telemedicine Practice: Review of the Current Ethical and Legal Challenges. Telemed J E Health 2020 Dec;26(12):1427-1437 [FREE Full text] [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
101. Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth 2017;8(4):266-276. [doi: 10.1080/23294515.2017.1403980] [Medline: 29125425]
102. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of Mobile Health in Communities Underrepresented in Biomedical Research: Barriers and Ethical Considerations for Scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
103. Torous J, Nebeker C. Navigating Ethics in the Digital Age: Introducing Connected and Open Research Ethics (CORE), a Tool for Researchers and Institutional Review Boards. J Med Internet Res 2017 Feb 08;19(2):e38 [FREE Full text] [doi: 10.2196/jmir.6793] [Medline: 28179216]
104. Dubben H, Beck-Bornholdt H. Systematic review of publication bias in studies on publication bias. BMJ 2005 Aug 20;331(7514):433-434 [FREE Full text] [doi: 10.1136/bmj.38478.497164.F7] [Medline: 15937056]

Abbreviations

GRADE: Grades of Recommendation, Assessment, Development, and Evaluation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.04.21; peer-reviewed by C Jonassaint, S Norell; comments to author 07.05.21; revised version received 21.05.21; accepted 31.05.21; published 08.07.21.

Please cite as:
Shah AC, O'Dwyer LC, Badawy SM
Telemedicine in Malignant and Nonmalignant Hematology: Systematic Review of Pediatric and Adult Studies
JMIR Mhealth Uhealth 2021;9(7):e29619
URL: https://mhealth.jmir.org/2021/7/e29619
doi:10.2196/29619
PMID:34255706

Original Paper

Assessing Apps for Health Care Workers Using the ISYScore-Pro Scale: Development and Validation Study

Inmaculada Grau-Corral[1,2], PhD; Percy Efrain Pantoja[1,3], MPH, MD; Francisco J Grajales III[1,4], MSc; Belchin Kostov[5], PhD; Valentín Aragunde[6], MD; Marta Puig-Soler[6], MD; Daria Roca[2,7], RN; Elvira Couto[2], MD; Antoni Sisó-Almirall[5,8], PhD

[1]Fundacion iSYS (internet, Salud y Sociedad), Barcelona, Spain

[2]Hospital Clínic Barcelona, Barcelona, Spain

[3]Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IRB Sant Pau), Barcelona, Spain

[4]University of British Columbia, Vancouver, BC, Canada

[5]Primary Healthcare Transversal Research Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[6]Casanova Primary Health-Care Center, Consorci d'Atenció Primària de Salut Barcelona Esquerra, Barcelona, Spain

[7]Universitat de Barcelona, Barcelona, Spain

[8]Les Corts Primary Health-Care Center, Consorci d'Atenció Primària de Salut Barcelona Esquerra, Barcelona, Spain

Corresponding Author:
Inmaculada Grau-Corral, PhD
Fundacion iSYS (internet, Salud y Sociedad)
c/ Mallorca 140 2-4
Barcelona, 08036
Spain
Phone: 34 692241233
Email: Imma.grau.corral@gmail.com

Abstract

Background: The presence of mobile phone and smart devices has allowed for the use of mobile apps to support patient care. However, there is a paucity in our knowledge regarding recommendations for mobile apps specific to health care professionals.

Objective: The aim of this study is to establish a validated instrument to assess mobile apps for health care providers and health systems. Our objective is to create and validate a tool that evaluates mobile health apps aimed at health care professionals based on a trust, utility, and interest scale.

Methods: A five-step methodology framework guided our approach. The first step consisted of building a scale to evaluate apps for health care professionals based on a literature review. This was followed with expert panel validation through a Delphi method of (rated) web-based questionnaires to empirically evaluate the inclusion and weight of the indicators identified through the literature review. Repeated iterations were followed until a consensus greater than 75% was reached. The scale was then tested using a pilot to assess reliability. Interrater agreement of the pilot was measured using a weighted Cohen kappa.

Results: Using a literature review, a first draft of the scale was developed. This was followed with two Delphi rounds between the local research group and an external panel of experts. After consensus was reached, the resulting ISYScore-Pro 17-item scale was tested. A total of 280 apps were originally identified for potential testing (140 iOS apps and 140 Android apps). These were categorized using International Statistical Classification of Diseases, Tenth Revision. Once duplicates were removed and they were downloaded to confirm their specificity to the target audience (ie, health care professionals), 66 remained. Of these, only 18 met the final criteria for inclusion in validating the ISYScore-Pro scale (interrator reliabilty 92.2%; kappa 0.840, 95% CI 0.834-0.847; $P<.001$).

Conclusions: We have developed a reproducible methodology to objectively evaluate mobile health apps targeted to health care professionals and providers, the ISYScore-Pro scale. Future research will be needed to adapt the scale to other languages and across other domains (eg, legal compliance or security).

(JMIR Mhealth Uhealth 2021;9(7):e17660) doi:10.2196/17660

KEYWORDS

assessment; mobile app; mobile application; mHealth; health care professionals; mobile application rating scale; scale development

Introduction

Information and communication technologies offer countless opportunities for knowledge management. Access, collection, and production of information are redefined by information technology with digitization [1]. In the health sector, where knowledge management is central to every process, the digitization of information has had an enormous impact. Moreover, health professionals are incorporating information technology into nearly every aspect of patient care and research.

As digitization of health care and health services expand, the landscape of digital health innovates, transforms, and scales for mass use and adoption. Dorsey and Topol [2] identified three new linked trends in how the health and medical community's concept of digital health evolves in this everchanging landscape. From the old paradigm in the concept of digital health—picturing increased accessibility, managing acute conditions, and facilitating communication between hospitals and health providers—towards a new scene where digital health is used for convenience services, management of patients with chronic or episodic disease, and for increased communication with patients, digital health enables the objective control of episodic and chronic conditions, and increased communication between providers and patients through their mobile devices. Mobile apps are a large driving force in improving (digital health) capacity for health systems.

Since the launch of mobile app platforms in 2008, people have had access to apps aimed at personal health management. As the popularity of these platforms has increased, their adoption has expanded. As of March 2021, the health and fitness apps represented 2.98% and those of Medicine, 1.88% [3]. Digital health apps have the potential to continue improving health and medical community concerns such as patient follow-up and monitoring, adherence to treatments, and promotion of healthy habits [4,5]. However, there is a need for a clearer understanding of best practices to evaluate the digital health apps.

Fieldwork has clarified the need for each professional and academic domain to understand and capture the needs of potential users. In the last few years, one of the gaps documented in the field is the need to develop and validate new mobile health (mHealth) assessment tools. New research needs to emerge from multidisciplinary and experienced teams to create convergence and integrate methodologies to improve consistency in the mHealth app market [6,7]. For example, theoretical frameworks like the Technology Acceptance Model (TAM) address some of the needs from behavioral research but fall short in terms of mHealth evaluation methods.

To expand on this, from the behavioral perspective, the TAM [8-11] describes the factors as to why users may uptake new technology. This model proposes that when users are faced with a new technology, perceived utility and ease of use will shape their decisions. From the lens of health care professionals, the TAM suggests that new technologies or apps should provide solutions for or to assist with the needs of the clinical practice. Enabling or enhancing patient care will thus become a top priority, ranging from diagnosis to socialization and during interparty communication.

A wide range of studies have evaluated the use of digital tools to enable or enhance patient care. These include, for example, the use of telemedicine to follow patients with chronic diseases (eg, diabetes [12] and lung cancer [13]), disease-relevant education (eg, lymphedema management [14]), and exercise and physical activity monitoring following a cardiac event for secondary prevention [15], along with the use of virtual communities to improve self-care and patient engagement, such as Forumclinic at Hospital Clínic de Barcelona [16].

Today, there is a great societal need to document how to create great experiences through digital health apps, to evaluate these interventions, and to solve the pain points of both patients and health care professionals. High-quality apps can catalyze the efficient translations of new research findings into daily clinical practice, strengthen knowledge translation from the lab to the bedside, and may even radically transform patient care. In this journey, establishing a validated scale to assess the trust, utility, and interest of mobile apps used by health care professionals is the first step to understand how these tools may impact the quality of care delivered and the outcomes of patients receiving care.

Despite multiple efforts to develop rating scales to evaluate apps used in health care [7,17,18], as of the writing of this paper, not a single one (to our knowledge) has empirically addressed how these tools may support day-to-day clinical practice.

The Internet, Health and Society Foundation (Fundación Internet, Salud y Sociedad [ISYS] in Spanish) works with journal editors and members of professional associations, and collaborates with a diverse group of experts to distil best practices in the digital health domain. In 2014, the ISYS developed a scale to evaluate the quality of mHealth apps for patients (the ISYScore for patients or ISYScore-Pac) [18]. This study builds on this earlier research and expands upon its applicability for health care professionals.

The goal of this study is to develop a tool that evaluates health care apps targeted to health care professionals and medical workers at the bedside. The scale was designed specifically to assess interest, trust, and usefulness, and allow for empirical replicability across these dimensions. The scoring system that supported the development of this scale is also presented.

Methods

The methodology proposed by Moher [19] in his work on health research reporting guidelines was used for the assessment portion. Limitations from other rating scales were considered [7]. This study involved an iterative sequence of a five-step methodology for the assessment of the scale: (1) investigation group definition, (2) theoretical framework and literature review, (3) scale draft, (4) expert panel definition and Delphi rating, and (5) pilot and first scale.

Investigation Group Definition

The local investigation group was defined as a local *Fundación ISYS* research team with experience in telemedicine and assessing other scales [20], and with strong knowledge on different domains (1 engineer, 5 medical specialists, 1 nurse, and 1 biostatistician). Profiles and expertise from the local investigation group members are available in Multimedia Appendix 1.

Theoretical Framework and Literature Review

A comprehensive literature search was conducted by the local investigation group to trace possible backgrounds, to collect a selection of theories that had been used in digital health interventions, and to perform a scope of other scales, focusing on the ones targeted specifically to health professionals. Papers were retrieved from the PubMed database with all articles that include "assessment" AND "healthcare professional" AND "mobile applications" with no language or time limitations.

The local investigation group used the TAM framework [8-11] and the vision of persuasive technology conceptualized by Fogg's functional triad: information and communication technology that function as tools, media, and social actors [21,22].

Draft of the Scale

The development, implementation, usability, viability, and acceptance information of existing models were extracted. The local investigation group classified assessment criteria into categories and subcategories, and developed the scale items and descriptors. Considerations for the final scale were the dimension values: trust, utility, and interest (Textbox 1). Iterative corrections were made until consensus was reached.

Textbox 1. The trust, utility, and interest model presented to the expert panel (ISYScore-Pro).

Trust

- A1. Validated by a health agency, scientific society, health care professional college, or nongovernmental organization
- A2. Authors are explicitly identified
- A3. Website is accessible (responsibility)
- A4. Cites peer-reviewed sources
- A5. Names the organization responsible
- A6. Updated within the last calendar year
- A7. Disclosure on how the app was financed

Utility

- Technology as a tool (increases capacity)
  - B1. Provides calculations and measurements
  - B2. Helps in a care procedure
  - B3. Archives data images
- Technology as a medium (increases the experience)
  - B4. Facilitates observation of cause-effect relationships and allows users to rehearse
  - B5. Facilitates observation of those who do well (vicarious learning)
  - B6. Facilitates patient follow-up
- Technology as social actor (increases social relationships)
  - B7. Obtains positive feedback
  - B8. Provides social content

Interest

- C1. Positive user ratings/downloads
- C2. Available on two platforms (18 items were selected from the review of the literature; this item was removed after local investigation group and external panel of experts agreement)
- C3. Content available in other formats (eg, web, tablet, or magazines)

The main purpose of this phase's outcome was to define objective indicators that cover all three dimensions of the assessment scale. *Trust* constitutes a dimension crucial for health apps. It would evaluate the robustness of the relationship that a mobile app has with scientific evidence, the frequency in which its information is updated, and the declaration of possible conflicts of interest. For *utility*, the local investigation group chose indicators inspired by the Fogg triad [21,22]. Within this

dimension, technology is evaluated as the capacity to enhance or improve experience or as a social actor. For health professionals, perceived ease of use falls within the user *interest* dimension of the ISYScore-Pro, such as evaluating the availability of the mobile app across different platforms, which is valued by some in the scientific community.

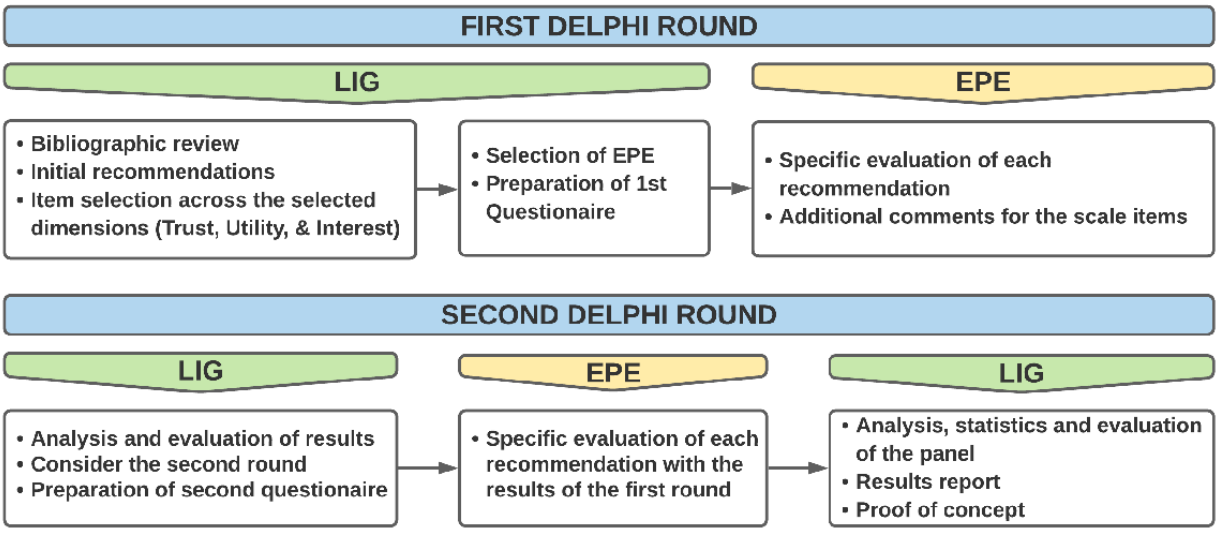
The External Panel of Experts and Delphi Rating

The external panel of experts was selected considering all backgrounds that can help us to validate the 18-item scale draft (Multimedia Appendix 2). An attempt was made to look for heterogeneity. Doctors of different specialties (cardiology, pneumology, pediatrics, surgery, public health, etc); nurses; psychologists and psychiatrists; and experts in physical education, pharmacy, and technology were recruited.

For the expertise profiles in panel selection, the approach [20] was made looking at Twitter, Facebook, and the local investigation group professional network. A wide range of health care professionals, all of them well-known influencers, key opinion leaders, and users of mobile technology in Spain and recognized as scientists and scientific or evidence-based disseminators on eHealth, were contacted. Due to distance and budget limitations, they were only contacted through email, social media, or other publicly available data. In total, 35 Spanish eHealth experts were invited to be part of the panel.

For feedback on the 18-item scale draft, a Delphi process [18,23,24] was followed. A minimum participation of 70% was needed [6] for each Delphi round. Participant agreement was also set at 75%, as advised by the literature [23-26]. Two rounds of questionnaires were sent out to the external panel of experts. Responses were aggregated and shared with the local investigation group after each round (Figure 1).

Figure 1. Delphi flowchart. EPE: external panel of experts; LIG: local investigation group.



The external panel of experts had to assess whether they would include each category in the draft scale (Textbox 1) and, if included, assign a weight to it on a scale from 1 to 5 (0 was used to exclude the category altogether). They were also asked to consider adding a new indicator not proposed in the first round only.

Pilot and Scale Testing

To test the scale's performance, we gathered a sample of apps to evaluate. For sampling, the local investigation group used the automated method for capturing apps with Google advanced search tools. For summary purposes, this method explores different results by disease clusters. For clustering, the local investigation group selected keywords from the disease groups in the *International Statistical Classification of Diseases, Tenth Revision* (*ICD-10*) [27]. Pregnancy-related apps and non-disease–related codes where discarded, which brought the total of disease groups explored to 14.

For each *ICD-10* cluster, the first 10 search results on iTunes and the first 10 in Google Play were kept. A total of 280 apps were collected: 140 from Google Play and 140 from iTunes. The strategy in Spanish for each cluster search can be found in Multimedia Appendix 3.

For this pilot, the inclusion criteria included that the app was in the local language (Spanish), the target audience was health care professionals, and its general availability did not require passwords or specific geographies. As exclusion criteria, accuracy was considered, excluding, for example, apps that mention cancer as horoscope. eBooks and podcasts were also excluded. The local investigation group also established that the most recent update had to be within the previous calendar year irrespective of the number of downloads at any time. After duplicates were removed and the inclusion and exclusion criteria were applied, 66 apps remained, and the scale was applied. By consensus the local investigation group established that apps had to meet a minimum cut-off score of no less than 4 items of the final ISYScore-Pro scale.

To evaluate the reliability of the scale the ISYScore-Pro was applied to the final sample of 66 apps. The local investigation group reviewers analyzed apps in pairs. In case of discrepancies, each subgroup discussed the findings and reached a consensus. If consensus was not reached, a third researcher solved the discrepancies [28,29].

Interrater reliability was measured on nominal variables using Cohen kappa with STATA v15.0 (StataCorp) [29]. The target aim was to reach >80% agreement. As the final score includes different categories, a weighted kappa calculation was used. The local investigation group considered the agreement *very low* for kappa results lower than 0.20, *low* for 0.20 to 0.39, *moderate* for 0.40 to 0.59, *high* for 0.60 to 0.79, and *very high or excellent* for 0.80 or higher.

Results

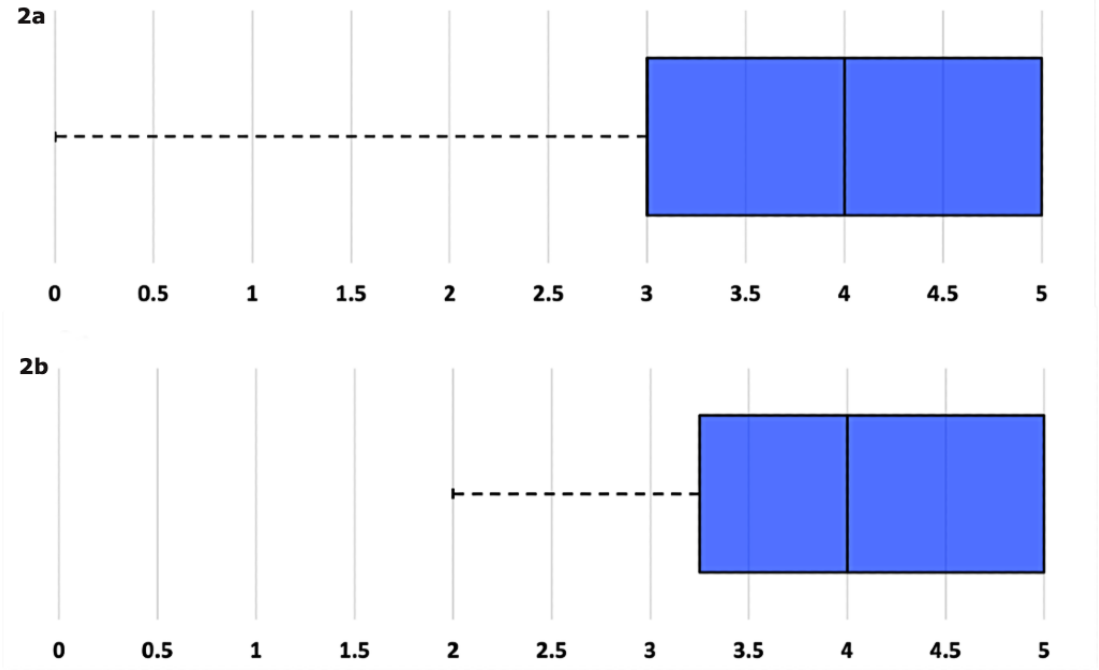
Scale Draft

The literature review indicated that existing models do not sufficiently overlap or integrate to provide a framework for rating mobile apps specific to health care professionals. The main mHealth domains were at the individual, organizational, and contextual levels. Existing scales included usability (perceived) and ease of use, design and technology aspects, cost, aesthetics, time, privacy and security, familiarity with technology, risk-benefit assessment, and interaction with others (colleagues, patients, and management) [7,8,30,31]. Based on these findings and the prior work on the ISYScore-Pa scale, our rating scale methodology tested three dimensions (Textbox 1): trust, utility, and interest [8-10,21,32-34].

Delphi Rating of the Draft Scale and Final Scale

The external panel of experts members completed the study's two rounds of measure ratings. From the 35 contacted people, 28 (80%) participated in the first round, and 25 (71.4%) in the second one. After the first interaction, results (ie, Figure 2) were sent for a second interaction to the external panel of experts members. From the 18 items originally proposed (Textbox 1), 17 (94.4%) were included in the final draft (C3 was discarded). The Delphi interactions database is available upon request from the corresponding author.

Figure 2. Examples of graphics sent to the external panel of experts (EPE) for the second interaction. After the first interaction, boxplots were sent to the EPE members for a second interaction. (2a) This boxplot example reflects an item with low consensual rate, catalogued by one with a zero value; later on, this item was not included. (2b) This boxplot is an example of an included item, well valuated in the first interaction.



Scale Reliability and Interrater Agreement

A total of 66 apps were used to test the reliability of the scale. Of these, 13 were excluded, as their score was equal to or lower than 12 (out of 17), which was deemed as the minimum cut-off. The breakdown of the apps evaluated is presented in Table 1. Apps were stratified according to their *ICD-10* disease cluster. Interrater reliability scores were also calculated.

Table 1. Evaluation of the agreement between raters, weighted by *ICD-10*[a] group cluster and app.

ICD-10 cluster	Apps included for score evaluation		
	Google Play, n	iTunes, n	Total[b], n
I. Certain infectious and parasitic diseases	2	2	3
II. Neoplasms	3	4	5
III. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	7	2	9
IV. Endocrine, nutritional, and metabolic diseases	2	2	2
V. Mental and behavioral disorders	2	2	2
VI. Diseases of the nervous system	1	0	1
VII. Diseases of the eye and adnexa	5	2	7
VIII. Diseases of the ear and mastoid process	0	0	0
IX. Diseases of the circulatory system	4	6	6
X. Diseases of the respiratory system	7	6	8
XI. Diseases of the digestive system	4	1	4
XII. Diseases of the skin and subcutaneous tissue	5	2	5
XIII. Diseases of the musculoskeletal system and connective tissue	8	10	11
XIV. Diseases of the genitourinary system	3	1	3
Total, n (%)	53 (80)	40 (61)	66 (100)

[a]*ICD-10*: *International Statistical Classification of Diseases, Tenth Revision*.

[b]Removing apps offered in both platforms but evaluated individually.

The flow of apps through the research process is shown on Figure 3 using a modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram.

Results of the evaluation of the agreement between raters, weighted by *ICD-10* group cluster and app, are shown in Table 2. Other analyses are shown in Multimedia Appendix 4. A 92.2% crude general interrater agreement was found, 93.7% and 91.0% when adjusted by cluster and app, respectively. Cohen kappa showed a significantly general agreement (0.84, 95% CI 0.834-0.847), without differences when weighted by app cluster or app evaluated.

Figure 3. Flow diagram of the app evaluation.

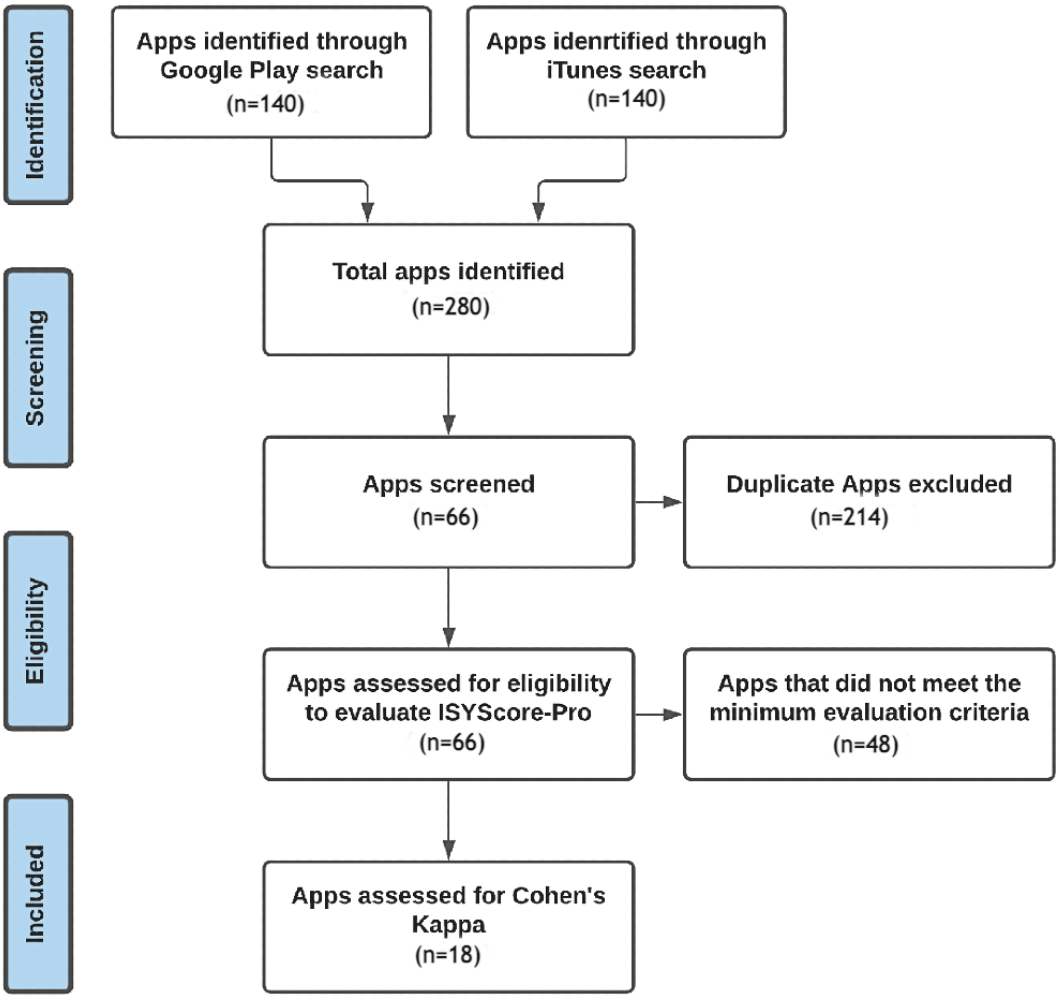


Table 2. Evaluation of the agreement between raters, weighted by *ICD-10*[a] group cluster and app.

Evaluation	Agreement (%)	Expected agreement (%)	Cohen kappa (95% CI)	P value
General (crude)	92.2	50.8	.84 (0.834-0.847)	<.001
Weighted by *ICD-10* cluster	93.7	53.1	.87 (0.865-0.867)	<.001
Weighted by app	91.0	52.5	.81 (0.809-0.810)	<.001

[a]*ICD-10*: *International Statistical Classification of Diseases, Tenth Revision*.

Discussion

ISYScore-Pro Tool Development

The 17-item scale ISYScore-Pro is specific to mobile apps targeted to health professionals. ISYScore-Pro assesses apps in the three dimensions that were identified by a literature review [8-10,21]. Other scales in existence consider other dimensions and are often centered upon perception and usability [32-34]. The prioritization of the dimensions on the ISYScore-Pro scale are specific to mobile apps targeted to health care professionals and their practice.

During the Delphi, all dimensions and items on the scale reached strong external panel of experts consensus within two rounds with the exception of C2 (Textbox 1), which was removed from the scale. No further rounds were considered necessary. During the pilot period, an emphasis was placed in selecting health care professionals with different backgrounds across each of the local investigation group team pairs. The piloting also revealed that the scale was easy to apply. The few discrepancies that were encountered in the process were discussed, and consensus was reached in all cases. These results are confirmed by the reported interrater agreement.

Our findings are congruent with previously published research [17,25]. Of particular relevance are the findings of Gagnon and colleagues [25] that found that the main adoption factors of mobile apps by health professionals are perceived usefulness and ease of use, design and technical concerns, cost, time, privacy and security issues, familiarity with the technology, risk-benefit assessment, and interaction with others. The ISYScore-Pro scale addresses a number of these factors including the risk-benefit assessment in the trust domain, perceived utility, and interactions with others (colleagues,

patients, and management). The Spanish apps in our sample had low perception of interest, an acceptable level of trust, and an improved perception of utility.

Validating and improving the development of scales specific to mobile apps targeted to health care workers is essential in the future. Ultimately, this will improve the quality, reliability, and usability of these apps, and may improve equity of information whenever and wherever care is delivered.

Limitations

It would have been difficult to develop the scale using only a systematic review due to the lack of peer-reviewed papers and our current state of knowledge on evaluation tools for smartphone apps targeted to health care professionals. Nevertheless, this issue was overcome with the use of a 5-step framework and a Delphi process.

A limitation of our method was using only volunteers in the external panel members and the investigator group. To mitigate this, a diverse set of professional practice and research experience was used to select the individuals that participated in the Delphi process. Additionally, no conflict of interest was reported by any of the researchers and experts in the study. Another limitation of the ISYScore-Pro scale is that the security, privacy, legal compliance, and efficiency [35] are not assessed by the scale.

It is important to note that it is difficult to find useful apps for health professionals in the Spanish language. As of the time of the study, the market is small, and penetration remains low when compared to the English language. The sample apps that were evaluated had few downloads and even fewer ratings (ie, usually between 500 and 1000).

Future Research

Although the ISYScore-Pro scale is specific to the Spanish language, future research should compare and contrast our findings with expert opinions in other languages and, in particular, clinicians who practice in the English language. Although our methodology is robust, it is also resource intensive, and the use of automated artificial intelligence and machine learning methods may facilitate and substantially reduce the level of resources needed to assess apps on the market.

The domain of data security in mobile apps was not evaluated. Some authors [7,36] have suggested using open-source developer codes to reduce the potential of malicious functionalities. Future research should evaluate the security, privacy, and integrity of apps and the quality of information contained within them.

Conclusion

Our research is the first empirical attempt at developing a scale that assesses mobile apps in Spanish targeted to health care professionals. The ISYScore-Pro scale uses a reliable and replicable methodology that standardizes the assessment of trust, utility, and interest using 17 criteria grounded on the existing peer-reviewed literature and the inputs of an expert panel of health care professionals.

Acknowledgments

We would like to acknowledge each member of the external panel of experts for their assistance with the development of the ISYScore-Pro: Manuel Escobar, Luis Fernández Luque, Victor Bautista, Jose Juan Gomez, Amalia Arce, Rosa Taberner, Joan Fontdevila, Joan Carles March, Frederic Llordachs, Joan Escarrabill, Mireia Sans, Antoni Benabarre, Jordi Vilardell, Anna Sort, Jose María Cepeda, Marga Jansa, Rosa Pérez, Marc Fortes, Lluís Gonzalez, Imma Male, Jordi Vilaro, Luna Couto Jaime, Javi Telez, Mónica Moro, Pau Gascón, Marisa Ara, Manuel Armayones, and Eulalia Hernández.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Local investigators group (LIG) profiles.
[DOCX File , 13 KB - mhealth_v9i7e17660_app1.docx]

Multimedia Appendix 2
External panel of experts (EPE) profiles.
[DOCX File , 14 KB - mhealth_v9i7e17660_app2.docx]

Multimedia Appendix 3
Strategy in Spanish for each International Statistical Classification of Diseases and Related Health Problems (Tenth Revision; ICD-10) cluster search on Google Play and iTunes.
[DOCX File , 16 KB - mhealth_v9i7e17660_app3.docx]

Multimedia Appendix 4
Evaluation of the agreement between raters, by groups of evaluation, by application, and by item.

[DOCX File , 18 KB - mhealth_v9i7e17660_app4.docx]

References

1. Grau I. Digital health information. Science Direct 2017 Feb;49(2):67-68 [FREE Full text] [doi: 10.1016/j.aprim.2016.10.004] [Medline: 28131323]
2. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/nejmra1601705]
3. Statista. Most popular app store categories. Statista. 2021 Mar. URL: http://www.statista.com/statistics/166976/popular-categories-in-the-app-store/ [accessed 2021-04-20]
4. Digital Doctor 2017 Internet. Ipsos Healthcare. 2017. URL: https://www.ipsos.com/sites/default/files/ct/publication/documents/2018-07/digital_doctor_2017_webinar_final_6sep17.pdf [accessed 2019-05-20]
5. Ipsos. Digital Doctor. Behavioural Science Internet. URL: https://www.ipsos.com/en/10-behavioural-science-principles-great-digital-creative [accessed 2019-05-20]
6. Ben-Zeev D, Schueller S, Begale M, Duffecy J, Kane J, Mohr D. 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]
7. 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]
8. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
9. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
10. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
11. Venkatesh V, Morris M, Davis G, Davis F. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
12. Bonoto B, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of Mobile Apps to Support the Care of Patients With Diabetes Mellitus: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
13. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Overall survival in patients with lung cancer using a web-application-guided follow-up compared to standard modalities: Results of phase III randomized trial. JCO 2016 May 20;34(15_suppl):LBA9006-LBA9006. [doi: 10.1200/jco.2016.34.15_suppl.lba9006]
14. Fu M, Axelrod D, Guth A, Rampertaap K, El-Shammaa N, Hiotis K, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. Mhealth 2016 Jul;2 [FREE Full text] [doi: 10.21037/mhealth.2016.07.03] [Medline: 27493951]
15. 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. JAHA 2015 Oct 29;4(11). [doi: 10.1161/jaha.115.002239]
16. Grau Corral I. La Comunicación en Comunidades Virtuales de Pacientes en un gran Hospital Universitario. In: Internet Interdisciplinary Institut (IN3). Barcelona: Universitat Oberta de Catalunya; Dec 11, 2011:327.
17. McKay FH, Slykerman S, Dunn M. The App Behavior Change Scale: Creation of a Scale to Assess the Potential of Apps to Promote Behavior Change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
18. Grau I, Kostov B, Gallego J, Grajales Iii F, Fernández-Luque L, Sisó-Almirall A. [Assessment method for mobile health applications in Spanish: The iSYScore index]. Semergen 2016;42(8):575-583. [doi: 10.1016/j.semerg.2015.12.001] [Medline: 26879598]
19. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010 Feb 16;7(2):e1000217 [FREE Full text] [doi: 10.1371/journal.pmed.1000217] [Medline: 20169112]
20. Nohara Y, Kai E, Ghosh PP, Islam R, Ahmed A, Kuroda M, et al. Health checkup and telemedical intervention program for preventive medicine in developing countries: verification study. J Med Internet Res 2015 Jan 28;17(1):e2 [FREE Full text] [doi: 10.2196/jmir.3705] [Medline: 25630348]
21. Fogg B. Persuasive Technology Using Computers to Change What We Think and Do. In: Morgan Kaufmann. Amsterdam: Morgan Kaufmann; 2003:312.
22. 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]
23. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]

24. Diamond I, Grant R, Feldman B, Pencharz P, Ling S, Moore A, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
25. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
26. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995 Aug 05;311(7001):376-380 [FREE Full text] [doi: 10.1136/bmj.311.7001.376] [Medline: 7640549]
27. Ministerio de Sanidad, Consumo y Bienestar Social. eCIE10ES Edición electrónica de la CIE-10-ES Diagnósticos. : Gobierno de Espana; 2018. URL: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_10_mc.html [accessed 2019-05-20]
28. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
29. Hallgren K. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
30. Christie D, Thompson R, Sawtell M, Allen E, Cairns J, Smith F, et al. Structured, intensive education maximising engagement, motivation and long-term change for children and young people with diabetes: a cluster randomised controlled trial with integral process and economic evaluation - the CASCADE study. Health Technol Assess 2014 Mar;18(20):1-202 [FREE Full text] [doi: 10.3310/hta18200] [Medline: 24690402]
31. Ellsworth RE, Valente AL, Shriver CD, Bittman B, Ellsworth DL. Impact of lifestyle factors on prognosis among breast cancer survivors in the USA. Expert Rev Pharmacoecon Outcomes Res 2012 Aug;12(4):451-464. [doi: 10.1586/erp.12.37] [Medline: 22971032]
32. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
33. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
34. Faruque MO, Holakouie Naieni K, Ardalan A, Ahmadnezhad E, Mohammadinia L. Feasibility assessment of using geoinformatics technology in disaster disease surveillance in a developing country, iran. PLoS Curr 2015 Apr 14;7 [FREE Full text] [doi: 10.1371/currents.dis.cbb7c1d062d4ec0646c2d67319a312f2] [Medline: 25932348]
35. Changes to how we assess apps and tools. NHS Apps Library.: NHS Digital URL: https://tinyurl.com/rzu3rnjp [accessed 2019-08-26]
36. Su W. A Preliminary Survey of Knowledge Discovery on Smartphone Applications (apps): Principles, Techniques and Research Directions for E - health. In: arXivLabs. 2014 Jul 28 Presented at: ICME Int Conf Complex Med Eng; 28 Jul 2014; Dortmund, Germany p. 369. [doi: 10.13140/RG.2.1.1941.6565]

Abbreviations

ICD-10*: *International Statistical Classification of Diseases, Tenth Revision
ISYS: Internet, Salud y Sociedad
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

Edited by CL Parra-Calderón; submitted 01.01.20; peer-reviewed by M Kimura, A Baumel, J Kim; comments to author 14.06.20; revised version received 22.10.20; accepted 15.01.21; published 21.07.21.

Please cite as:
Grau-Corral I, Pantoja PE, Grajales III FJ, Kostov B, Aragunde V, Puig-Soler M, Roca D, Couto E, Sisó-Almirall A
Assessing Apps for Health Care Workers Using the ISYScore-Pro Scale: Development and Validation Study
JMIR Mhealth Uhealth 2021;9(7):e17660
URL: https://mhealth.jmir.org/2021/7/e17660
doi:10.2196/17660
PMID:34287216

Original Paper

Acceptability of Intervention Design Factors in mHealth Intervention Research: Experimental Factorial Study

Frank T Materia[1], MHS, PhD; Joshua M Smyth[2], PhD

[1]Division of Health Services and Outcomes Research, Children's Mercy Kansas City, Kansas City, MO, United States

[2]Departments of Biobehavioral Health and Medicine, The Pennsylvania State University, University Park, PA, United States

Corresponding Author:
Frank T Materia, MHS, PhD
Division of Health Services and Outcomes Research
Children's Mercy Kansas City
2401 Gillham Road
Kansas City, MO, 64108
United States
Phone: 1 (816) 731 7123
Email: ftmateria@cmh.edu

Abstract

Background: With the growing interest in mobile health (mHealth), behavioral medicine researchers are increasingly conducting intervention studies that use mobile technology (eg, to support healthy behavior change). Such studies' scientific premises are often sound, yet there is a dearth of implementational data on which to base mHealth research methodologies. Notably, mHealth approaches must be designed to be acceptable to research participants to support meaningful engagement, but little empirical data about design factors influencing acceptability in such studies exist.

Objective: This study aims to evaluate the impact of two common design factors in mHealth intervention research—requiring multiple devices (eg, a study smartphone and wrist sensor) relative to requiring a single device and providing individually tailored feedback as opposed to generic content—on reported participant acceptability.

Methods: A diverse US adult convenience sample (female: 104/255, 40.8%; White: 208/255, 81.6%; aged 18-74 years) was recruited to complete a web-based experiment. A 2×2 factorial design (number of devices×nature of feedback) was used. A learning module explaining the necessary concepts (eg, behavior change interventions, acceptability, and tailored content) was presented, followed by four vignettes (representing each factorial cell) that were presented to participants in a random order. The vignettes each described a hypothetical mHealth intervention study featuring different combinations of the two design factors (requiring a single device vs multiple devices and providing tailored vs generic content). Participants rated acceptability dimensions (interest, benefit, enjoyment, utility, confidence, difficulty, and overall likelihood of participating) for each study presented.

Results: Reported interest, benefit, enjoyment, confidence in completing study requirements, and perceived utility were each significantly higher for studies featuring tailored (vs generic) content, and the overall estimate of the likelihood of participation was significantly higher. Ratings of interest, benefit, and perceived utility were significantly higher for studies requiring multiple devices (vs a single device); however, multiple device studies also had significantly lower ratings of confidence in completing study requirements, and participation was seen as more difficult and was associated with a lower estimated likelihood of participation. The two factors did not exhibit any evidence of statistical interactions in any of the outcomes tested.

Conclusions: The results suggest that potential research participants are sensitive to mHealth design factors. These mHealth intervention design factors may be important for initial perceptions of acceptability (in research or clinical settings). This, in turn, may be associated with participant (eg, self) selection processes, differential compliance with study or treatment processes, or retention over time.

(JMIR Mhealth Uhealth 2021;9(7):e23303) doi:10.2196/23303

KEYWORDS

mHealth; acceptability; implementation; health behavior; smartphone; mobile phone; wearable

Introduction

Background

Most health behavior change programs have historically required participants to attend in-person appointments or sessions, with a trained clinician or facilitator guiding intervention delivery. However, in-person delivery modes can be costly, time consuming, and burdensome for both participants and providers. The field of mobile health (mHealth) broadly examines how portable, wireless technologies (eg, smartphones, wearable fitness trackers, and smartwatches) can be used in an effort to reduce costs, enhance access, and increase the reach of health behavior change interventions. The ubiquity of mobile technology ownership across populations and contexts makes mHealth uniquely positioned to supplement or replace behavior change programs traditionally delivered in person [1].

mHealth Design and Implementation Science

The field of mHealth is still relatively new as a scientific discipline (ie, approximately 2 decades old) [2]. Recent literature has demonstrated the efficacy of some mHealth approaches for improving behavioral health outcomes [3-5]. A key factor in mHealth implementation underlying potential efficacy is the willingness and capability of users (eg, patients and research participants) to successfully engage with the mHealth delivery system. There is still very little research formally examining evidence-based methods for designing mHealth to support uptake and adherence with end users (eg, participants enrolled in a technology-based behavior change program) or, conversely, what study design features may be inhibiting engagement (particularly *early on* such as at initial study enrollment). This issue is particularly vital in mHealth research, given the large number of design factors and considerations in play. Some common design decisions mHealth intervention researchers must consider, among many, including selecting which technologies (eg, smartphone vs tablet) will be used to deliver intervention content, determining how to deliver intervention messaging (eg, text vs videos), deciding on the style of messaging (eg, personalized vs generic content), and requiring persistent internet connectivity (ie, if internet connectivity is necessary to push content and share data or if content and data stay native to the device).

Rigorous research evaluating how different design decisions may enhance or hinder end user engagement with mHealth is limited. One possible solution for filling these gaps is integrating research approaches from implementation science. The goal of implementation science is to evaluate methods for integrating evidence-based supports (eg, mHealth) into practice, with the goal of enhancing the successful delivery and effectiveness of new (or underutilized) behavior change approaches [6]. One key aim of implementation science involves evaluating different intervention design features (eg, varying choices for delivery modes and content selection) and understanding how they affect successful program implementation (eg, support uptake and adherence to the intervention among participants [7]).

Acceptability in mHealth Design and Implementation

One implementation factor directly related to the intervention design is participant *acceptability*. One broad definition of acceptability is "the quality of being tolerated or allowed" [8], and the other is "the quality or state of meeting one's needs adequately" [9]. In terms of mHealth, one proposed definition of acceptability is "an end-user's subjective perceptions of, and measurable sustained engagement with, a mobile health intervention system...(including) perceived satisfaction, willingness, and agreeability" [10].

The premise of studying acceptability as part of mHealth implementation research is rooted in the theory of behavioral science. For example, the theory of planned behavior (TPB) explains how individuals' beliefs are associated with their actions [11]. The TPB proposes that individuals' motivations to adopt certain behaviors are driven by behavioral (eg, advantages to behavior change), normative (eg, external expectations), and control (eg, barriers to behavior change) beliefs. Tenets of the TPB can be applied to inform the study of acceptability in mHealth, especially for informing which dimensions of acceptability are important to evaluate. As the TPB proposes, there are multiple psychological factors that influence behavioral adoption. Similarly, acceptability is a multidimensional construct that includes perceptions of interest (piques participant curiosity), enjoyment (participation is pleasurable), difficulty (how tough participation will be), utility (individual value in participating), benefit (participation is advantageous), and confidence (perceived ability to accomplish participation) [10,12,13]. Researchers evaluating these dimensions can determine which mHealth intervention design elements are preferred by participants and which may limit motivation for participation. Such research allows for a better understanding of participant selection processes, differential compliance, or adherence over time in mHealth intervention research, which are factors related to effective and meaningful engagement with behavior change interventions [13-15].

Acceptability can also be assessed at different time points in the user experience—a priori and posteriori. Evaluating a priori acceptability (eg, *pre* or *prospective* acceptability) involves assessing end user acceptability for the design of a system before actually engaging with it. An a priori assessment of acceptability is particularly useful during the intervention design process, as it provides researchers with proxy ratings for how design decisions may influence individuals' initial beliefs, perceptions, motivation, and likelihood of uptake and adherence to program requirements in daily life (eg, participation in a research study). To assess a priori acceptability, researchers typically introduce a description or prototype of the intervention to the participant and evaluate their perceived acceptability of the intervention's various requirements. This is most commonly measured via self-report surveys and measures or through semistructured focus group interviews [12-14].

Posteriori acceptability (ie, *post* or *retrospective* acceptability) is evaluated after an end user has been introduced to a system and they have spent time engaging with it [14,15]. As opposed to a priori assessment (which is hypothetical and only predictive of intervention uptake), a posteriori assessment can inform

researchers of how acceptability may change over the course of intervention participation. Posteriori acceptability can also serve as a measure of ongoing participant engagement and buy-in with the program [14,15].

Limitations in Evaluating Acceptability in mHealth Design

Given the recent proliferation of mHealth technologies worldwide (eg, smartphones and wearables [1,16]), researchers from multiple disciplines (eg, behavioral medicine and computer science) have expressed an imperative need to evaluate acceptability as part of the mHealth intervention design process [17-22]. Acceptability information informs design decisions that may enhance participant motivation to meaningfully engage with the technology and may positively impact their likelihood of adopting effective behavioral change [2,23]. However, despite the recognition of a need, research specifically focused on evaluating how different design decisions influence mHealth intervention acceptability is limited. The studies that have been conducted have mostly used small, homogenous samples with limited generalizability [24]. Most design studies have also been descriptive (eg, using focus groups or brief surveys to evaluate acceptability) and lacked experimental designs. Experimental methods (eg, randomization and manipulation of technology design features) can be particularly informative for studying a priori acceptability because they afford controls that allow researchers to assess whether significant differences in acceptability ratings before participating in a mobile intervention are because of varying conditions of mHealth design factors [25,26]. Surprisingly, many mHealth design studies to date have also lacked attention to theory from implementation and behavioral sciences to inform evaluation plans [27]. The literature is unclear regarding how various mHealth design conditions differentially influence participant acceptance of the intervention and its participation requirements [10,28].

Common mHealth Design Features

To experimentally evaluate how mHealth design components influence acceptability (specifically, a priori acceptability), an approach was taken to identify a small number of representative design features to study (from the large array of possible design options), which allows for the examination of several intervention features that are already extensively used in mHealth and also serves as an example (*proof of concept*) for this method more generally (that may be applicable to other design features as well). Two mHealth intervention methods of interest are requiring a single mobile device rather than multiple devices (eg, for broader data collection or alternate modes of content delivery) and providing tailored intervention content to individual participants (rather than providing the same generic content to all participants).

An advantage of requiring participants to use multiple devices (eg, smartphones, smartwatches, and ambulatory heart rate monitors) in a study is that researchers can obtain diverse data streams and use a variety of interfaces to deliver content [29-31]. A commonsense worry about this approach, partially supported by previous posteriori research, is that end users report feeling overwhelmed and exhibit lower acceptability after participating in a program that requires the concurrent use of multiple devices [32]. However, research examining a priori acceptability for participating in an mHealth program that requires multiple devices is less clear. If requiring multiple devices reduces a priori acceptability, participants may be less likely to participate, but research examining this is scant.

Although much intervention content is standardized and delivered generically to all participants, there is growing interest in providing intervention elements tailored for each participant. In brief, this involves providing content that is specific to each individual end user (eg, prompting participants by name and generating messaging based on person-specific health indicators). Previous work suggests that tailoring may have beneficial effects on intervention outcomes [1,33,34]. In addition, previous work measuring posteriori acceptability has demonstrated that participants positively perceive tailored content as enhancing their satisfaction with behavioral interventions; however, little empirical work has focused on investigating how this provision may influence a priori acceptability [33,35].

There is no existing research that has specifically examined the dimensions of a priori acceptability for participating in an mHealth intervention that requires multiple device utilization and provides tailored intervention content. Previous posteriori studies have examined these two factors separately after intervention engagement and found that they are meaningful for acceptability after participation is complete, but there is little research focused on understanding how potential mHealth participants may respond to such design components in the early stages of recruitment, as the a priori acceptability of device requirements and tailoring may influence individuals' motivation for selecting to participate in mHealth research. Furthermore, given the general lack of experimental research in this space, the aspects of such designs that are reliably related to positive perceptions of a priori acceptability are largely unknown (past observational inference). These are important factors for researchers to understand and consider in designing mHealth interventions to support participant motivation, uptake, engagement, and adherence to program requirements. More generally, this method serves as proof of concept for how other design features can be tested—singly or in combinations of factors—and using assessments of a priori acceptability, posteriori acceptability, or both.

This Study

This study sought to understand the impact that two common design decisions in mHealth behavior change interventions—requiring a single versus multiple devices (a study smartphone and a wrist sensor) and providing tailored as opposed to generic content—have on a priori participant acceptability. Specifically, the goal was to understand if participants were sensitive to changes in factors of these mHealth design decisions and how these changes may affect various dimensions of a priori acceptability, which act as proxy indicators of participant motivation and likelihood for meaningful engagement with mHealth intervention content and requirements. Previous research and theory from behavioral science (ie, TPB) informed the design of the study instrument and the dimensions of acceptability measured (ie, perceived

interest, benefit, utility, enjoyment, confidence and difficulty in participating, and perceived likelihood of participating [12,13,36]). This was an experimental 2×2 factorial design study using cross-sectional web-based survey methods. One factor was the number of devices required (1 vs 2), and the second factor was the nature of feedback (personalized vs generic); thus, each cell represents a different mHealth study scenario that featured different combinations of the two mHealth design factors of the study. As each participant received all four scenarios, the presentation order of the study vignettes was randomized to alleviate possible ordering and conditioning effects.

Methods

Recruitment and Study Design

An experimental 2×2 factorial web-based study was conducted in September 2018. Participants were recruited via Amazon Mechanical Turk (MTurk), a web-based crowdsourcing platform. MTurk provides a portal through which participants (ie, US citizens aged >18 years) find paid web-based research study opportunities. Participants have dedicated MTurk profiles that contain their demographic information (eg, age, gender, and income). Researchers can indicate specific inclusion and exclusion criteria for their MTurk study and advertise the study on MTurk, and participants are able to choose to participate in studies that interest them in MTurk and also indicate their desired sample size for the study and provide Amazon directly with a payment for all participants' compensation. Participants were provided compensation on completion of the study's survey; payment was transferred directly via Amazon to the participants' preferred accounts.

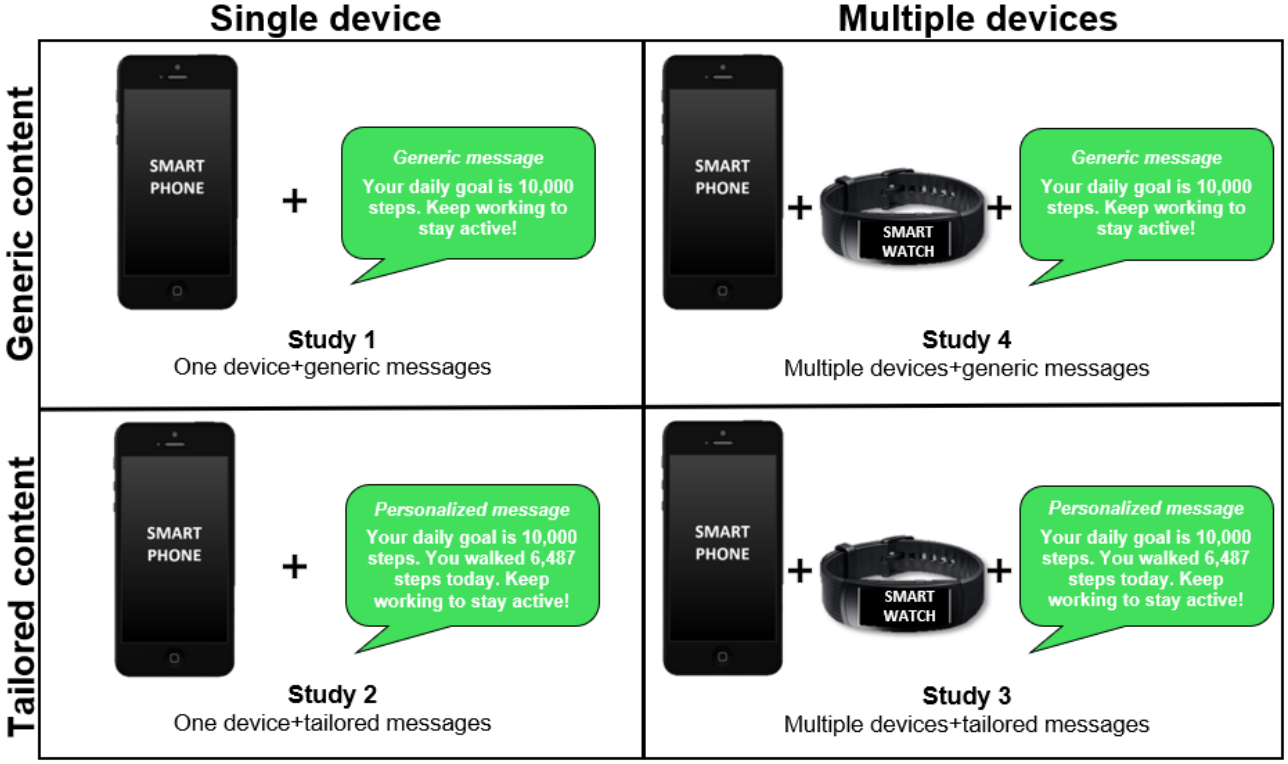
When verified, eligible participants log in to their MTurk account on the web, and they see a list of eligible survey and study opportunities or human interaction tasks (HITs). The compensation provided for completing the study is listed, along with the approximate length of time that the HIT will take to complete, and a brief description of the study. Participants were free to choose the HIT that they were interested in and eligible for from the list.

The name of this study's HIT was *The SmartHealth Preferences Study*. The HIT indicated that the study takes approximately 12 minutes to complete, and participants were paid US $5.00 to complete the entire study. Once they clicked on the HIT in MTurk, participants were taken directly to the study, and informed consent was obtained. The study was hosted on Qualtrics' CoreXM (SAP SE) web-based survey platform via a secure university account.

After obtaining informed consent, participants were presented with a learning module that ensured that they were familiar with the relevant concepts and terminology necessary for completing the study. The learning module defined, in lay terms, a health behavior, specifically, "any action that a person takes that affects their health." Examples included smoking cigarettes—BAD health behavior, healthy eating—GOOD health behavior, binge drinking alcohol—BAD health behavior, staying fit and active—GOOD health behavior. Then, a broad definition of a health behavior change intervention program was described, that is, "any planned series of events, like classes, routines, or group meetings, that aims to encourage its participants to live healthier lifestyles and make healthier decisions for themselves." The example provided described how Jenny Craig and WeightWatchers are health behavior change programs aimed at promoting nutritional and physical health to help participants with weight loss. This was followed by an explanation of what a mobile device is, that is, "examples [of a mobile device] include a smartphone or mobile phone, FitBit or smartwatch, or iPad or tablet computer." A *single device* was described as only needing to carry or use one mobile device (eg, a smartphone). *Multiple devices* was described as needing to carry or use more than one mobile device (eg, a smartphone plus wear a smartwatch). Then, the learning module explained the difference between generic versus personalized program messages, broadly using physical activity as an exemplar behavior for which these intervention messages could be delivered to support. Generic (ie, nontailored) messages were described as, "when all participants in a health behavior change program receive the same messages and notifications." An example included a generic message from a fitness program: "Keep up the good work staying active, you can do it!" Personalized (ie, tailored) messages were described as, "when participants in a health behavior change program receive messages and notifications that are specific to them and their personal goals." An example provided demonstrated a personalized message from a fitness program: "Mary, you walked 10,500 steps today—you beat your personal goal! Keep up the good work staying active!"

As additional examples to visually reinforce and remind participants about the differences between single versus multiple device requirements and generic versus personalized content, participants were shown various graphics. A graphic of just a smartphone was presented to indicate the meaning of *single device* and a graphic depicting a smartphone, plus sign (+), and then a coupled smartwatch was presented to indicate the meaning of *multiple devices*. Two graphics designed to look like a text message notification on a smartphone were displayed to further reinforce the concept of generic versus tailored (ie, personalized) content. Participants were able to read examples of both types of content in the graphics. The generic message picture example read, "Your daily goal is 10,000 steps. Keep working to stay active!" indicating a nontailored message. The personalized message picture example read, "Your daily goal is 10,000 steps. You walked 6,487 steps today. Keep working to stay active!" indicating a tailored message. Figure 1 shows the graphics shown to the participants to reinforce these concepts.

Figure 1. A 2×2 factorial design exploring multiple dimensions of acceptability for mobile health study scenarios featuring different combinations of a single device versus multiple devices and tailored versus generic content.



After completing the learning modules, participants advanced to a new screen in the survey, where they were instructed to imagine themselves being involved in a health behavior change intervention. They were told that they would be presented with different program scenarios where requirements for the number of devices and whether they receive generic or personalized content would be different. Then, four vignettes describing each cell of a 2×2 factorial were delivered in a random sequence for each participant. Hypothetical mHealth intervention studies were described in each vignette, featuring varying combinations of the design factors being studied. Vignette 1 described an mHealth study scenario requiring a single device and provided generic content, vignette 2 was a single device with tailored content, vignette 3 required multiple devices and provided tailored content, and vignette 4 required multiple devices and provided generic content. Each vignette also displayed appropriate graphics from the earlier learning module that were provided on the screen for each hypothetical study scenario to further ensure participants understood the combination of tailoring and device factors featured in the vignette on their computer screen (Figure 1).

Measures

At the end of each mHealth design vignette, participants were asked to rate five dimensions of their a priori acceptability for participating in that particular mHealth design scenario. The specific dimensions chosen for assessment (ie, interest, benefit, enjoyment, utility, confidence, and difficulty) were based on constructs from the TPB (ie, behavioral intentions and behavioral control) as well as previous work assessing behavioral motivation and a priori mHealth design preferences [12]. Participants rated each acceptability dimension using a 5-point Likert scale (eg, “How interesting would participating in this study be?”; 1=not interesting; 5=very interesting). Participants also rated their overall likelihood of participating in the study on a 100-point visual analog slider scale that displayed their indicated value (ie, “if given the chance, how likely would you be to participate in this study?” 0=not at all likely to 100=extremely likely), which was broadly based on the TPB’s construct of intentions (ie, intended likelihood of participating) predicting behavioral engagement (ie, actual participation in the study) [10,12].

Attention checks were also administered to assess data quality and participant comprehension while completing the web-based study. Five careless responder questions were included in the study. These questions asked participants to indicate information that they should have read, had they been paying attention (and not simply clicking through the questions). Using careless responder items in web-based survey research has been shown to be an effective method for confirming the reliability and accuracy of participant data [37]. After reading each vignette, the participant was presented with a careless responder question that asked them to confirm what they just reviewed. For example, after reading the vignette description for the *single device and generic content* scenario, participants were asked the following question: “Please select the response that best describes the requirements for participating in this study.” Then, they were prompted to select the correct response from a list of five possible scenarios.

Hypotheses and Data Analysis

The following hypotheses focused on the two factors of study and were generated before data collection.

Hypothesis 1

There will be a main effect of the number of device requirements factor on all mHealth acceptability dimensions, such that requiring participants to only use a single mobile device, as opposed to requiring multiple devices, will elicit higher levels of a priori acceptability for mHealth intervention participation across seven acceptability dimensions (ie, perceived interest, benefit, utility, enjoyment, confidence, difficulty in participating, and perceived likelihood of participating).

Hypothesis 2

There will be a main effect of the tailoring factor on all mHealth acceptability dimensions, such that providing tailored personalized content to participants, as opposed to generic content, will elicit higher levels of a priori acceptability for mHealth intervention participation across seven acceptability dimensions (ie, perceived interest, benefit, utility, enjoyment, confidence, difficulty in participating, and perceived likelihood of participating).

Exploratory analyses were conducted to examine if there were any significant interaction effects between the different levels of the two factors (ie, tailoring vs generic and single vs multiple devices) and acceptability ratings. Given the lack of previous evidence or strong theory on this issue, no hypotheses regarding interaction effects were made before the study design, data collection, and analysis.

Descriptive statistics were used to evaluate the sample characteristics and the careless responder items. SAS 9.4 (English) data analytic software was used to conduct analyses of variance to evaluate the effect of each hypothetical design scenario on the dimensions of a priori acceptability as well as the overall perceived likelihood of participating in the study. SAS's PROC GLM procedure was used to generate models of the main effects of tailoring and device factors on acceptability dimensions as well as explore any possible interaction effects. No other factors (besides those indicated in the hypotheses and in the description of the exploratory analyses) were included in these models.

Results

Sample Characteristics

A diverse US sample (N=255) of English-speaking adults aged >18 years (female: 104/255, 40.8%; White: 208/255, 81.6%; aged 18-74 years) was recruited (Table 1).

Table 1. Sample characteristics (N=255).

Characteristics	Values
Sex, n (%)	
Female	104 (40.8)
Male	155 (59.2)
Race, n (%)	
People of color	47 (18.4)
White	208 (81.6)
Ethnicity, n (%)	
Hispanic or Latino	15 (5.9)
Non-Hispanic or non-Latino	240 (94.1)
Age (years)	
Range	18-74
Family income (US $), n (%)	
10,000-40,000	107 (41.9)
40,000-100,000	130 (50.9)
>100,000	18 (7.1)
Education, n (%)	
No Bachelor's degree	128 (50.2)
Bachelor's degree	117 (45.9)
Master's degree	10 (3.9)

Study Length and Careless Responder Results

Data collection for all participants was completed within 9 days. Participants answered 95.52% (1218/1275) of the careless responder questions correctly (Table 2).

Table 2. Careless responder question results (N=255).

Questions	Participants, n (%)
Question 1: learning module	
Correct	249 (97.6)
Incorrect	6 (2.4)
Question 2: study scenario 1	
Correct	249 (97.6)
Incorrect	6 (2.4)
Question 3: study scenario 2	
Correct	235 (92.2)
Incorrect	20 (7.8)
Question 4: study scenario 3	
Correct	247 (96.9)
Incorrect	6 (2.4)
Question 5: study scenario 4	
Correct	238 (93.3)
Incorrect	17 (6.7)
Overall correct careless responder questions (n=1275)	1218 (95.52)

Effect of Design Factors on Dimensions of Acceptability

Table 3 provides the main effect results, and Table 4 provides the mean ratings across dimensions of acceptability. There were strong main effects of requiring multiple (ie, 2) devices on reports of the dimensions of acceptability. Requiring multiple devices (vs a single device) led to higher interest in the study (*P*=.008), more perceived benefit (*P*=.04), and greater anticipated utility (*P*=.004) but also lower confidence in performing study requirements (*P*<.001) and significantly higher perceived study difficulty associated with participation (*P*<.001). Requiring multiple devices was unrelated to perceived enjoyment (*P*=.59). Finally, requiring multiple devices was related to a lower likelihood of anticipated study participation (*P*=.02).

Table 3. Main effects of the number of devices and tailoring on dimensions of acceptability.

Acceptability dimension	Main effect (devices)[a]			Main effect (tailoring)[a]		
	Mean square	*F* test (*df*)	*P* value	Mean square	*F* test (*df*)	*P* value
Interest	8.99	7.02 (1)	.008	198.43	154.85 (1)	<.001
Benefit	4.58	4.16 (1)	.04	120.27	109.33 (1)	<.001
Usefulness	9.33	8.30 (1)	.004	163.43	145.43 (1)	<.001
Enjoyment	0.35	0.30 (1)	.59	132.34	111.70 (1)	<.001
Confidence	47.26	40.69 (1)	<.001	17.06	14.69 (1)	<.001
Difficulty	100.06	81.93 (1)	<.001	1.57	1.28 (1)	.26
Likelihood to participate	5305.45	5.83 (1)	.02	58,210.97	63.96 (1)	<.001

[a]The manipulations of tailoring and the number of devices across dimensions of acceptability did not reveal any significant interaction effects.

Table 4. Means across dimensions of acceptability across each cell of the 2 (number of devices)×2 (tailoring) factorial design.

Acceptability dimension and tailoring	Devices, n	
	Single, mean[a] (SD)	Two, mean (SD)
Interest		
Generic	2.77 (1.13)	3.02 (1.23)
Tailored	3.71 (1.05)	3.85 (1.12)
Benefit		
Generic	3.22 (1.10)	3.45 (1.13)
Tailored	4.00 (0.95)	4.05 (1.01)
Usefulness		
Generic	3.09 (1.10)	3.37 (1.16)
Tailored	3.98 (1.00)	4.10 (0.97)
Enjoyment		
Generic	2.78 (1.05)	2.82 (1.14)
Tailored	3.58 (0.98)	3.47 (1.17)
Confidence		
Generic	4.15 (0.99)	3.72 (1.29)
Tailored	4.41 (0.85)	3.98 (1.12)
Difficulty		
Generic	1.84 (1.06)	2.52 (1.23)
Tailored	1.81 (1.04)	2.39 (1.08)
Likelihood to participate		
Generic	53.73 (30.18)	51.65 (32.19)
Tailored	71.40 (27.32)	64.55 (30.79)

[a]All mean ratings are on a 1-5 scale (1=not; 5=very) except for likelihood to participate, which is on a 1-100 scale.

Similarly, there was a robust main effect of tailoring on acceptability, such that interest in the study, perceived benefit, enjoyment, confidence in performing study activities, perceived utility, and overall likelihood of participating in the study were each rated as significantly higher for studies featuring tailored versus generic content (all $P<.001$); in contrast, tailoring had no effect on ratings of perceived difficulty ($P=.26$).

There were no significant interactions between requiring multiple devices and providing tailored content noted for interest ($P=.43$), benefit ($P=.17$), enjoyment ($P=.26$), confidence ($P=.94$), utility ($P=.20$), and difficulty ($P=.47$). Similarly, the interaction of tailored content and requiring multiple devices was not related to the reported overall likelihood of participation ($P=.21$).

Discussion

Principal Findings

The overarching goal of this study was to examine if design factors influence a priori acceptability for participating in mHealth interventions and if this factorial- and vignette-based experimental design approach may be a useful method for examining such factors. More specifically, the study focused on evaluating how two common mobile intervention design elements—requiring multiple devices and providing tailored intervention content—affect a priori participant acceptability for participating in behavior change programs via mobile technology. Given its focus on how thoughts and beliefs may influence behavior, the TPB was used as a theoretical basis for conceptualizing how design decisions may predict acceptability (and subsequent intervention engagement) and informed this study's hypotheses. Overall, the results of this study suggest that potential research participants are sensitive to mHealth study design components and participation requirements.

To date, the limited mHealth implementation research that has focused on mobile design and acceptability has been mostly observational and had participants respond to a particular array of features [12,25,26]. In contrast, this study experimentally manipulated two representative features and obtained ratings on each, thus combining the strengths of experimental and within-person designs. A risk associated with this approach was that viewing multiple arrays of design features (represented in the four vignettes or cells) would lead to order or conditioning effects (eg, some preference for the earlier or later vignettes presented or a carryover effect from one vignette to the next). To minimize this concern, this study presented four design scenarios in random sequences for all participants.

Multiple types of thoughts, some of which are inherently positive (eg, thoughts about beneficial experiences like enjoyment) and negative (eg, thoughts about burdensome experiences like difficulty) can combine to form larger overall beliefs about acceptability. This study focused on this by taking a multidimensional approach to the evaluation of multiple psychological constructs that may be at play when forming perceptions of acceptability. The goal of this study was to evaluate the directionality of these relationships and better understand which elements of mobile design factors (ie, content and number of devices) predict multiple dimensions of a priori acceptability, positively or negatively, in the context of being presented an opportunity to join an mHealth intervention.

Hypotheses

Findings related to the first hypothesis (ie, requiring participants to use multiple mobile devices, as opposed to a single device, would elicit lower levels of a priori acceptability for participating in mHealth) were mixed. Intervention scenarios requiring multiple devices had clear and somewhat unexpected main effects on increasing (as opposed to lowering) perceived interest, benefit, and utility. In line with predictions, the results indicate that mHealth intervention scenarios that require multiple devices increase the perceived burden of participating, as they elicit higher levels of perceived difficulty and lower levels of confidence in fulfilling program requirements and lead to a lower reported overall likelihood of participating. There was also no main effect of requiring multiple- versus single-device requirements on anticipated enjoyment.

Why did the requirement of multiple devices lead to increased ratings on some dimensions of acceptability (ie, interest, benefit, and utility)? Perhaps participants generally view interventions with multiple devices as superior; perhaps such studies are seen as more scientific, legitimate, or reflect more innovative and potentially useful approaches to tracking and addressing behavior change. If present, such effects may have possibly been enhanced by the within-person factorial design; that is, as each person saw all the vignettes, it may have implicitly led to more direct contrasts between the design features (and led to a *more devices must be better* evaluation). Other dimensions of acceptability were in accordance with predictions, and common sense, in that participants appeared appropriately sensitive to the additional requirements of caring for and managing multiple devices over time (eg, charging smartphone batteries frequently and taking off wearable sensors to bathe and then replacing them) as functional barriers to joining an mHealth study. This is similar to previous posteriori research that suggests that requiring participants to use multiple devices over the course of a program may reduce acceptability [32]. More contextual data (eg, qualitative feedback) in future studies may be helpful to understand the effect requiring multiple devices, compared with a single device, has on a priori mHealth acceptability.

Findings related to the second hypothesis (ie, providing tailored content to participants, as opposed to generic content, would elicit higher levels of a priori acceptability for participating in mHealth) were largely supported by these results. Intervention scenarios featuring tailored content had strong main effects for enhancing perceived interest, benefit, enjoyment, utility, and higher intended participation compared with interventions providing only generic content. Research suggests that tailored content may be perceived by participants as having enhanced value (compared with generic content) because it can provide information specific to their individual health behavior needs [1,33,34]. Tenets of the TPB can be used to interpret these findings, such that designing mHealth interventions to provide tailored content to individuals (as opposed to generic messaging) may elicit personally beneficial thoughts and beliefs (ie, high acceptability) about participating. This in turn may enhance prospective participant motivation for signing up for mHealth studies featuring tailored content, which may translate to a higher likelihood of engaging with behavior change delivered via mobile technology. Although plausible, future research is needed to determine if high acceptability for tailored interventions at baseline before the intervention (a priori) actually translates to actual persistence and effective adoption of behavior change strategies throughout mHealth study participation.

There was no main effect of tailored content on perceived difficulty. This suggests that providing personalized versus generic content does not affect perceptions of effort or burden for participating in mHealth research. The hypothetical manipulation of tailoring in this study was a requirement for reading intervention messaging that is either tailored or generic. It is reasonable that participants do not perceive such tasks as being differentially burdensome, as there were no meaningful differences (eg, in length) between the two conditions. It is also possible, however, that the learning module did not provide enough information about additional requirements that sometimes accompany participation in tailored mHealth interventions. For example, these interventions may entail components such as active tracking, additional app installations, and syncing devices. These (and similar) tasks may increase the perceived burden of participating in tailored mHealth programs, and future studies may provide more specificity about such requirements to further investigate how tailoring may affect perceived difficulty.

There were no significant interaction effects between requiring multiple devices and providing tailored content noted for any of the dimensions of acceptability. Given the limited past research examining these two specific design factors together in an experimental fashion, this was an exploratory hypothesis. The results suggest that participants' perceptions of acceptability for multiple devices and tailoring factors are not contingent on the presence of one factor or the other. Although no evidence was found about these two design elements in particular interacting with one another, perhaps other design features (or more extreme versions of these features; eg, carrying 4 devices) may show interactions. As such, this is seen as an important open issue for future mHealth design research, and additional theoretical considerations (eg, what elements of design features and their implications are thought to interrelate) and empirical evidence (eg, additional features, combinations, and levels) should be considered.

Limitations and Future Directions

There are notable limitations to this study. Amazon's web-based MTurk platform was used for the identification, enrollment, and compensation of research participants. MTurk served as a rapid recruitment service (N=255 participants within 9 days). It afforded access to a diverse sample of respondents (female: 104/255, 40.8%; White: 208/255, 81.6%; aged 18-74 years), and our findings (ie, careless responder items answered correctly: 1218/1275, 95.52%) suggest high data quality in this web-based experiment. In recent years, MTurk has been increasingly used in social and behavioral science research because of its ability to rapidly recruit and compensate large pools of research subjects [38]. However, there is also an ongoing debate as to whether research findings from MTurk participant pools are generalizable to the rest of the US population [39,40]. One issue is that samples recruited from MTurk generally exhibit higher educational attainment than the actual US population (ie, 35% of the US population possesses bachelor's degrees as of 2018 [41]). This study followed a similar suit, as 49.8% (127/255) of the sample recruited via MTurk possessed a bachelor degree or higher.

This study investigated hypothetical mHealth design scenarios using web-based vignettes. This was useful for examining how design decisions may affect a priori (ie, prospective) acceptability, which may be related to self-selection into mHealth research or motivation to fulfill intervention requirements. Given the hypothetical nature of the vignettes and a priori design, one limitation is that this study did not allow for examination of acceptability posteriori (ie, after participants actually engage with a real mHealth intervention in the field). It is unclear how acceptability (even if it is high before engagement) changes over time and whether differences in design factors affect posteriori acceptability in a similar direction as a priori acceptability. For example, perhaps participants are highly motivated to participate in an intervention that provides tailored content, but over time, their acceptance and motivation may decline. Real-world engagement with mHealth behavior change programs may also enhance acceptability above and beyond baseline measures. In addition, because of the hypothetical nature of the study vignettes, these findings do not directly provide information about how a priori acceptability is associated with other measurable implementation outcomes (eg, actual adherence and sustained engagement). Longitudinal experimental designs within randomized controlled mHealth trials (eg, as opposed to hypothetical web-based study scenarios, as used here) are needed to better understand these relationships. A quasi-experimental study of mHealth design and acceptability (ie, manipulating mHealth design features and assessing both a priori and posteriori acceptability) may also help understand how mHealth acceptability may change over time between different design conditions. Each of these are important questions that could be examined in future real-world mHealth implementation research.

Another limitation to the design of this study was that participant health was not assessed. As such, it is unclear whether the sample was at risk of behavioral health challenges. This makes further generalizability of these findings to patient populations with health behavior challenges unclear, as there may be reliable differences across meaningful subsamples (eg, sick individuals may be more highly motivated to pursue treatment and thus more *tolerant* of high burden design features). Future mHealth implementation studies (including those conducted via web-based surveys) should screen and select participants to ensure that the findings are applicable to the target population of interest. Another approach would be to consider enrolling healthy controls to match patient participant groups to investigate whether differences in mHealth acceptability exist depending on the saliency of health behavior change needs.

Finally, to convey the meaning of generic versus tailored content delivered via mHealth to support behavior change, participants were provided with a relatable example of a behavior to which this messaging could be applied. Physical activity (in the form of *step count*) was used as an exemplar health behavior to describe differences between generic and tailored messaging. This behavior was selected under the assumption that most participants would likely be at least vaguely familiar with the concept of tracking activity, especially given increasing trends and publicity worldwide for wearing fitness trackers (eg, Fitbit) and smartwatches (eg, Apple Watch) to monitor health [42]. However, physical activity is a complex health behavior, and it is possible that participants perceived activity-related behaviors as more difficult to change compared with other familiar health behaviors (eg, medication adherence and applying sunscreen). Acceptability ratings and the impact of design factors may be differentially related across different behaviors, and the features and characteristics of these behaviors that may drive any divergent associations is an important topic for future consideration. More generally, it is unclear which behavior change outcomes mHealth is best suited to address to support buy-in and engagement in target populations (eg, Is an mHealth intervention targeting physical activity more or less acceptable than one focused on healthy eating?). Some research has shown disparities in participant engagement with mHealth-based behavior change interventions based on demographic moderators (eg, age and health literacy) [43], and such differences between participants may also play a role in perceptions of acceptability for supporting different types of behavior change with technology. Future research should further evaluate how acceptability for participating in mHealth may change differentially depending on the intervention's target behaviors, as well as explore other possible moderators inherent to target health behaviors that may influence this relationship (eg, frequency of behavior, intensity of behavior, and perceived behavioral norms).

Conclusions

This study showed that differences in intervention design factors, specifically the number of devices required and tailoring, affect various dimensions of participant acceptability for engaging in behavior change programming via mHealth. Some limited previous work examining acceptability for these design factors was conducted posteriori (and studied these factors separately); however, understanding a priori perceptions is advantageous for designing mobile interventions that support up-front buy-in and acceptability from individuals for participating in mHealth research. Previous implementation research and theoretical models from behavioral science suggest that acceptability is a

predictor of the uptake and utilization of behavior change interventions. This study supports the hypothesis that participants are sensitive to the mHealth design decisions made by researchers, specifically those related to the provision of the number of devices required for participation and tailored content. The results show that requiring participants to use multiple devices increases perceived burden and reduces their overall reported likelihood of participating and that providing participants with tailoring in mobile behavior change interventions is beneficial for enhancing a priori acceptability.

This work also contributes to the early development of evidence-based recommendations in mHealth design to support participant acceptability. More generally, this method serves as proof of concept for how other design features can be tested—singly or in combinations of factors—and using assessments of a priori acceptability, posteriori acceptability, or both. As the need for investigating mHealth implementation factors continues to emerge, researchers in this space may also consider MTurk as an inexpensive and effective recruitment tool with a wide reach and reliable, attentive respondents in cases where unique sample characteristics are not essential. Future studies on the relationship between mHealth design elements and participant motivation to engage with mHealth should further explore multiple dimensions of acceptability experimentally in controlled trials in both clinical and community samples. Overall, this will afford more reliable data on the effects these common mHealth design elements have on influential mHealth implementation factors (eg, not just a priori acceptability, but also posteriori acceptability, real-world compliance, and adherence).

Authors' Contributions

FTM and JMS contributed to the design of this study, data collection, cleaning and analysis, and writing and revisions of this manuscript before submission.

Conflicts of Interest

None declared.

References

1. 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]
2. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83. [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
3. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232]
4. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. J Pers Med 2014;4(3):412-423 [FREE Full text] [doi: 10.3390/jpm4030412] [Medline: 25563359]
5. 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 Sep;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
6. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015 Sep 16;3:32 [FREE Full text] [doi: 10.1186/s40359-015-0089-9] [Medline: 26376626]
7. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010 Mar;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
8. Acceptability. Oxford Dictionary. 2018. URL: https://en.oxforddictionaries.com/definition/acceptability [accessed 2018-11-13]
9. Synonyms and antonyms. Merriam-Webster Dictionary. 2018. URL: https://www.merriam-webster.com/thesaurus/acceptability [accessed 2018-11-12]
10. Materia F, Downs DS, Smyth J. Conceptualizing acceptability in mHealth research. In: Proceedings of the 6th Biannual Meeting of the Society for Ambulatory Assessment. 2019 Presented at: 6th Biannual Meeting of the Society for Ambulatory Assessmen; June 19-22, 2019; Syracuse, NY.
11. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
12. Downs DS, Smyth JM, Heron KE, Feinberg ME, Hillemeier M, Materia FT. Beliefs about using smartphones for health behavior change: an elicitation study with overweight and obese rural women. J Technol Behav Sci 2018 Dec 15;4(1):33-41. [doi: 10.1007/s41347-018-0081-3]
13. Materia FT, Smyth JM, Heron KE, Hillemeier M, Feinberg ME, Fonzi P, et al. Preconceptional health behavior change in women with overweight and obesity: prototype for SMART strong healthy women intervention. mHealth 2018 Jul;4:24 [FREE Full text] [doi: 10.21037/mhealth.2018.06.06] [Medline: 30148139]
14. Adell E, Varhelyi A, Nilsson L. The definition of acceptance and acceptability. In: Driver Acceptance of New Technology. Boca Raton, Florida, United States: CRC Press; 2014:11-19.

15. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121. [doi: 10.2196/jmir.5250] [Medline: 27245693]
16. Bull S. Beyond acceptability and feasibility: moving mHealth into impact. mHealth 2016 Dec 19;2:45 [FREE Full text] [doi: 10.21037/mhealth.2016.12.02] [Medline: 28293615]
17. 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]
18. Chau JP, Lee DT, Yu DS, Chow AY, Yu W, Chair S, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. Int J Med Inform 2012 Oct;81(10):674-682. [doi: 10.1016/j.ijmedinf.2012.06.003] [Medline: 22789911]
19. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015 Sep 22;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
20. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014 Aug 05;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
21. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
22. 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]
23. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
24. 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 15;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
25. Cajita MI, Gleason KT, Han H. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):10-22. [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
26. Miah SJ, Gammack J, Hasan N. Extending the framework for mobile health information systems Research: a content analysis. Inf Sys 2017 Sep;69:1-24. [doi: 10.1016/j.is.2017.04.001]
27. Kalke K, Ginossar T, Bentley J, Carver H, Shah S, Kinney A. Use of evidence-based best practices and behavior change techniques in breast cancer apps: systematic analysis. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e14082 [FREE Full text] [doi: 10.2196/14082] [Medline: 32012084]
28. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
29. Jones V, Gay V, Leijdekkers P. Body sensor networks for mobile health monitoring: experience in Europe and Australia. In: Proceedings of the Fourth International Conference on Digital Society. 2010 Presented at: Fourth International Conference on Digital Society; Feb. 10-16, 2010; Saint Maarten, Netherlands Antilles. [doi: 10.1109/icds.2010.41]
30. Lymberis A, Dittmar A. Advanced wearable health systems and applications. IEEE Eng Med Biol Mag 2007 May;26(3):29-33. [doi: 10.1109/memb.2007.364926] [Medline: 17549917]
31. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. Pharm Ther 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
32. Shaw RJ, Steinberg DM, Bonnet J, Modarai F, George A, Cunningham T, et al. Mobile health devices: will patients actually use them? J Am Med Inform Assoc 2016 Jan 17;23(3):462-466. [doi: 10.1093/jamia/ocv186]
33. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychological Bulletin 2007;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673]
34. Rimer B, Kreuter M. Advancing tailored health communication: a persuasion and message effects perspective. J Commun 2006;56(1):184-201. [doi: 10.1111/j.1460-2466.2006.00289.x]
35. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013 Sep;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
36. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR mHealth uHealth 2016 Jun 09;4(2):e29. [doi: 10.2196/mhealth.5127] [Medline: 27282195]
37. Schneider S, May M, Stone AA. Careless responding in internet-based quality of life assessments. Qual Life Res 2018 Apr 16;27(4):1077-1088. [doi: 10.1007/s11136-017-1767-2] [Medline: 29248996]
38. Chan C, Holosko MJ. An overview of the use of mechanical turk in behavioral sciences. Res Soc Work Pract 2015 Jul 06;26(4):441-448. [doi: 10.1177/1049731515594024]

39. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018 Apr 4;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
40. Mullinix KJ, Leeper TJ, Druckman JN, Freese J. The generalizability of survey experiments. J Exp Polit Sci 2016 Jan 12;2(2):109-138. [doi: 10.1017/XPS.2015.19]
41. Rates of high school completion and bachelor's degree attainment among persons age 25 and over, by race/ethnicity and sex: selected years, 1910 through 2018. National Center for Education Statistics. URL: https://nces.ed.gov/programs/digest/d18/tables/dt18_104.10.asp [accessed 2019-12-12]
42. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2020-07-14]
43. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101. [doi: 10.2196/jmir.6838] [Medline: 28381394]

Abbreviations

HIT: human interaction task
mHealth: mobile health
MTurk: Mechanical Turk
TPB: theory of planned behavior

Edited by L Buis; submitted 06.08.20; peer-reviewed by I Kim, H Jiang; comments to author 06.10.20; revised version received 28.03.21; accepted 14.05.21; published 26.07.21.

Please cite as:
Materia FT, Smyth JM
Acceptability of Intervention Design Factors in mHealth Intervention Research: Experimental Factorial Study
JMIR Mhealth Uhealth 2021;9(7):e23303
URL: https://mhealth.jmir.org/2021/7/e23303
doi:10.2196/23303
PMID:34309563

Original Paper

A Smart Shoe Insole to Monitor Frail Older Adults' Walking Speed: Results of Two Evaluation Phases Completed in a Living Lab and Through a 12-Week Pilot Study

Antoine Piau[1,2,3], MD, PhD; Zara Steinmeyer[1], MD; Yoann Charlon[2], PhD; Laetitia Courbet[4], PhD; Vincent Rialle[4], PhD; Benoit Lepage[5], PhD; Eric Campo[2], PhD; Fati Nourhashemi[1,3], MD, PhD

[1]Gerontopole, University Hospital of Toulouse, Toulouse, France

[2]Laboratoire d'analyse et d'architecture des systèmes, Centre national de la Recherche Scientifique, Toulouse, France

[3]Unité 1295, Institut National de la Santé Et de la Recherche Médicale, Toulouse, France

[4]Autonomie, Gérontologie, E-santé, Imagerie et Société, Grenoble Alps University, Grenoble, France

[5]Department of Medical Information, University Hospital of Toulouse, Toulouse, France

Corresponding Author:
Antoine Piau, MD, PhD
Gerontopole
University Hospital of Toulouse
24 rue du pont St Pierre
Toulouse, 31400
France
Phone: 33 561323010
Fax: 33 561325265
Email: antoinepiau@hotmail.com

Abstract

Background: Recent World Health Organization reports propose wearable devices to collect information on activity and walking speed as innovative health indicators. However, mainstream consumer-grade tracking devices and smartphone apps are often inaccurate and require long-term acceptability assessment.

Objective: Our aim is to assess the user acceptability of an instrumented shoe insole in frail older adults. This device monitors participants' walking speed and differentiates active walking from shuffling after step length calibration.

Methods: A multiphase evaluation has been designed: 9 older adults were evaluated in a living lab for a day, 3 older adults were evaluated at home for a month, and a prospective randomized trial included 35 older adults at home for 3 months. A qualitative research design using face-to-face and phone semistructured interviews was performed. Our hypothesis was that this shoe insole was acceptable in monitoring long-term outdoor and indoor walking. The primary outcome was participants' acceptability, measured by a qualitative questionnaire and average time of insole wearing per day. The secondary outcome described physical frailty evolution in both groups.

Results: Living lab results confirmed the importance of a multiphase design study with participant involvement. Participants proposed insole modifications. Overall acceptability had mixed results: low scores for reliability (2.1 out of 6) and high scores for usability (4.3 out of 6) outcomes. The calibration phase raised no particular concern. During the field test, a majority of participants (mean age 79 years) were very (10/16) or quite satisfied (3/16) with the insole's comfort at the end of the follow-up. Participant insole acceptability evolved as follows: 63% (12/19) at 1 month, 50% (9/18) at 2 months, and 75% (12/16) at 3 months. A total of 9 participants in the intervention group discontinued the intervention because of technical issues. All participants equipped for more than a week reported wearing the insole every day at 1 month, 83% (15/18) at 2 months, and 94% (15/16) at 3 months for 5.8, 6.3, and 5.1 hours per day, respectively. Insole data confirmed that participants effectively wore the insole without significant decline during follow-up for an average of 13.5 days per 4 months and 5.6 hours per day. For secondary end points, the change in frailty parameters or quality of life did not differ for those randomly assigned to the intervention group compared to usual care.

Conclusions: Our study reports acceptability data on an instrumented insole in indoor and outdoor walking with remote monitoring in frail older adults under real-life conditions. To date, there is limited data in this population set. This thin

instrumentation, including a flexible battery, was a technical challenge and seems to provide an acceptable solution over time that is valued by participants. However, users still raised certain acceptability issues. Given the growing interest in wearable health care devices, these results will be useful for future developments.

Trial Registration: ClinicalTrials.gov NCT02316600; https://clinicaltrials.gov/ct2/show/NCT02316600

(JMIR Mhealth Uhealth 2021;9(7):e15641) doi:10.2196/15641

KEYWORDS

frail older adults; walking speed; outpatient monitoring; activity tracker; shoe insert

Introduction

Frailty is an age-related syndrome characterized by a decline in biological reserves with an increased risk of impaired autonomy and death [1]. As such, implementing intervention programs promoting physical activity is essential in preventing functional decline [2-4]. To measure the efficacy of these programs, it seems necessary to monitor frailty indicators and adherence over time [5,6]. Frailty was defined according to the five Fried criteria (slow gait speed, low physical activity, unintentional weight loss, exhaustion, and muscle weakness) [1]. Participants with a score of 0 were robust, 1 to 2 were considered prefrail, and 3 to 5 frail.

Among these criteria, consistent data indicates that walking speed is one of the strongest to predict adverse outcomes [7,8]. Currently, a patient walking speed is evaluated during clinical consultations by manually measuring the time the patient takes to walk 15 feet. This discrete assessment often fails to detect changes in day-to-day walking speeds and does not reflect walking speeds in everyday environments. As such, continuous ambulatory monitoring would ensure precise monitoring of a patient's health and better support medical diagnosis, especially by capturing a patient's decrease in physical activity and walking speed profile [9].

Currently, physical activity assessments are often based on self-reported questionnaires with poor or inconsistent reliability [10,11], highlighting the importance of objective measures. The World Health Organization's reports on aging and health propose the use of wearable devices to collect information on physical activity and gait speed as health indicators [5]. Digital technologies allow the monitoring of patient's physiological data in their environment and thus tracking of subtle changes over time [9,12-14]. For example, accelerometers provide an objective measure of physical activity over a few days compared to standard physical performance measures [15]. Moreover, physical activity feedback with wearable sensors may also be incentive to increase daily activity [16-19].

Several sensor-based tools have been proposed to assess frailty and walking speed. However, they do not allow monitoring walking speed and activity in real-life conditions over long periods of time both indoors and outdoors [20,21]. To date, walking analysis research in patients who are frail is limited, and most studies involve electronic walkways, camera systems, or force plates, which limits real-life monitoring [9,22].

Thus, assessing a device specifically designed to monitor this population is relevant. Currently, multiple consumer-grade monitoring devices are commercialized, such as wrist-worn fitness trackers or smartphone apps. However, their accuracy is under debate especially in monitoring gait speed in older populations [19,23-25]. As such, their results must be interpreted cautiously. A majority of these devices evaluate activity by a built-in sensor counting one's steps and thus does not help characterize a patient's type of walk from normal walking to shuffling.

Our hypothesis is that a shoe insole is acceptable for long-term monitoring of indoors and outdoors walking speed in real-life conditions. A wireless insole is discrete and does not stigmatize the patient, users do not have to remember to wear it every day because it is placed in the patient's walking shoes, and studies have shown that inertial feet sensors can accurately measure walking speed [25-27]. However, acceptability beyond a few hours of testing is not yet reported. The objectives of our multiphase study is to assess the technical feasibility (eg, wireless transmission and calibration protocol) and the acceptability from the user's perspective. The secondary objective is to describe the evolution of a patient's frailty syndrome and functional autonomy in both groups, quality of life, and health costs.

Methods

Study Design and Setting

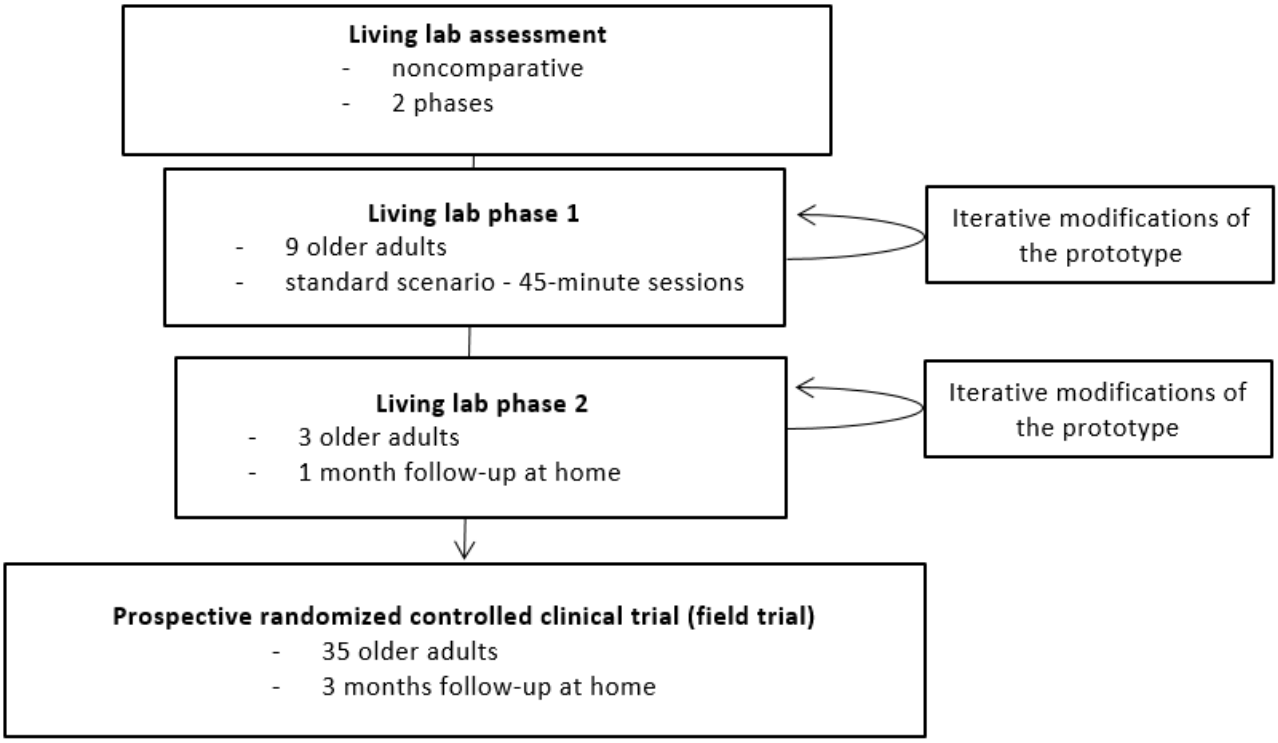
In line with codevelopment and health technology assessment recommendations [28-30], we designed a two-phase pilot study [31] involving community-dwelling prefrail and frail older participants remotely followed at home. It consisted of, first, a noncomparative trial in a living lab and a 12-week prospective, parallel, randomized controlled clinical trial (field trial). The first trial lasted from September 2015 to January 2016. The second lasted from October 2016 to January 2019.

The living lab experiment was set up in a living lab at the University Institute of Technology in Blagnac, France (Maison Intelligente de Blagnac) located on the university campus [26]. This flat of 70 m^2 is equipped with a networking infrastructure accessible to valid, frail, or disabled persons. It enables testing of technological devices in an environment similar to one's home setting but with controlled technical conditions. The living lab phase was carried out in two subphases: (1) 9 older adults were evaluated in the Maison Intelligente de Blagnac living lab and (2) 3 older adults among the 9 participants of the living lab were followed up with at home for a month.

During the first phase, participants were invited to complete a single session standard scenario in the Maison Intelligente de Blagnac living lab (45-minute sessions consisting of a walking

tour in the living lab and outside, as per usual) after calibrating the insole to their own step length and after a 10-minute presentation of the product's objectives. During the second phase, volunteers were asked to use the device over a 1-month time period in their own homes (Figure 1). Participants were invited to wear the insole and use a touchpad without any additional instructions or training.

Figure 1. Description of the evaluation phases.



For the field trial, 35 participants were randomized following a 2:1 allocation into two distinct arms. Randomization was conducted independent of recruitment by the hospital's epidemiology department. The aim of randomization was to describe the secondary objectives for future effect size calculation. Participants were randomized in either the intervention (smart insole follow-up) or control group (standard follow-up) and enrolled for 12 weeks.

Intervention Description

Participants in the living lab phase and those randomized in the intervention group in the field trial phase were equipped with the instrumented shoe insole and were given a touchpad feedback app. The technological devices tested were a pair of insoles (only one insole is instrumented), a touch pad to collect data from the insole and inform the users about their activity (Bluetooth communication), and an induction charger to charge the shoe every night (Figure 2). The insole thickness is less than 2.5 millimeters at its thickest point (arch). The insole measures according to time (day, week, and month) the number of steps and walking distance, the average walking speed during walking periods, and active walking duration (as opposed to shuffling). Active walking was defined as continuous walking for at least 5 minutes with a tolerance of 1 minute (see Figure 3) considering that health benefits of aerobic exercise begin with any increase above the lowest levels of activity [32]. Moreover, the insole is calibrated on each participant's step length for an accurate measurement of walking speed, unlike consumer-grade devices. The walking speed measure algorithms and the lab test results of the insole have been previously described [26].

Both groups benefitted from a frailty assessment at a geriatric day hospital for frailty (GHF) as per usual follow-up. Indeed, GHFs are developed nationwide in France following the French national authority recommendations and may accommodate up to five patients per day. Patients benefit from a follow-up phone call at 3 months and at 1 year by a nurse.

Figure 2. Overall technological device description. It includes a pair of insoles, an induction charger fitting in the shoe, a touchpad to collect data from the insole and provide feedback to the user, a secure remote database, and a web application for the patient and the physician. The insole is 2.5 mm at its thickest point (arch); it has a buffer memory and a flexible battery for walking comfort. If the battery is not recharged, an alert is issued to the user. The touchpad is presented here with a diagram of average walking speed and a diagram reporting active walking minutes.

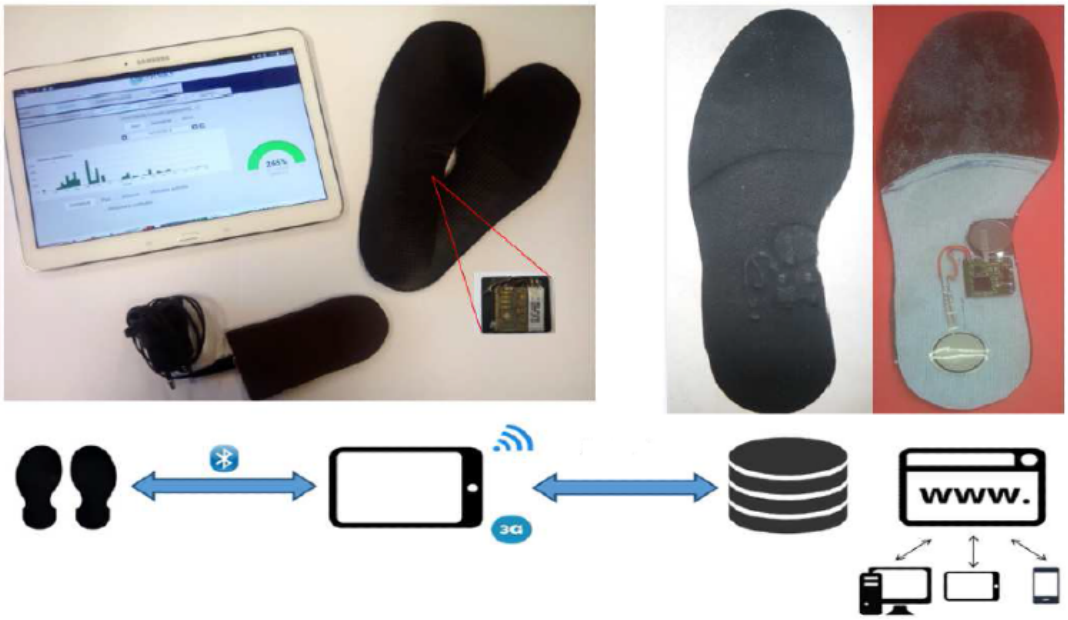
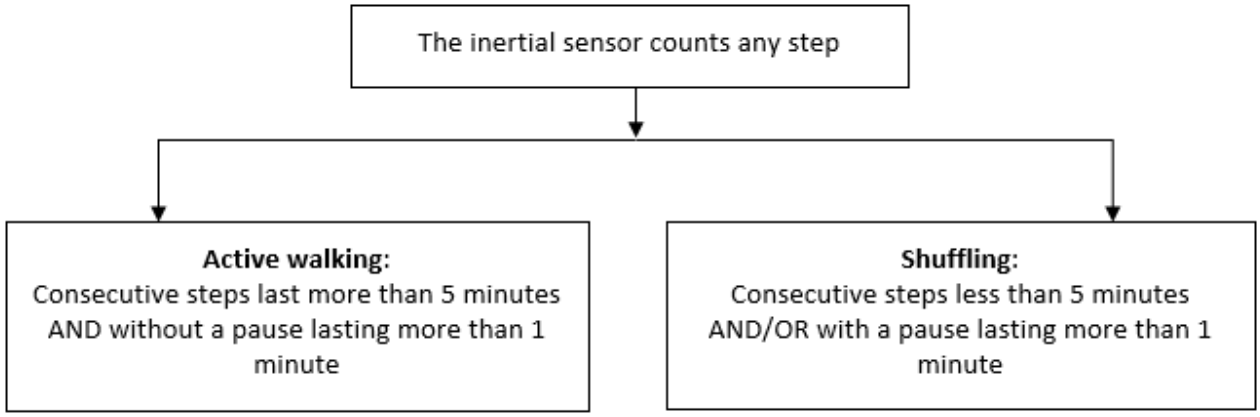


Figure 3. Active walking definition. The insole accounts for steps in any case.



Recruitment and Eligibility Criteria

Study participants were recruited through the Toulouse University Hospital GHF (France). Ethical approval for the study was obtained from the regional independent ethics committee in September 2014 (ID-RCB: 2014-A00523-440). The trial is registered on ClinicalTrials.gov (NCT02316600). All participants provided written consent. Baseline assessments were conducted in-person at the GHF. The inclusion criteria were (both phases):

- Patients 65 years or older living independently at home
- Activity of Daily Living (ADL) score (ranges from 0 to 6; the higher the score, the higher the level of functional autonomy in daily life is; eg, walking or dressing) [33] of 4 or higher
- Mini-Mental State Examination (MMSE; ranges from 0 to 30; the higher the score, the higher the level of global cognition) [34] of 24 or higher
- Prefrail or frail according to Fried criteria [1], for the field trial only

The sole exclusion criteria was life expectancy of less than 12 months. We did not request any specific level of computer literacy, and none of the participants were familiar with digital tools. There was no financial compensation for participation.

Data Collection Procedure

At inclusion, sociodemographic data (gender, age, marital status, living place), frailty status [1], functional abilities with the ADL [33], and cognition with MMSE [34] were collected. For the field trial, other functional and physical scores were also assessed (Instrumental Activities of Daily Living [IADL] [35] and Short Physical Performance Battery [SPPB] scores [36], respectively).

In the living lab phase, semistructured interviews were conducted, focusing on technical feasibility and acceptability. We also scored the 3 *home participants'* satisfaction with the

device at the end of the 1-month follow-up using the main themes identified during the interviews.

For the field trial, the primary outcome measured the acceptability of the device as follows:

- Qualitative indicators: acceptability assessment questionnaire based on the Quebec User Evaluation of Satisfaction With Assistive Technology scale [37] aimed to evaluate technology satisfaction degree from 1 to 5 (1: not satisfied at all; 5: very satisfied) and semistructured interview of the first 10 participants who completed follow-up.
- Quantitative indicators: number of days wearing the insole and average time of wearing the insole per day as declared by the participants and as measured by the insole and the number of connections to the app

The secondary outcome of the field trial measured:

- Frailty status according to the Fried criteria (slow gait speed, low physical activity, unintentional weight loss, exhaustion, and muscle weakness) [1]
- Lower extremity physical performance assessed with the SPPB (consisting of a balance test, a 3-meter gait test, and a 5 chair rises test; score ranging from 0 to 12; 12 indicating the highest degree of functioning) [36]
- ADL [33]
- IADL (ranging from 0 to 8; the higher the score is, the better the participant's functional autonomy in daily life is; eg, driving) [35]
- Quality of life questionnaire (36-Item Short Form Health Survey [SF-36]; ranges between 0 and 100; greater score indicates better health-related quality of life) [38]
- EQ-5D-3L index, recording the patient's self-rated health in a five-digit health state profile, each comprised of three levels (the score is converted into a single summary number ranging from −0.59 to 1; 1 indicating the highest level of perceived health) [39]
- Major medical events defined as any event leading to a hospitalization or an emergency admission
- Health costs

The semistructured interview was proposed to each participant at the end of the follow-up date (45 minutes for the living lab test, 1 month for the 3 living lab *home testers*, 3 months for the field test). The interview was face-to-face for the living lab group and by phone call for both home evaluations. Two researchers of the Age-Imaging-Modelization Laboratory (LC and VR; Joseph Fourier University Sociology Laboratory, Grenoble, France) conducted them. Each interview lasted 2 hours and was designed to explore key questions relating to acceptability. For the first phase of the study (living lab), nondirective exchanges were conducted to identify recurring themes. These themes guided the interviews of the field trial participants (n=10). All interviews were transcribed. The transcripts were analyzed using a conventional content analysis along with a summative qualitative content analysis [40].

For the field phase, a clinical research assistant also contacted participants at 1, 2, and 3 months to evaluate secondary outcomes. For acceptability, the following question was systematically asked: "how well did you tolerate wearing the insole during the past month?" (5 possible answers were proposed: from totally tolerable to totally intolerable).

Statistical Analysis

For the field trial, we planned to include 35 participants complying with pilot study recommendations [41]. Qualitative variables were presented by effectives and percentages, quantitative variables by means and SDs. Likert-type items were handled as continuous variables.

Comparison tests were performed: quantitative variables with Wilcoxon-Mann-Whitney test and chi-square or Fisher tests for qualitative variables. Missing data was replaced with the mean values of the groups, allowing complete case analysis. A drawback of this approach is reduced variability and weakening of covariance and correlation estimates in the data.

The Department of Epidemiology of the Toulouse University Hospital in Toulouse conducted statistical analyses using Stata version 14.2 (Stata Corp).

Results

Living Lab

Study Sample Characteristics

A total of 9 participants were included, 6 (67%) women and 3 men. The mean age was 70.1 (SD 2.3) years (range 65-75), and none of them presented any functional disability (Table 1).

Table 1. Characteristics of participants (n=9).

Characteristics	Participants
Gender (female), n (%)	6 (67)
Age (years), mean (SD)	70.1 (2.3)
Activity of Daily Living score, mean (SD)	5.9 (0.3)
Mini-Mental State Evaluation score, mean (SD)	29.6 (0.5)
Frailty status (frail or prefrail), n (%)	1 (11)

Acceptability

The research assistant of the study informed participants in the living lab in the following terms: "Walking every day is beneficial to your health (e.g., independence, cognition) and the insole is designed to quantify your activity and provide feedback that can help you progress and make you want to walk more." Interview feedback revealed participants' understanding

of the device's objectives: 7 participants out of 9 cited *prevent the risk of becoming dependent* as a main objective, 5 *physical activity follow-up*, and 4 *motivate us to be physically active*. A total of 6 participants stated that wearing the insole could potentially encourage them to walk more. Indeed, real-time feedback would "motivate them to go out from home" and "stimulate their desire to adopt a healthier lifestyle." A total of 4 participants suggested that reminder texts would increase physical activity motivation if activity fell below objectives. There were 5 other participants that were confident in achieving the walking threshold without any help. One person reported that such a reminder would be intrusive. None of the participants expressed concerned with the calibration protocol. Most participants (n=8) claimed that the insole was comfortable, light, and robust, and did not cause any discomfort. Nevertheless, flexibility and thickness were negatively highlighted by 2 of them with the fear of possible long-term use discomfort. Concerning the user touchpad interface, 5 participants expressed the need for a longer time practicing to assess acceptability. Several major themes were identified (example codes are given in parenthesis):

- Understanding and adhesion to the device's objectives (eg, facilitating role)
- Device acceptability (eg, comfort)
- Device usability (eg, reliability)
- Device adherence (eg, time wearing the insole)

The 3 *home participants* reported wearing the insole for 1 month without early dropout. According to the insole data, they wore the insole for 22, 13, and 13 days out of 30, respectively. Total "active walking" time (as opposed to shuffling and therefore different from the number of steps) was 29 hours and 33 minutes (7288 average steps per day), 4 hours and 11 minutes (average steps per day 1748), and 3 hours and 52 minutes (average steps per day 1574), respectively. Average walking speed (calculated from the participants average step length obtained during the calibration protocol and cadence measurement) was 0.90 ms^{-1}, 0.75 ms^{-1}, and 0.69 ms^{-1}.

All participants claimed that the insole was comfortable and that "once they were placed in the subject's shoes, you tend to forget them." According to users, the most interesting feedback information was walking distance (more than number of steps, walking speed, or the active walking time). Some texts of the interface were considered too small and colors not appropriate by 1 participant with age-related visual impairment. Of the 3 participants, 2 declared that they would use this device if it was commercialized. No harms or unintended side effects were reported. Table 2 summarizes the 3 *home participants'* satisfaction with the device at the end of the 1-month follow-up. At the end of the interview, we asked them to rate their overall satisfaction with the device (score ranging from 0 "strongly disagree" to 6 "strongly agree"). The mean score was 3.1 (out of 6) for the *facilitating role*, 4.3 for *user friendliness*, and 2.1 for *reliability outcomes*.

Concerning technical feasibility, battery autonomy at home ranged from 30 to 48 hours, and no serious concern was raised about the induction charging system. Participants were satisfied with the daily routine, especially since there is no connection to be made for charging. One insole instrumentation was broken after 3 weeks of use. The 3 participants encountered synchronization problems between the insole and the touchpad (long latency).

Table 2. Participants satisfaction with the device at the end of the 1-month follow-up.

Participants satisfaction	Participant ratings[a], mean (individual)
The device helps me to achieve my objectives (facilitating role)	3.7 (5, 3, 3)
It motivates me to complete my activities (facilitating role)	3.0 (1, 3, 5)
It helps me to be more efficient (facilitating role)	3.0 (5, 2, 2)
The device is easy to install (user friendliness)	2.7 (0, 5, 3)
The device is fun to use (user friendliness)	5.0 (5, 5, 5)
Using it is effortless (user friendliness)	5.3 (5, 5, 6)
I don’t need written instructions (user friendliness)	3.0 (1, 5, 3)
I easily learned to use it (user friendliness)	4.7 (5, 5, 4)
I quickly became an expert in its use (user friendliness)	4.7 (6, 4, 4)
It is easy to use (user friendliness)	4.0 (4, 4, 4)
It is user-friendly (user friendliness)	4.7 (5, 4, 5)
It is suitable for both frequent and infrequent users (user friendliness)	5.0 (5, 5, 5)
I always remember how to use it (user friendliness)	3.7 (6, 2, 3)
It is pleasant to use (user friendliness)	3.3 (5, 2, 3)
I am always able to use the device (reliability)	2.3 (1, 4, 2)
It always works as desired (reliability)	2.7 (3, 1, 4)
It always does exactly what I want (reliability)	1.3 (0, 2, 2)
It perfectly fits my needs	2.0 (1, 0, 5)
I need to have one	2.7 (1, 4, 3)
I will recommend it to a friend	3.0 (1, 3, 5)

[a]Answers range from 0 to 6: 0=strongly disagree, 1=disagree, 2=somewhat disagree, 3=no opinion, 4=somewhat agree, 5=agree, 6=strongly agree. The higher the rating, the better the satisfaction.

End Users’ Propositions

Participants proposed changing the insole’s design, the app, and the synchronization protocol. As a result, improvements were made before the field trial phase. To improve comfort and strength, the insole was modified by thermo-moulding, and the thickness of the electronics was reduced. The circuit was then varnished and encapsulated in epoxy glue and neoprene to protect it from impact and avoid friction. The participant’s interface was also modified to improve text and graphics readability. Upon user request, a light encoder was added on the induction charger that changes from red to green when the charger is rightly positioned in the shoe. Finally, the communication protocol between the tablet and the insole was optimized, and a bug disrupting data transmission (synchronization after several days of use) was fixed.

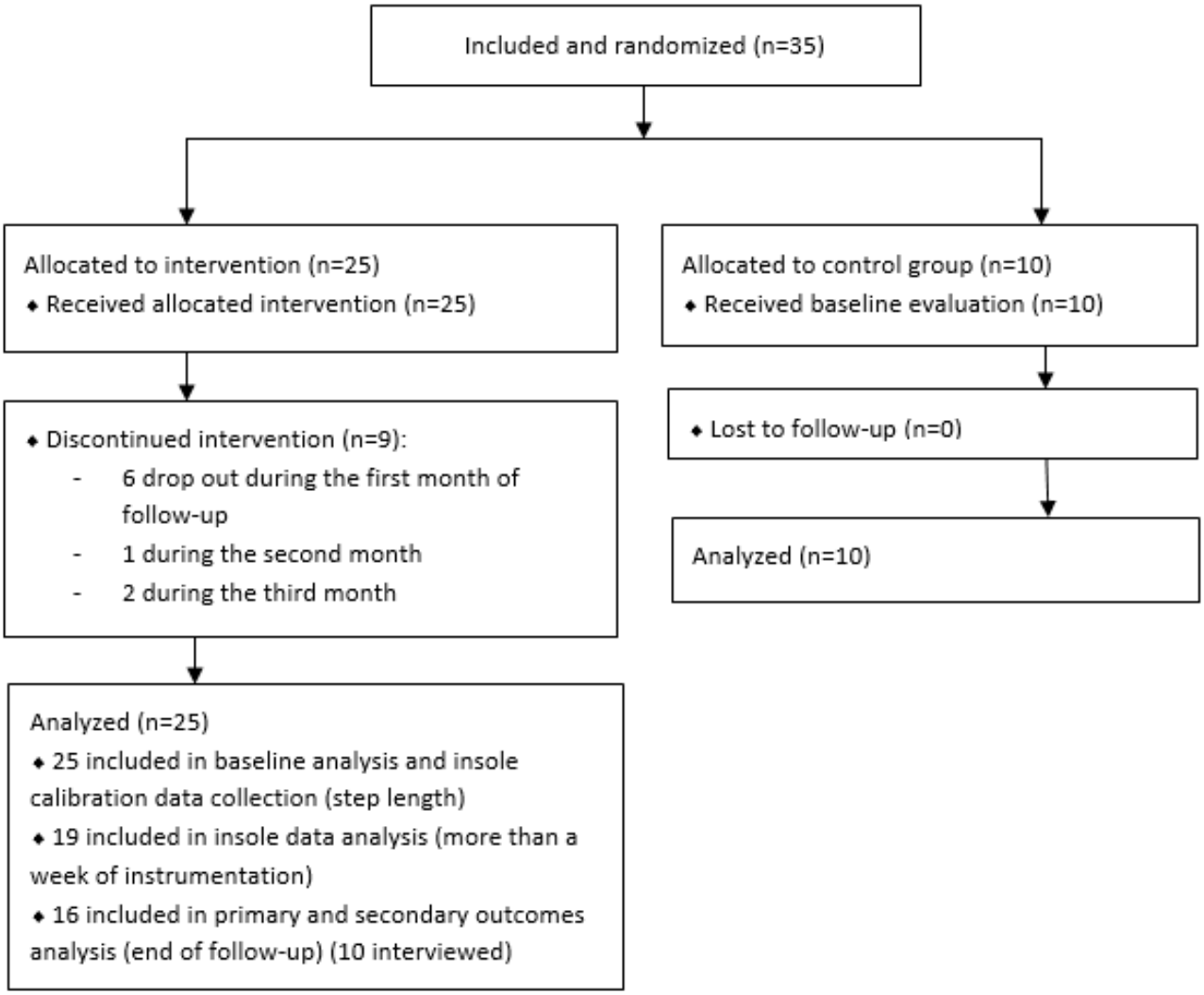
Field Trial

Study Sample Characteristics

A total of 35 participants were included, 10 in the control group and 25 in the intervention group. In the intervention group, 6 participants left the study between visit one and visit two (3 because of defective equipment), 1 participant left the study between visit two and visit three because of defective equipment, and 2 participants left the study between visit three and visit four (1 for defective equipment).

The CONSORT (Consolidated Standards of Reporting Trials) flow diagram is presented in Figure 4 [42].

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



A total of 80% (8/10) of participants were women in the *control* group and 64% (16/25) in the *intervention* group. The average age was 79 (SD 5.8, range 70-89) years. One-third lived in collective housing and two-thirds in individual housing. Participants were quite active; 72% (18/25) reported walking every day in the intervention group compared to 70% (7/10; *P*=.92) in the control one, and less than 10% reported not walking at all in both groups (Table 3).

Table 3. Characteristics of participants (n=35).

Characteristics	Control group (n=10)	Intervention group (n=25)	*P* value[a]
Age (years), mean (SD)	77.8 (5.9)	79.3 (5.9)	.05
Gender (female), n (%)	8 (80)	16 (64)	.44
Education, n (%)			.44
Low level	1 (10)	0 (0)	
Middle level	5 (50)	12 (48)	
High level	4 (40)	13 (52)	
Marital status, n (%)			.23
Married	3 (30)	12 (48)	
Single, divorced, widower	7 (70)	13 (52)	
Living arrangements, n (%)			>.99
Alone (n=9 for control)	6 (67)	15 (60)	
With other (n=9 for control)	3 (33)	10 (40)	
ADL[b] score, mean (SD)	4.8 (1.5)	5.0 (1.4)	.79
IADL[c] score, mean (SD)	7.9 (0.3)	7.9 (0.2)	.59
MMSE[d] score, mean (SD)	27.0 (1.4)	29.1 (1.4)	.09
SPPB[e] score, mean (SD)	10.2 (2.8)	11.2 (1.1)	.42

[a]Fisher test or Wilcoxon rank sum test.

[b]ADL: Activity of Daily Living.

[c]IADL: Instrumental Activities of Daily Living.

[d]MMSE: Mini-Mental State Evaluation.

[e]SPPB: Short Physical Performance Battery.

Acceptability Results (Primary Outcome)

Qualitative Indicators

Semistructured home interviews at the end of the 3-month follow-up (n=10) reported that the insole was well tolerated. A total of 10 (100%) participants declared wearing them every day. Most participants (n=7, 70%) affirmed that they did not need any incentive to wear the insole. Most claimed that the insole was comfortable, light, and robust, and did not cause any discomfort (n=7, 70%; 3 participants found them too thick). A total of 4 participants complained that the insoles did not fit in every type of shoe. The 3 participants who did not walk regularly (walk over short distances on a daily basis), declared that wearing the insole encouraged them to walk because “it stimulated their desire to surpass themselves” and “go out from home.” Concerning the user interface on the touchpad, users expressed some difficulties in handling the devices due to lack of habit (6 participants). As a result, they failed to read their data and did not understand its usefulness. These shortcomings led half of them (5/10) to almost abandon the tablet.

Regarding the tolerance of wearing the insoles during follow-up, 63% (12/19) of participants found it totally tolerable, 37% (7/19) quite tolerable at 1 month, 50% (9/18) and 44% (8/18) at 2 months, and finally increasing to 75% (12/16) and 19% (3/16) at 3 months. Concerning the acceptability questionnaire results at the end of the follow-up (n=16; Table 4), the overall answer was “quite satisfied” or “very satisfied.” The participants were “very satisfied” for weight; between “more or less satisfied” and “very satisfied” for dimensions, ease of adjusting, safety, robustness, and comfort; and less satisfied for ease of use and effectiveness in meeting their needs.

Table 4. Questionnaire results in the intervention group at the end of the follow-up (n=16).

How satisfied are you with (...) your device ?	Not at all, n (%)	Not very, n (%)	More or less, n (%)	Quite, n (%)	Very, n (%)
...the dimensions of...	0 (0)	0 (0)	3 (19)	4 (25)	9 (56)
...the weight of...	0 (0)	0 (0)	2 (12)	0 (0)	14 (88)
...the ease in adjusting...	0 (0)	0 (0)	2 (12)	4 (25)	10 (63)
...the safety of...	0 (0)	0 (0)	4 (25)	2 (12)	10 (63)
...the robustness...	1 (6)	0 (0)	3 (19)	3 (19)	9 (56)
...the ease of use...	1 (6)	4 (25)	1 (6)	4 (25)	6 (38)
...the comfort...	0 (0)	1 (6)	2 (12)	3 (19)	10 (63)
...the effectiveness...	4 (25)	1 (6)	4 (25)	1 (6)	6 (38)

Quantitative Indicators

A total of 25 participants were equipped with the insole, of which 6 for a duration of less than 7 days because of technical problems (Bluetooth communication with the touchpad) at the beginning of the study. Mean step length during calibration was 0.54 (SD 0.16) meters (n=25).

Apart from these 6 participants, 100% (n=19) of participants reported wearing the device every day at 1 month, 83% (15/18) at 2 months, and 94% (15/16) at 3 months of follow-up. Participants reported that the device was worn on average between 5.8 (SD 2.9), 6.3 (SD 6.4), and 5.1 (SD 3.7) hours per day at 1 month, 2 months, and 3 months, respectively. The mean number of days of wearing the insole according to the sensors data in a 3-month period was 29.2 (SD 28.7). If the participants with less than a week of instrumentation (n=6) were excluded, the mean increased to 40.4 (SD 28.8) without significant decline during follow-up (14.2, 12.7, and 13.5, respectively). On average, participants wore the insole for 5.6 (SD 3.7) hours a day. These figures only take into account the days when the insole effectively transmitted data to the server, which excludes connection failure periods or days with insufficient battery charging. The participants connected to the web application on average 45.4 (SD 68.3) times during the follow-up, which corresponds to a mean number of 4.3 (SD 10.6) connections per day.

Secondary Outcomes

Health Outcomes

For health outcomes, there were no statistically significant differences between the two groups at baseline and at the end of follow-up. At baseline, there were 0% (0/10) frail and 70% (7/10) prefrail in the control group, compared to 0% (0/10) and 83% (21/25), respectively, in the intervention group (P=.39). At the end of follow-up, 40% (4/10) frail and 40% (4/10) prefrail were found in the control group, compared to 19% (3/16) and 62% (10/16), respectively, in the intervention group (P=.58). Between visit one and visit four, 10% (1/10) improved their frailty status in the control group versus 19% (3/16) in the intervention group. However, these differences were not significant. Concerning the evolution of the physical activity criterion during follow-up, there was no significant differences despite a trend toward a more sedentary lifestyle in both groups. The overall EQ-5D score, SF-36 score, and functional scores (ADL, IADL, and SPPB) did not show any significant difference between the two groups.

Two notable adverse events not attributable to the intervention were reported, one in each group: 1 participant in the control group had a fall and 1 participant in the intervention group had a fracture.

Health Costs, Installation, and Maintenance Costs

Intend to treat analysis results showed a trend in favor of the intervention group in terms of costs (Table 5).

Table 5. Descriptive cost data.

Cost data	Control	Intervention	*P* value
Medical visits total costs (€[a])	2001.00	1051.00	N/A
Hospitalization total cost (€)	15,374.60	6751.10	N/A
Total costs (€)	17,375.60	7802.10	N/A
Medical visits, n	87	45	N/A
Hospitalization stays, n	11	7	N/A
Medical visits cost (€), mean (SD)	111.20 (124.8)	27.70 (120.9)	.03
Hospitalization cost (€), mean (SD)	854.10 (1245.20)	177.70 (1202.80)	.21
Total cost (€), mean (SD)	965.30 (1329.90)	205,32 (1284.90)	.049

[a]A currency exchange rate of €1=US $1.2 is applicable.

A commercial company (SADIR Assistance) set up home installation and calibration protocol. They spent on average 33.9 (SD 13.1) minutes per visit per person during the study, for an average number of 3.8 (SD 0.6) visits per person (including the installation visit). The mean cost for installation and maintenance communicated by the company was €47.7 (SD €9.7) per month per person (a currency exchange rate of €1=US $1.2 is applicable).

Discussion

Principal Results and Limitations

Our 12-week field trial is the first to assess participant's acceptability of an instrumented insole over 12 weeks in frail older adults in a real-life setting. Difficulty of physical activity follow-up in older individuals and insufficient in-person measures and self-reported data limit clinical research in this domain [9,22]. Currently, there is limited data on the use of instrumented insoles beyond a few hours of laboratory testing.

This study confirmed the importance of a multiphase design for health technologies. The living lab participants valued the study's participative aspect and proposed modifications concerning both the software and the hardware, including the manufacturing method of the insole. Our living lab tests introduced an *optimized reality* between technical lab tests [26] and an ongoing *real-life* field trial. User feedback provided technical and acceptability issues, which were fixed before the field trial (low reliability scores, mainly associated with synchronization problems between the insole and the touchpad). However, despite these precautions, certain technical problems affected the field trial.

Participants validated the device's design: induction charger, charging routine, battery autonomy, and data transfer automation. The insole calibration phase raised no user concern. These are important results as the device was specifically designed to ensure its unobtrusiveness. There were several reasons to choose an insole: it is unobtrusive and can be worn without disturbing or stigmatizing the person, several studies showed that inertial sensors worn on the feet allow accurate walking speed measurement [27,43,44], and the user does not need to remember wearing it.

Semistructured interviews and questionnaires at the end of follow-up reported that the insole was actually worn, unobtrusive, and well tolerated. Moreover, participants were compliant during the 3-month follow-up; this was confirmed by objective data measured by the insole. Most participants claimed that the insole was comfortable and did not cause any discomfort. This result was innovative, as it is one of the first to describe insole long-term wearing acceptability in older adults who are frail. Thus, a thin instrumentation including a flexible battery was designed. This technical challenge remains and must be considered when instrumenting an insole because it is one of the main participant complaints. Acceptability of wearable devices is the cornerstone of large implementation in real-life settings. Finally, those who did not walk regularly also expressed the fact that wearing a smart insole could encourage walking, which remains to be proven.

We found no significant differences between groups in terms of physical activity or frailty evolution. This could be explained by the low study power and by the fact that our touchpad did not offer any incentive or educational content. Information and communication technology–supported lifestyle programs and motion sensing–based monitoring can influence daily physical activity thanks to feedback [45]. However, adding educational content is more efficient [46,47].

This study included a small sample of volunteer older participants. Even if a few exclusion criteria were applied, there was probably a recruitment bias due to highly motivated participants. Moreover, a large proportion of participants in the intervention group discontinued the intervention because of technical issues.

The results obtained by this semiqualitative approach are not free of the usual biases, in particular external validity. There are other limitations to the generalizability of our results. Certain participants brought up the difficulty of integrating an insole in their shoes (eg, sandals), which could be even more problematic in countries where people walk barefoot indoor or wear outdoor shoes.

Moreover, the use of the touchpad interface was not satisfying because of lack of technical support for using it and participant's computer literacy, web interface ergonomic issues, and the absence of educational content (eg, video tutorial). These shortcomings led most of the participants not to use the touchpad. The motivational aspect of our device was poor despite the potential interest this could have [48]. In a second development phase, an educational and motivational network should be developed to increase user's adherence. It seems important to also include participant's computer literacy when developing health devices in this population to ensure better acceptability.

Previous work has highlighted the difficulty of implementing such technical devices in health care practice due to the importance of material and human investment [49]. The results of this study give a glimpse into health device development and improvement, but further studies are required in a larger population to improve large-scale implementation.

Comparison With Prior Work

Previous studies have shown the possibility to monitor mobility-related activities based on motion sensors, but few explicitly mentioned acceptability issues or used experimental research designs to evaluate clinical applications in older people [16,22,50,51]. Most systems consisting of multiple sensors or devices are difficult when applied in long-term monitoring in real-life (eg, the DynaPort MoveMonitor weighs 44.5 g and is fixed with an elastic belt [15]). A few studies evaluated accelerometer-based devices on short treadmill walks [52] on a short time period at home, ranging from a few minutes to days [15,53-56] with up to 20 older participants with unknown frailty status.

Most of these studies explore various aspects of walking such as depth posture and activity detection. Nevertheless, none of these studies evaluated long-term acceptability of such wearable

devices in monitoring walking activity in an older participant sample.

Kaye and colleagues [57] conducted the most notable study. They evaluated in-home walking monitoring using embedded sensors in autonomous living older individuals during a 4-week period. The study did not specifically target frailty status monitoring and was limited to indoor activity. They were able to accurately monitor mean walking speed and even predict falls [9]. Embedded and wearable tools to monitor walking activity provide complementary information. Outdoor walking speed analysis has advantages on indoor measurements because walking distance may be longer and informs global aerobic physical activity [22,58].

Consumer-grade activity trackers (connected devices or smartphone apps) are increasingly popular. However, there is a lack of real-life research on the device's performance in older adults. Given the inaccuracy of these applications, caution is required in promoting self-monitoring physical activity and their use for health prediction [19,23,24]. Indeed, the absence of step length calibration does not allow accurate measures of walking speed and walking characterization. Another major concern is their lack of acceptability in frail older adults. Most of them require minimal computer literacy, and users have to remember to wear it every day. Moreover, there is many constraints related to obtrusiveness (eg, device charging and data transfer). Lastly, the algorithms used to measure steps and other metrics are typically proprietary and may not be available to investigators [59].

Conclusion and Perspectives

Wearable connected sensors are promising for real-life monitoring and appear to be a solution in improving physical activity promotion in frail older adults. However, optimal deployment of wearable health devices will require further research conducted in real-life conditions to test acceptability, effectiveness, and costs. This study reports real-life acceptability data on an instrumented insole in frail older participants over a 12-week period. These results are informative in terms of technical choices for those who wish to instrument a shoe insole.

This field of research is essential and offers interesting perspectives. Along physical activity promotion, these tools would improve detection of early preclinical health transitions implicated in decreased physical performance [9,57,60,61] (eg, gait speed variability over time). Thus, continuous measurements would also enable identification of innovative "digital biomarkers" as a complementary solution to "traditional" biomarkers, leading to more personalized interventions.

Acknowledgments

This study was funded by the French National Research Agency and National Health and Autonomy Funding Agency through the TECSAN program (Project Number ANR-13-TECS-0007-2013-RESPECT). The funder does not have any roles in the design of this study, execution of the trial, analysis and interpretation of the trial data, and research publications.

Conflicts of Interest

None declared.

References

1. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
2. Metzelthin SF, van Rossum E, de Witte LP, Ambergen AW, Hobma SO, Sipers W, et al. Effectiveness of interdisciplinary primary care approach to reduce disability in community dwelling frail older people: cluster randomised controlled trial. BMJ 2013 Sep 10;347:f5264 [FREE Full text] [doi: 10.1136/bmj.f5264] [Medline: 24022033]
3. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, LIFE study investigators. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. JAMA 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]
4. Dedeyne L, Deschodt M, Verschueren S, Tournoy J, Gielen E. Effects of multi-domain interventions in (pre)frail elderly on frailty, functional, and cognitive status: a systematic review. Clin Interv Aging 2017;12:873-896. [doi: 10.2147/CIA.S130794] [Medline: 28579766]
5. Beard JR, Officer A, de Carvalho IA, Sadana R, Pot AM, Michel J, et al. The World report on ageing and health: a policy framework for healthy ageing. Lancet 2016 May 21;387(10033):2145-2154 [FREE Full text] [doi: 10.1016/S0140-6736(15)00516-4] [Medline: 26520231]
6. Cesari M, Araujo de Carvalho I, Amuthavalli Thiyagarajan J, Cooper C, Martin FC, Reginster J, et al. Evidence for the domains supporting the construct of intrinsic capacity. J Gerontol A Biol Sci Med Sci 2018 Nov 10;73(12):1653-1660. [doi: 10.1093/gerona/gly011] [Medline: 29408961]
7. Cesari M, Kritchevsky SB, Penninx BWHJ, Nicklas BJ, Simonsick EM, Newman AB, et al. Prognostic value of usual gait speed in well-functioning older people--results from the Health, Aging and Body Composition Study. J Am Geriatr Soc 2005 Oct;53(10):1675-1680. [doi: 10.1111/j.1532-5415.2005.53501.x] [Medline: 16181165]

8. Purser JL, Kuchibhatla MN, Fillenbaum GG, Harding T, Peterson ED, Alexander KP. Identifying frailty in hospitalized older adults with significant coronary artery disease. J Am Geriatr Soc 2006 Nov;54(11):1674-1681. [doi: 10.1111/j.1532-5415.2006.00914.x] [Medline: 17087693]
9. Piau A, Mattek N, Crissey R, Beattie Z, Dodge H, Kaye J. When will my patient fall? Sensor-based in-home walking speed identifies future falls in older adults. J Gerontol A Biol Sci Med Sci 2020 Apr 17;75(5):968-973 [FREE Full text] [doi: 10.1093/gerona/glz128] [Medline: 31095283]
10. Wild KV, Mattek N, Austin D, Kaye JA. "Are You Sure?": lapses in self-reported activities among healthy older adults reporting online. J Appl Gerontol 2016 Jun;35(6):627-641 [FREE Full text] [doi: 10.1177/0733464815570667] [Medline: 25669877]
11. Liu S, Eaton CB, Driban JB, McAlindon TE, Lapane KL. Comparison of self-report and objective measures of physical activity in US adults with osteoarthritis. Rheumatol Int 2016 Oct;36(10):1355-1364. [doi: 10.1007/s00296-016-3537-9] [Medline: 27435920]
12. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. Front Aging Neurosci 2015;7:102. [doi: 10.3389/fnagi.2015.00102] [Medline: 26113819]
13. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66 Suppl 1:i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]
14. Seelye A, Mattek N, Sharma N, Riley T, Austin J, Wild K, et al. Weekly observations of online survey metadata obtained through home computer use allow for detection of changes in everyday cognition before transition to mild cognitive impairment. Alzheimers Dement 2018 Feb;14(2):187-194 [FREE Full text] [doi: 10.1016/j.jalz.2017.07.756] [Medline: 29107052]
15. Taylor LM, Klenk J, Maney AJ, Kerse N, Macdonald BM, Maddison R. Validation of a body-worn accelerometer to measure activity patterns in octogenarians. Arch Phys Med Rehabil 2014 May;95(5):930-934. [doi: 10.1016/j.apmr.2014.01.013] [Medline: 24486241]
16. de Bruin ED, Hartmann A, Uebelhart D, Murer K, Zijlstra W. Wearable systems for monitoring mobility-related activities in older people: a systematic review. Clin Rehabil 2008;22(10-11):878-895. [doi: 10.1177/0269215508090675] [Medline: 18955420]
17. Kramer J, Kowatsch T. Using feedback to promote physical activity: the role of the feedback sign. J Med Internet Res 2017 Jun 02;19(6):e192 [FREE Full text] [doi: 10.2196/jmir.7012] [Medline: 28576757]
18. Geraedts H, Zijlstra A, Bulstra SK, Stevens M, Zijlstra W. Effects of remote feedback in home-based physical activity interventions for older adults: a systematic review. Patient Educ Couns 2013 Apr;91(1):14-24. [doi: 10.1016/j.pec.2012.10.018] [Medline: 23194823]
19. Orr K, Howe HS, Omran J, Smith KA, Palmateer TM, Ma AE, et al. Validity of smartphone pedometer applications. BMC Res Notes 2015 Nov 30;8:733 [FREE Full text] [doi: 10.1186/s13104-015-1705-8] [Medline: 26621351]
20. Toosizadeh N, Joseph B, Heusser MR, Orouji Jokar T, Mohler J, Phelan HA, et al. Assessing upper-extremity motion: an innovative, objective method to identify frailty in older bed-bound trauma patients. J Am Coll Surg 2016 Aug;223(2):240-248 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2016.03.030] [Medline: 27155751]
21. Piau A, Wild K. Performance of e-health devices for frailty evaluation in real-life settings is far from being demonstrated. Gerontology 2019;65(3):309-310 [FREE Full text] [doi: 10.1159/000495208] [Medline: 30783068]
22. Schwenk M, Howe C, Saleh A, Mohler J, Grewal G, Armstrong D, et al. Frailty and technology: a systematic review of gait analysis in those with frailty. Gerontology 2014;60(1):79-89 [FREE Full text] [doi: 10.1159/000354211] [Medline: 23949441]
23. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]
24. Block VAJ, Pitsch E, Tahir P, Cree BAC, Allen DD, Gelfand JM. Remote physical activity monitoring in neurological disease: a systematic review. PLoS One 2016;11(4):e0154335 [FREE Full text] [doi: 10.1371/journal.pone.0154335] [Medline: 27124611]
25. Ngueleu AM, Blanchette AK, Maltais D, Moffet H, McFadyen BJ, Bouyer L, et al. Validity of instrumented insoles for step counting, posture and activity recognition: a systematic review. Sensors (Basel) 2019 May 28;19(11):2438 [FREE Full text] [doi: 10.3390/s19112438] [Medline: 31141973]
26. Piau A, Charlon Y, Campo E, Vellas B, Nourhashemi F. A smart insole to promote healthy aging for frail elderly individuals: specifications, design, and preliminary results. JMIR Rehabil Assist Technol 2015 May 25;2(1):e5 [FREE Full text] [doi: 10.2196/rehab.4084] [Medline: 28582238]
27. Mariani B, Hoskovec C, Rochat S, Büla C, Penders J, Aminian K. 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. J Biomech 2010 Nov 16;43(15):2999-3006. [doi: 10.1016/j.jbiomech.2010.07.003] [Medline: 20656291]

28. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 03;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
29. Sorenson C, Drummond M, Borlum KF, Busse R. How can the impact of health technology assessments be enhanced? Policy Brief – WHO European Ministerial Conference on Health Systems. 2008 Feb Presented at: Health Systems and Policy Analysis; February 2008; Geneva.
30. Chilcott J, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Paisley S, et al. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. Health Technol Assess 2010 May;14(25):iii-iv, ix. [doi: 10.3310/hta14250] [Medline: 20501062]
31. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010 Jul 16;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
32. Sparling PB, Howard BJ, Dunstan DW, Owen N. Recommendations for physical activity in older adults. BMJ 2015 Jan 21;350:h100. [doi: 10.1136/bmj.h100] [Medline: 25608694]
33. Katz S, Ford A, Moskiwitz R. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963 Sep 21;185:914-919. [Medline: 14044222]
34. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
35. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
36. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 1995 Mar 02;332(9):556-561. [doi: 10.1056/NEJM199503023320902] [Medline: 7838189]
37. Demers L, Monette M, Lapierre Y, Arnold DL, Wolfson C. Reliability, validity, and applicability of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) for adults with multiple sclerosis. Disabil Rehabil 2002;24(1-3):21-30. [doi: 10.1080/09638280110066352] [Medline: 11827151]
38. Ware JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998 Nov;51(11):903-912. [doi: 10.1016/s0895-4356(98)00081-x] [Medline: 9817107]
39. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108. [doi: 10.1097/00005650-199711000-00002] [Medline: 9366889]
40. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
41. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j..2002.384.doc.x] [Medline: 15189396]
42. 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]
43. Martin Schepers H, van Asseldonk EHF, Baten CTM, Veltink PH. Ambulatory estimation of foot placement during walking using inertial sensors. J Biomech 2010 Dec 01;43(16):3138-3143. [doi: 10.1016/j.jbiomech.2010.07.039] [Medline: 20723901]
44. Tien I, Glaser SD, Bajcsy R, Goodin DS, Aminoff MJ. Results of using a wireless inertial measuring system to quantify gait motions in control subjects. IEEE Trans Inf Technol Biomed 2010 Jul;14(4):904-915. [doi: 10.1109/TITB.2009.2021650] [Medline: 19423449]
45. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
46. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
47. van het Reve E, Silveira P, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: part 2 of a phase II preclinical exploratory trial. J Med Internet Res 2014 Jun 25;16(6):e159 [FREE Full text] [doi: 10.2196/jmir.3055] [Medline: 24966165]
48. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth brisk walking intervention on increasing physical activity in older people with cognitive frailty: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16596 [FREE Full text] [doi: 10.2196/16596] [Medline: 32735218]
49. Adlam T, Ulrich E, Kent M, Malinzak L. Frailty testing pilot study: pros and pitfalls. J Clin Med Res 2018 Feb;10(2):82-87 [FREE Full text] [doi: 10.14740/jocmr3203w] [Medline: 29317953]

50. Gilgen-Ammann R, Schweizer T, Wyss T. Accuracy of distance recordings in eight positioning-enabled sport watches: instrument validation study. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e17118 [FREE Full text] [doi: 10.2196/17118] [Medline: 32396865]
51. Katzan I, Schuster A, Kinzy T. Physical activity monitoring using a Fitbit device in ischemic stroke patients: prospective cohort feasibility study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e14494 [FREE Full text] [doi: 10.2196/14494] [Medline: 33464213]
52. Huijben B, van Schooten K, van Dieën JH, Pijnappels M. The effect of walking speed on quality of gait in older adults. Gait Posture 2018 Sep;65:112-116. [doi: 10.1016/j.gaitpost.2018.07.004] [Medline: 30558916]
53. Dijkstra B, Kamsma Y, Zijlstra W. Detection of gait and postures using a miniaturised triaxial accelerometer-based system: accuracy in community-dwelling older adults. Age Ageing 2010 Mar;39(2):259-262. [doi: 10.1093/ageing/afp249] [Medline: 20083616]
54. Hollewand AM, Spijkerman AG, Bilo HJG, Kleefstra N, Kamsma Y, van Hateren KJJ. Validity of an accelerometer-based activity monitor system for measuring physical activity in frail older adults. J Aging Phys Act 2016 Oct;24(4):555-558. [doi: 10.1123/japa.2014-0290] [Medline: 26964560]
55. Culhane KM, Lyons GM, Hilton D, Grace PA, Lyons D. Long-term mobility monitoring of older adults using accelerometers in a clinical environment. Clin Rehabil 2004 May;18(3):335-343. [doi: 10.1191/0269215504cr734oa] [Medline: 15137565]
56. Kim B, McKay SM, Lee J. Consumer-grade wearable device for predicting frailty in Canadian home care service clients: prospective Observational proof-of-concept study. J Med Internet Res 2020 Sep 03;22(9):e19732 [FREE Full text] [doi: 10.2196/19732] [Medline: 32880582]
57. Kaye J, Mattek N, Dodge H, Buracchio T, Austin D, Hagler S, et al. One walk a year to 1000 within a year: continuous in-home unobtrusive gait assessment of older adults. Gait Posture 2012 Feb;35(2):197-202 [FREE Full text] [doi: 10.1016/j.gaitpost.2011.09.006] [Medline: 22047773]
58. Najafi B, Khan T, Wrobel J. Laboratory in a box: wearable sensors and its advantages for gait analysis. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:6507-6510. [doi: 10.1109/IEMBS.2011.6091605] [Medline: 22255829]
59. Seo M, Shin M, Park TS, Park J. Clinometric gait analysis using smart insoles in patients with hemiplegia after stroke: pilot study. JMIR Mhealth Uhealth 2020 Sep 10;8(9):e22208 [FREE Full text] [doi: 10.2196/22208] [Medline: 32909949]
60. Dodge HH, Zhu J, Harvey D, Saito N, Silbert LC, Kaye JA, Alzheimer's Disease Neuroimaging Initiative. Biomarker progressions explain higher variability in stage-specific cognitive decline than baseline values in Alzheimer disease. Alzheimers Dement 2014 Nov;10(6):690-703 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.513] [Medline: 25022534]
61. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]

Abbreviations

ADL: Activity of Daily Living
CONSORT: Consolidated Standards of Reporting Trials
GHF: geriatric day hospital for frailty
IADL: Instrumental Activities of Daily Living
MMSE: Mini-Mental State Examination
SF-36: 36-Item Short Form Health Survey
SPPB: Short Physical Performance Battery

Edited by L Buis; submitted 25.07.19; peer-reviewed by M Yuhas, A Gros, P Robert, H Verloo, E Sadeghi-Demneh, D López López; comments to author 06.09.20; revised version received 02.02.21; accepted 14.04.21; published 05.07.21.

Please cite as:
Piau A, Steinmeyer Z, Charlon Y, Courbet L, Rialle V, Lepage B, Campo E, Nourhashemi F
A Smart Shoe Insole to Monitor Frail Older Adults' Walking Speed: Results of Two Evaluation Phases Completed in a Living Lab and Through a 12-Week Pilot Study
JMIR Mhealth Uhealth 2021;9(7):e15641
URL: https://mhealth.jmir.org/2021/7/e15641
doi:10.2196/15641
PMID:36260404

Original Paper

The Utility of Real-Time Remote Auscultation Using a Bluetooth-Connected Electronic Stethoscope: Open-Label Randomized Controlled Pilot Trial

Takanobu Hirosawa[1*], MD; Yukinori Harada[1*], MD, PhD; Kohei Ikenoya[2*], MD; Shintaro Kakimoto[1*], MD; Yuki Aizawa[1*], MD; Taro Shimizu[1], MD, MBA, MPH, PhD

[1]Department of Diagnostic and Generalist Medicine, Dokkyo Medical University, Tochigi, Japan

[2]Docodemo Clinic Mashiko, Tochigi, Japan

[*]these authors contributed equally

Corresponding Author:
Taro Shimizu, MD, MBA, MPH, PhD
Department of Diagnostic and Generalist Medicine
Dokkyo Medical University
880 Kitakobayashi, Mibu-cho, Shimotsuga
Tochigi, 321-0293
Japan
Phone: 81 282 87 2498
Email: shimizutaro7@gmail.com

Abstract

Background: The urgent need for telemedicine has become clear in the COVID-19 pandemic. To facilitate telemedicine, the development and improvement of remote examination systems are required. A system combining an electronic stethoscope and Bluetooth connectivity is a promising option for remote auscultation in clinics and hospitals. However, the utility of such systems remains unknown.

Objective: This study was conducted to assess the utility of real-time auscultation using a Bluetooth-connected electronic stethoscope compared to that of classical auscultation, using lung and cardiology patient simulators.

Methods: This was an open-label, randomized controlled trial including senior residents and faculty in the department of general internal medicine of a university hospital. The only exclusion criterion was a refusal to participate. This study consisted of 2 parts: lung auscultation and cardiac auscultation. Each part contained a tutorial session and a test session. All participants attended a tutorial session, in which they listened to 15 sounds on the simulator using a classic stethoscope and were told the correct classification. Thereafter, participants were randomly assigned to either the real-time remote auscultation group (intervention group) or the classical auscultation group (control group) for test sessions. In the test sessions, participants had to classify a series of 10 lung sounds and 10 cardiac sounds, depending on the study part. The intervention group listened to the sounds remotely using the electronic stethoscope, a Bluetooth transmitter, and a wireless, noise-canceling, stereo headset. The control group listened to the sounds directly using a traditional stethoscope. The primary outcome was the test score, and the secondary outcomes were the rates of correct answers for each sound.

Results: In total, 20 participants were included. There were no differences in age, sex, and years from graduation between the 2 groups in each part. The overall test score of lung auscultation in the intervention group (80/110, 72.7%) was not different from that in the control group (71/90, 78.9%; *P*=.32). The only lung sound for which the correct answer rate differed between groups was that of pleural friction rubs (*P*=.03); it was lower in the intervention group (3/11, 27%) than in the control group (7/9, 78%). The overall test score for cardiac auscultation in the intervention group (50/60, 83.3%) was not different from that in the control group (119/140, 85.0%; *P*=.77). There was no cardiac sound for which the correct answer rate differed between groups.

Conclusions: The utility of a real-time remote auscultation system using a Bluetooth-connected electronic stethoscope was comparable to that of direct auscultation using a classic stethoscope, except for classification of pleural friction rubs. This means that most of the real world's essential cardiopulmonary sounds could be classified by a real-time remote auscultation system using a Bluetooth-connected electronic stethoscope.

Trial Registration: UMIN-CTR UMIN000040828; https://tinyurl.com/r24j2p6s and UMIN-CTR UMIN000041601; https://tinyurl.com/bsax3j5f

(JMIR Mhealth Uhealth 2021;9(7):e23109) doi:10.2196/23109

KEYWORDS

telemedicine; electronic stethoscope; simulator; remote auscultation; lung auscultation; cardiac auscultation; physical examination

Introduction

Since the French physician René Laennec invented the stethoscope in 1816 [1], auscultation has been an essential component of clinical examination [2]. Auscultation is not only a highly cost-effective screening tool to detect abnormal clinical signs but also a useful means to build a good relationship between physician and patient. Although there has been a concerning decline in physicians' auscultatory skills [2], auscultation will remain important in the 2020s, as cardiopulmonary diseases are important underlying or direct causes of death and morbidity, substantially impacting quality of life and health care costs [3,4].

However, auscultation became challenging during the COVID-19 pandemic. As medical staff also need to be protected from infection during an outbreak, the need for telemedicine is growing rapidly worldwide [5]. Therefore, remote auscultation should be developed for patients with acute or chronic cardiopulmonary diseases. However, several barriers exist for implementing telemedicine in clinical practice [6], including a lack of effectiveness via remote examination that can act as a substitute for direct physical examination. Auscultation is not one of the procedures typically mentioned in discussions of and research on telemedicine [2], which makes telemedicine seem impractical for patients with cardiopulmonary diseases.

Electronic stethoscopes are promising options to solve the problem of the lack of remote auscultation systems [7]. The electronic stethoscope not only is useful for remote auscultation but also has other advantages over the classic stethoscope. It can facilitate the differentiation of several types of cardiac and lung sounds by visualization of their sonograms during auscultation [2]; it can improve sound quality [8]; and it can contribute to better performance on auscultation, as personalized adjustments can be made [9]. Therefore, the electronic stethoscope can be used in telemedicine without a reduction in the quality of auscultation.

To the best of our knowledge, only a few studies have been conducted to determine the utility of real-time remote auscultation using an electronic stethoscope [7]. However, we are not aware of studies conducted for the direct comparison between real-time remote auscultation and direct auscultation. Therefore, in this study, we tested the hypothesis that the utility of real-time remote auscultation by a physician would be comparable to that of classical auscultation by a physician.

Methods

Study Design, Setting, and Participants

An open-label randomized controlled trial was designed to assess the utility of real-time remote auscultation using a Bluetooth-connected electronic stethoscope. Direct auscultation using a traditional stethoscope was used as control. To standardize and enhance the reliability of the assessment, we used a lung simulator for lung auscultation and a cardiology patient simulator for cardiac auscultation [10]. All sessions in this study were conducted at the skills laboratory at Dokkyo Medical University. As study participants, we recruited senior residents and faculty in the Department of Diagnostic and Generalist Medicine, as they were considered representative of general physicians working in community hospitals or clinics, the main population expected to perform auscultation in routine clinical practice. The only exclusion criterion was refusal to participate in this study. At baseline, no physicians exhibited hearing loss at their previous annual health checkup. This study was conducted in accordance with the current version of the Declaration of Helsinki. The study protocols were approved by the institutional ethics committee of Dokkyo Medical University, Tochigi, Japan (No. R-33-20J and R-37-19J). Written informed consent was obtained from all participants after a detailed explanation of the study and before participation.

Procedures

Study Flow and Randomization

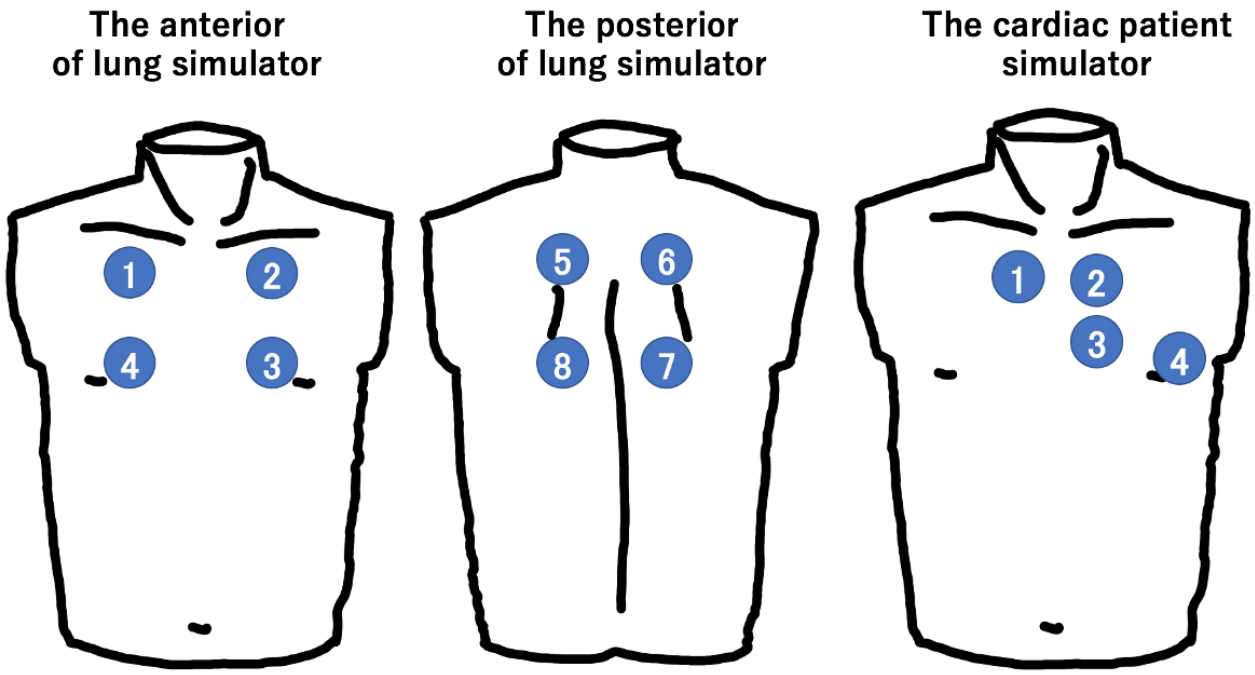
This study consisted of a lung auscultation part and a cardiac auscultation part. Each part contained a tutorial and a test session. Prior to the test session, all participants attended a tutorial session to become familiar with the device. Thereafter, participants were randomly assigned (simple randomization) to either the real-time remote auscultation group (intervention group) or the classical auscultation group (control group). The randomization was conducted separately in the lung and cardiac parts. Researchers were blinded in terms of allocation, by using a computer-generated allocation table to assign participants.

Tutorial Session

In the tutorial sessions, all participants took part in auscultation using a traditional stethoscope (Littmann Cardiology III, 3M, St Paul, MN) on the lung simulator (Lung Sound Auscultation Trainer "LSAT" ver.2, model #MW28, Kyoto Kagaku Co, Ltd, Kyoto, Japan) and on the cardiac patient simulator (Cardiology Patient Simulator "K" ver.2, model #MW10, Kyoto Kagaku Co, Ltd). In each tutorial session, the participant listened to 15 sounds, with the correct classification being provided to participants. A short instruction for the simulator and the correct placements for auscultation were provided. In the tutorial session

for lung auscultation, the following 15 sounds were played in a random order: normal lung sounds (standard, loud), wheezes (350-450 Hz, 600-700 Hz, 200-1000 Hz), 2 different rhonchi, stridor (twice), 2 different coarse crackles, 2 different fine crackles, and 2 different pleural friction rubs. In the tutorial session for cardiac auscultation, the following 15 sounds were played in random order: 3 different normal cardiac sounds (no S_2 split, S_2 split, and S_1 split), 3 different third cardiac sounds (S_3 gallop enhanced, S_3 gallop, and S_3-S_4 gallop), aortic stenosis (twice), aortic regurgitation (twice), mitral regurgitation (3 times), mitral stenosis, and atrial fibrillation. Each participant was instructed to auscultate in the standardized positions on the simulators: 4 on the anterior and 4 on the posterior on the lung simulator; 4 on the cardiology patient simulator (Figure 1). Each sound was played for a maximum of 1 minute.

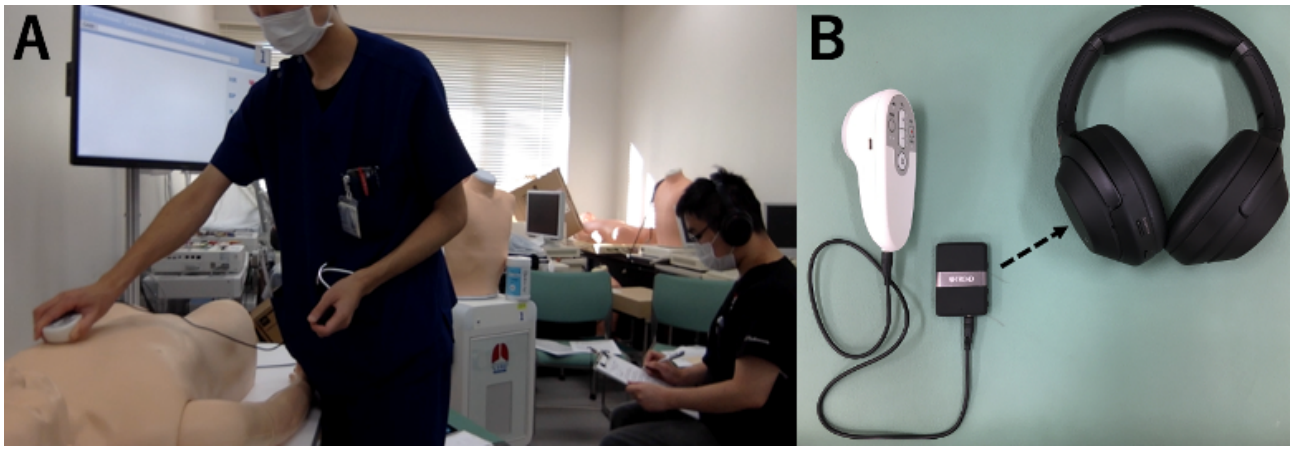
Figure 1. The 8 different areas of auscultation on the lung simulator and the 4 different areas of auscultation on the cardiac patient simulator.



Test Session

In the test sessions, participants in the intervention group auscultated all sounds remotely using an electronic stethoscope (JPES-01, MEMS CORE Co, Ltd, Miyagi, Japan), a Bluetooth transmitter and receiver (BT-DUO, TROND, Eastvale, CA), and a wireless, noise-canceling, stereo headset (WH-1000XM3, Sony Co, Tokyo, Japan), as depicted in Figures 2A and 2B. For this group, the researcher placed the electronic stethoscope and transmitter on the simulator, and participants could monitor the placement of the electronic stethoscope in real time. In the lung simulator, the LED panel on the side of the simulator indicated whether it was the inspiration or expiration phase. The monitoring screen of the cardiology patient simulator was modified to display only a heartbeat icon. Participants in the control group auscultated all sounds directly using the traditional stethoscope, placing it by themselves. For lung auscultation, of the 15 sounds played during the tutorial session, the following 10 sounds were played in a random order during the test session: normal lung sounds (normal, loud), wheezes (350-450 Hz, 600-700 Hz, 200-1000 Hz), rhonchi, stridor, coarse and fine crackles, and pleural friction rubs. For cardiac auscultation, of the 15 sounds played during the tutorial session, the following 10 sounds were played in random order during the cardiac test session: 2 different normal cardiac sounds (no S_2 split, S_2 split), 2 different S_3 (S_3 gallop enhanced, S_3 gallop), aortic stenosis, aortic regurgitation, mitral regurgitation (twice), mitral stenosis, and atrial fibrillation. Auscultation positions and length were also the same as those in the tutorial session. During the test session, all participants filled in forms to indicate the types of sounds they recognized (Multimedia Appendix 1 and Multimedia Appendix 2).

Figure 2. The remote auscultation processes: (A) The researcher is on the left, placing the electronic stethoscope on the cardiac patient simulator, and the participant is on the right listening to the cardiac sounds via a wireless, noise-canceling, stereo headset and Bluetooth transmitter; (B) the remote auscultation equipment including an electronic stethoscope; wireless, noise-canceling, stereo headset; and Bluetooth transmitter (Bluetooth connection is indicated with a dashed double arrow).



Lung Simulator

For lung auscultation, the same lung simulator (MW28) was used in the tutorial and test sessions. This simulator was designed for medical education training and includes 34 samples of lung sounds recorded from actual patients and reproduced using a high-quality sound system. These lung sounds were classified, according to the American Thoracic Society classification system, as continuous (wheezes, rhonchi, or stridor) or discontinuous (fine or coarse crackles) [11]. Fine crackles were defined as "relatively high-pitched sounds usually heard at the end of inspiration as air enters the acinar unit." Coarse crackles were defined as "the low-pitched, bubbling sounds that result from the accumulation of secretions in larger bronchi and trachea" [11]. The default respiratory rate of 15 breaths/minute was used during the study.

Cardiology Patient Simulator

For cardiac auscultation, the same cardiology patient simulator (MW10) was used in the tutorial and test sessions. This simulator was designed for medical education training and includes 88 cases of cardiac sounds recorded from actual patients and reproduced using a high-quality sound system.

Electronic Stethoscope

The electronic stethoscope is equipped with pressure-sensitive sensors, and the signals are converted into sound waves. It is also equipped with a volume regulator and a frequency filter. The filter has a bell mode, diaphragm mode, and wide mode, which enhance the 20-100 Hz, 200-2000 Hz, and 20-2000 Hz frequency bands, respectively. In the lung and cardiac parts, we used the diaphragm mode and the bell mode, respectively. The transmitter transferred the sounds from the lung simulator to the headset via Bluetooth (A2DP: Advanced Audio Distribution Profile).

Data Collection and Outcome Measures

Age, sex, and years since obtaining a degree in medicine were collected from all participants as baseline demographic data. All participants' answers for each sound in the test session were collected. The primary outcome measure was the test score in each group. The rates of correct answers for each sound were the secondary outcome measures.

Statistical Analysis

The correct answer in each group was compared using the Fisher exact test for primary and secondary outcome measures. Continuous variables for participant baseline characteristics are presented as medians (IQRs) and were compared using the Mann-Whitney *U* test. Categorical and binary variables for participant baseline characteristics are presented as numbers (percentages) and were compared using the Fisher exact test. A *P* value <.05 was considered statistically significant. All statistical tests were performed using R 3.6.0 for MacOS X (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Participant Profiles

In total, 20 physicians in the Department of Diagnostic and Generalist Medicine of Dokkyo Medical University were enrolled in the final analysis (Figure 3). Of these, 7 (35%) were senior residents (3-5 years after graduation), and 13 (65%) were faculty (≥6 years after graduation). The median age of all participants was 32 (IQR 8.5) years; 16 (16/20, 80%) were male, and the median time since graduation was 7.5 (IQR 7.5) years. In the lung part, 11 and 9 participants were assigned to the intervention and control groups, respectively. There were no statistically significant differences in participant age (*P*=.25), sex (*P*=.82), and years since graduation (*P*=.15) between the groups (Table 1). In the cardiac part, 6 and 14 participants were assigned to the intervention and control groups, respectively. There were no statistically significant differences in participant age (*P*=.99), sex (*P*=.99), and years since graduation (*P*=.78) between the groups (Table 1).

Figure 3. Flowchart of participant inclusion in the study.

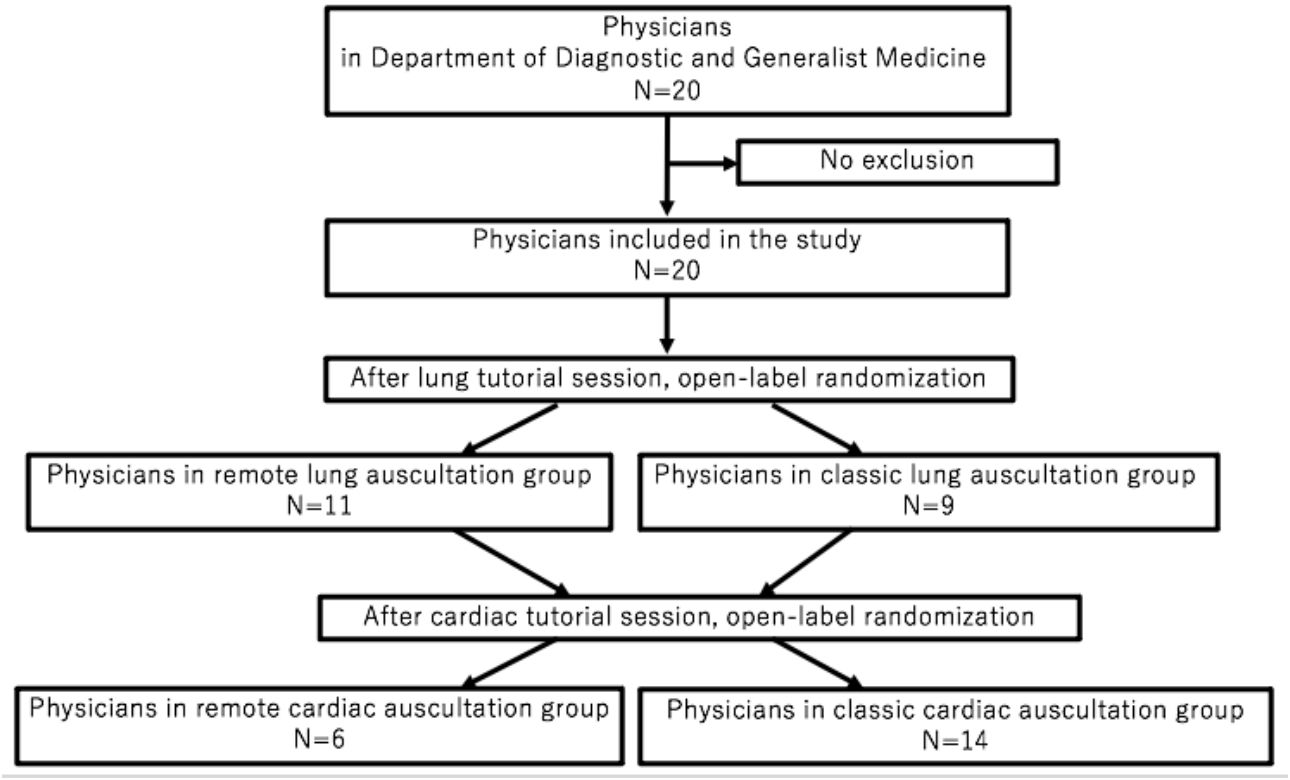


Table 1. Baseline characteristics of participants in the classical and remote cardiopulmonary auscultation groups.

Variable	Lung auscultation			Cardiac auscultation		
	Remote lung auscultation (n=11)	Classical lung auscultation (n=9)	*P* value	Remote cardiac auscultation (n=6)	Classical cardiac auscultation (n=14)	*P* value
Age (years), median (IQR)	34.0 (7.0)	29.0 (9.0)	.25[a]	32.5 (6.3)	32.5 (6.5)	.99[a]
Men, n (%)	9 (82)	7 (78)	.82[b]	6 (100)	11 (79)	.99[b]
Years after graduation (years), median (IQR)	10.0 (6.0)	4.0 (5.0)	.15[a]	8.0 (4.8)	7.4 (5.5)	.78[a]

[a]Mann-Whitney *U* test

[b]Fisher exact test.

Diagnostic Performance

Test scores and rates of correct answers for each lung sound are summarized in Table 2. The total combined test score was 80/110 (72.7%) in the intervention group and 71/90 (78.9%) in the control group, with no differences between the groups (*P*=.32). There were no differences between the groups for normal lung sounds, wheezes, rhonchi, coarse crackles, fine crackles, or stridor. Only 3/11 (27%) participants in the intervention group correctly auscultated pleural friction rubs, whereas 7/9 (78%) in the control group did (*P*=.03).

Details of the answers for lung auscultation are supplied in Table 3 and Table 4. During exploratory data analysis, we determined that 4/22 (18%) instances of normal lung sounds were assigned as pleural friction rubs in the intervention group. On the other hand, 5/11 (45%) instances of pleural friction rubs were assigned as normal lung sounds in the same group.

Table 2. Lung sounds correctly identified.

Variable	Remote auscultation (n=11)	Classical auscultation (n=9)	*P* value[a]
Total, n/N (%)	80/110 (72.7)	71/90 (78.9)	.32
Normal, n/N (%)	16/22 (72.7)	18/18 (100)	.99
Wheezes, n/N (%)	25/33 (75.8)	19/27 (70.4)	.64
Rhonchi, n/N (%)	9/11 (81.8)	5/9 (55.6)	.21
Coarse crackles, n/N (%)	7/11 (63.6)	7/9 (77.8)	.50
Fine crackles, n/N (%)	9/11 (81.8)	6/9 (66.7)	.44
Pleural friction rubs, n/N (%)	3/11 (27.3)	7/9 (77.8)	.03
Stridor, n/N (%)	11/11 (100)	9/9 (100)	N/A[b]

[a]Fisher exact test.

[b]N/A: not applicable.

Table 3. Details of the participants' answers for lung auscultation in the intervention (remote lung auscultation) group.

Correct answer	Normal	Wheezes	Rhonchi	Coarse crackles	Fine crackles	Pleural friction rubs	Stridor
Normal (n=22)	16	0	0	2	0	4	0
Wheezes (n=33)	2	25	3	0	0	1	2
Rhonchi (n=11)	0	1	9	0	0	0	1
Coarse crackles (n=11)	2	0	0	7	0	2	0
Fine crackles (n=11)	0	0	0	2	9	0	0
Pleural friction rubs (n=11)	5	1	0	2	0	3	0
Stridor (n=11)	0	0	0	0	0	0	11

Table 4. Details of the participants' answers for lung auscultation in the control (traditional lung auscultation) group.

Correct answer	Normal	Wheezes	Rhonchi	Coarse crackles	Fine crackles	Pleural friction rubs	Stridor
Normal (n=18)	18	0	0	0	0	0	0
Wheezes (n=27)	0	19	6	0	0	0	2
Rhonchi (n=9)	0	2	5	0	0	0	2
Coarse crackles (n=9)	0	0	0	7	2	0	0
Fine crackles (n=9)	0	0	0	2	6	1	0
Pleural friction rubs (n=9)	1	0	1	0	0	7	0
Stridor (n=9)	0	0	0	0	0	0	9

Test scores and rates of correct answers for each cardiac sound are summarized in Table 5. The total combined test score was 50/60 (83.3%) in the intervention group and 119/140 (85.0%) in the control group, with no differences between the groups (*P*=.77). There were no differences between the groups for normal cardiac sounds, S_3, aortic stenosis, aortic regurgitation, mitral stenosis, mitral regurgitation, and atrial fibrillation. Although not statistically significant, there was over a 30% difference in the score for mitral stenosis between the 2 groups: While 5 of 6 (84%) participants in the intervention group correctly auscultated, only 7 of 14 (50%) participants in the control group did (*P*=.19).

Details of the answers in the remote cardiac auscultation group are provided in Table 6. Misinterpretations between normal cardiac sounds and S_3 sounds were frequently observed in the intervention group: 2 of 12 (17%) instances of normal cardiac

sounds were assigned as S_3 sounds. On the other hand, 4 of 12 (33%) instances of S_3 sounds were assigned as normal cardiac sounds. Compared to the intervention group, participants in the control group made fewer misclassifications between normal cardiac sounds and S_3 sounds (Table 7): Misclassifications of normal cardiac sound as S_3 sounds occurred in 2 of 28 (7%) sounds, and misclassifications of S_3 sounds as normal cardiac sounds occurred in 5 of 28 (17%) sounds. On the other hand, misinterpretations between mitral stenosis sounds and mitral regurgitation sounds were frequently observed in the control group: 6 of 14 (43%) instances of mitral stenosis sounds were assigned as mitral regurgitation sounds. In comparison, 4 of 28 (14%) instances of mitral regurgitation sounds were assigned as mitral stenosis sounds.

Table 5. Cardiac sounds correctly identified.

Variable	Remote auscultation (n=6)	Classical auscultation (n=14)	*P* value[a]
Total, n/N (%)	50/60 (83.8)	119/140 (85.0)	.77
Normal, n/N (%)	9/12 (75.0)	26/28 (92.9)	.14
S_3, n/N (%)	8/12 (66.7)	22/28 (78.6)	.43
Aortic stenosis, n/N (%)	5/6 (83.3)	14/14 (100)	.99
Aortic regurgitation, n/N (%)	6/6 (100)	13/14 (92.9)	.99
Mitral stenosis, n/N (%)	5/6 (83.8)	7/14 (50.0)	.19
Mitral regurgitation, n/N (%)	11/12 (91.7)	23/28 (82.1)	.45
Atrial fibrillation, n/N (%)	6/6 (100)	14/14 (100)	N/A[b]

[a]Fisher exact test.

[b]N/A: not applicable.

Table 6. Details of the participants' answers for cardiac auscultation in the intervention (remote cardiac auscultation) group.

Correct answer	Normal	S_3	Aortic stenosis	Aortic regurgitation	Mitral stenosis	Mitral regurgitation	Atrial fibrillation
Normal (n=12)	9	2	0	0	1	0	0
S_3 (n=12)	4	8	0	0	0	0	0
Aortic stenosis (n=6)	0	0	5	0	0	1	0
Aortic regurgitation (n=6)	0	0	0	6	0	0	0
Mitral stenosis (n=6)	0	0	0	0	5	1	0
Mitral regurgitation (n=12)	0	0	0	0	1	11	0
Atrial fibrillation (n=6)	0	0	0	0	0	0	6

Table 7. Details of the participants' answers for cardiac auscultation in the control (traditional cardiac auscultation) group.

Correct answer	Normal	S_3	Aortic stenosis	Aortic regurgitation	Mitral stenosis	Mitral regurgitation	Atrial fibrillation
Normal (n=28)	26	2	0	0	0	0	0
S_3 (n=28)	5	22	0	0	1/28	0	0
Aortic stenosis (n=14)	0	0	14	0	0	0	0
Aortic regurgitation (n=14)	0	0	0	13	1	0	0
Mitral stenosis (n=14)	0	0	0	1	7	6	0
Mitral regurgitation (n=28)	0	0	1	0	4	23	0
Atrial fibrillation (n=14)	0	0	0	0	0	0	14

Discussion

Principal Findings

In this study, there were 4 main findings. First, using a simulator, we demonstrated that the utility of real-time remote auscultation using a Bluetooth-connected electronic stethoscope was comparable to that of direct auscultation using a classic stethoscope. From previous finding of lung auscultation, coarse crackles, fine crackles, wheezes, and stridor are useful for diagnosing bronchitis or pneumonia [12-14], interstitial pulmonary fibrosis or pneumonia [12,15-19], exacerbation of asthma or chronic obstructive pulmonary disease [12,19], and upper-airway obstruction [19,20], respectively. For cardiac auscultation, valvular cardiac diseases and irregular rhythm disease were not misclassified as normal cardiac sounds in real-time remote auscultation and classical groups. The correct detection of S_3 is also an essential skill in diagnosing various cardiac diseases such as congestive heart failure, ischemic heart disease, cardiomyopathies, myocarditis, cor pulmonale, high-output states, left-to-right intracardiac shunts, and complete atrioventricular block [21]. Therefore, the results of this study suggest that real-time remote auscultation would be useful for the diagnosis of these diseases. At the bedside, this would allow medical staff to screen patients for adventitious sounds at a distance, protecting themselves from infectious diseases such as COVID-19.

Second, the rate of correct answers for pleural friction rubs was lower in the real-time remote lung auscultation group than in the direct lung auscultation group. Therefore, it would be challenging to diagnose pleuritis with real-time remote lung auscultation [12].

Third, in the real-time remote lung auscultation group, we observed a trend for confusion of normal lung sounds and pleural friction rubs. Respecting pleural friction rubs, 45% were misclassified as normal lung sounds in the remote auscultation group in this study. According to the participants, placement of the electronic stethoscope to the surface of the lung simulator caused a bit of noise. Electronic stethoscopes are sensitive to electronic and ambient noise, and this placement noise may be the cause of the difficulty observed in auscultation of pleural friction rub.

Fourth, the rate of correct answers for mitral stenosis was higher in the real-time remote cardiac auscultation than in the direct cardiac auscultation group. According to the participants, the monitoring screen of the simulator was on the caudal side. The screen showed a heartbeat icon regardless of whether it was the systolic or diastolic phase without any waveform. In the direct cardiac auscultation group, it was difficult to watch a display with auscultation. On the other hand, in the real-time remote auscultation group, the participants could watch the screen to detect the systolic or diastolic phase with auscultation. This may have been the cause for the misinterpretations between mitral stenosis sounds and mitral regurgitation sounds in the direct auscultation group.

Strengths

There were 3 strengths of note in this study. First, use of simulators allowed us to gather standardized data, reducing bias. Second, the direct comparison of a novel auscultation technology with classical auscultation as control adds value to this randomized controlled study. Third, all participants were physicians who specialized in general internal medicine. Thus, the study participants were representative of general physicians working in community hospitals or clinics, the main population expected to perform auscultation in routine clinical practice.

Limitations

This study was a pilot study and had several limitations. First, the sample size was small and did not include real patient data. Fully powered trials need to be conducted to better show equivalence. There was grouping variation through simple randomization, especially in the cardiac part, with 6 in the intervention group vs 14 in the control group. This grouping variation may have affected the detection power in this study. Therefore, in future studies, the efficacy of real-time remote auscultation has to be confirmed at the bedside with a larger sample size. Second, as the researcher who placed the electronic stethoscope on the simulators could not hear the sounds while doing so, the timing of the change in auscultation position could not be adjusted for optimum auscultation. Participants in the classical auscultation group were able to change the auscultation position on their own, which may have given this group an advantage over the remote auscultation group. Third, the respiratory rate and phase of crackle could not be adjusted with the lung simulator. The default respiratory rate was slower than that of real patients with respiratory diseases. Therefore, the results from this study may not be generalizable to auscultation at the bedside. Fourth, in this study, the technique used for classical auscultation was not standardized. This may have limited the reproducibility of our results. Fifth, there might be some dependences among answers within-subject.

Comparison With Prior Work

To the best of our knowledge, there has been no other study in which real-time remote auscultation and classical auscultation were directly compared. In terms of the accuracy of classical auscultation using the simulator, our results, which showed variable accuracy depending on the type of sounds, were consistent with previous studies [10,22]. In both previous studies, the total accuracies varied from 62% to 89.7%, which may have depended on the difficulty of tests. For example, one study that showed a low accuracy rate included 3 types of S_2, S_4, and S_3+S_4, whereas another study that showed a high accuracy rate only included S_2 split, mitral regurgitation, and aortic stenosis. Regarding the difficulty of testing, our study may be close to the previous 2 studies.

In a previous study of the accuracy of identifying lung sounds using classical auscultation, stridor was not included in the evaluation [10]. Stridor can be identified without a stethoscope. However, we included stridor in this study, as its detection with electronic stethoscopes was reported in other studies [23].

We are aware of 1 study reporting results regarding the utility of real-time remote auscultation [7]. In that study, the

interobserver concordance of remote auscultation via the internet (93.7%-94.0%) was lower than that of direct auscultation (98.4%-98.9%), but not statistically significant. Although remote auscultation in that study was not conducted in real time, the results are similar to ours in terms of a slightly lower accuracy in the remote auscultation group than in the conventional auscultation group.

Conclusions

This study demonstrates that the utility of a Bluetooth-connected, real-time remote auscultation system is comparable to that of classical, direct auscultation, except for pleural friction rubs. Future studies focused on real-time auscultation through Wi-Fi or the internet are warranted. Furthermore, this study leads the way for further studies in real patients with fully powered trials. In a future study, visualized waves of sounds [24] or artificial intelligence [25] would be supported to detect abnormal sounds.

Acknowledgments

This study was made possible using the resources from the Department of Diagnostic and Generalist Medicine, Dokkyo Medical University. Special thanks to Mr Masahiro Sato, manager of the MEMS CORE, Miyagi, Japan, who advised to adjust the electronic stethoscope. Special thanks to Mr Yukio Hanatani, chief executive officer of SMART GATE, Tokyo, Japan, who advised to adjust the electronic devices. An electronic stethoscope (JPES-01) used in this study was provided by MEMS CORE, Miyagi, Japan.

Authors' Contributions

TH, YH, KI, SK, YA, and TS contributed to the study concept and design. TH and YH performed the statistical analyses. TH contributed to the drafting of the manuscript. YH, KI, and TS contributed to the critical revision of the manuscript for relevant intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The format of the questionnaire for identification of the 10 lung sounds.
[PDF File (Adobe PDF File), 46 KB - mhealth_v9i7e23109_app1.pdf]

Multimedia Appendix 2
The format of the questionnaire for identification of the 10 cardiac sounds.
[PDF File (Adobe PDF File), 45 KB - mhealth_v9i7e23109_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1516 KB - mhealth_v9i7e23109_app3.pdf]

References

1. Morris JS. Laennec's stethoscope--the Welsh connection. J R Soc Med 2004 Mar;97(3):137-141 [FREE Full text] [doi: 10.1258/jrsm.97.3.137] [Medline: 14996966]
2. Pereira D, Castro A, Gomes P, Areias JCNC, Reis ZSN, Coimbra MT, et al. Digital Auscultation: Challenges and Perspectives. In: Cruz-Cunha MM, Miranda IM, Martinho R, Rijo R, editors. Encyclopedia of E-Health and Telemedicine. Hershey, PA: IGI Global; 2016:910-927.
3. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease, Second Edition. World Health Organization. 2017. URL: https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory_Disease.pdf [accessed 2021-07-14]
4. NCD Countdown 2030 collaborators. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. Lancet 2018 Sep 22;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 30264707]
5. Omboni S. Telemedicine During the COVID-19 in Italy: A Missed Opportunity? Telemed J E Health 2020 Aug 01;26(8):973-975 [FREE Full text] [doi: 10.1089/tmj.2020.0106] [Medline: 32324109]
6. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]

7. Marani R, Gelao G, Perri AG. High quality heart and lung auscultation system for diagnostic use on remote patients in real time. Open Biomed Eng J 2010 Nov 03;4(1):250-256 [FREE Full text] [doi: 10.2174/1874120701004010250] [Medline: 21379392]
8. Tavel ME. Cardiac auscultation. A glorious past--but does it have a future? Circulation 1996 Mar 15;93(6):1250-1253. [doi: 10.1161/01.cir.93.6.1250] [Medline: 8653848]
9. Høyte H, Jensen T, Gjesdal K. Cardiac auscultation training of medical students: a comparison of electronic sensor-based and acoustic stethoscopes. BMC Med Educ 2005 May 09;5(1):14 [FREE Full text] [doi: 10.1186/1472-6920-5-14] [Medline: 15882458]
10. Bernardi S, Giudici F, Leone MF, Zuolo G, Furlotti S, Carretta R, et al. A prospective study on the efficacy of patient simulation in heart and lung auscultation. BMC Med Educ 2019 Jul 23;19(1):275 [FREE Full text] [doi: 10.1186/s12909-019-1708-6] [Medline: 31337361]
11. American Thoracic Society. Updated nomenclature for membership reaction- Reports from the ATS AdHoc Committee on Pulmonary. Am Thoracic Soc News 1977;3:5-6.
12. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. N Engl J Med 2014 Feb 20;370(8):744-751. [doi: 10.1056/NEJMra1302901] [Medline: 24552321]
13. Piirilä P. Changes in crackle characteristics during the clinical course of pneumonia. Chest 1992 Jul;102(1):176-183. [doi: 10.1378/chest.102.1.176] [Medline: 1623749]
14. Nakanishi M, Yoshida Y, Takeda N, Hirana H, Horita T, Shimizu K, et al. Significance of the progression of respiratory symptoms for predicting community-acquired pneumonia in general practice. Respirology 2010 Aug;15(6):969-974. [doi: 10.1111/j.1440-1843.2010.01807.x] [Medline: 20630031]
15. Cottin V, Cordier J. Velcro crackles: the key for early diagnosis of idiopathic pulmonary fibrosis? Eur Respir J 2012 Sep 31;40(3):519-521 [FREE Full text] [doi: 10.1183/09031936.00001612] [Medline: 22941541]
16. Melbye H, Straume B, Aasebø U, Brox J. The diagnosis of adult pneumonia in general practice. The diagnostic value of history, physical examination and some blood tests. Scand J Prim Health Care 1988 May 12;6(2):111-117. [doi: 10.3109/02813438809009300] [Medline: 3387706]
17. Gennis P, Gallagher J, Falvo C, Baker S, Than W. Clinical criteria for the detection of pneumonia in adults: guidelines for ordering chest roentgenograms in the emergency department. J Emerg Med 1989;7(3):263-268. [doi: 10.1016/0736-4679(89)90358-2] [Medline: 2745948]
18. Melbye H, Straume B, Aasebø U, Dale K. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. Scand J Prim Health Care 1992 Sep 12;10(3):226-233. [doi: 10.3109/02813439209014066] [Medline: 1410955]
19. Reichert S, Gass R, Brandt C, Andrès E. Analysis of respiratory sounds: state of the art. Clin Med Circ Respirat Pulm Med 2008 May 16;2:45-58 [FREE Full text] [doi: 10.4137/ccrpm.s530] [Medline: 21157521]
20. Forgacs P. The functional basis of pulmonary sounds. Chest 1978 Mar;73(3):399-405. [doi: 10.1378/chest.73.3.399] [Medline: 630938]
21. Silverman ME. The Third Heart Sound. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. Boston, MA: Butterworths; 1990.
22. Kagaya Y, Tabata M, Arata Y, Kameoka J, Ishii S. Variation in effectiveness of a cardiac auscultation training class with a cardiology patient simulator among heart sounds and murmurs. J Cardiol 2017 Aug;70(2):192-198 [FREE Full text] [doi: 10.1016/j.jjcc.2016.10.011] [Medline: 27916238]
23. Gurung A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. Respir Med 2011 Sep;105(9):1396-1403 [FREE Full text] [doi: 10.1016/j.rmed.2011.05.007] [Medline: 21676606]
24. Kudriavtsev V, Polyshchuk V, Roy DL. Heart energy signature spectrogram for cardiovascular diagnosis. Biomed Eng Online 2007 May 04;6:16 [FREE Full text] [doi: 10.1186/1475-925X-6-16] [Medline: 17480232]
25. Li S, Li F, Tang S, Xiong W. A Review of Computer-Aided Heart Sound Detection Techniques. Biomed Res Int 2020;2020:5846191 [FREE Full text] [doi: 10.1155/2020/5846191] [Medline: 32420352]

Edited by L Buis; submitted 01.08.20; peer-reviewed by G Gelao, B Bente; comments to author 05.11.20; revised version received 09.11.20; accepted 24.06.21; published 27.07.21.

Please cite as:
Hirosawa T, Harada Y, Ikenoya K, Kakimoto S, Aizawa Y, Shimizu T
The Utility of Real-Time Remote Auscultation Using a Bluetooth-Connected Electronic Stethoscope: Open-Label Randomized Controlled Pilot Trial
JMIR Mhealth Uhealth 2021;9(7):e23109
URL: https://mhealth.jmir.org/2021/7/e23109
doi:10.2196/23109
PMID:34313598

Corrigenda and Addenda

Correction: Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study

Mariska Weenk[1], MD; Harry van Goor[1], MD, PhD, FRCS; Bas Frietman[1], MD; Lucien JLPG Engelen[2]; Cornelis JHM van Laarhoven[1], MD, PhD, MSc; Jan Smit[3], MD, PhD; Sebastian JH Bredie[3], MD, PhD; Tom H van de Belt[2], PhD

[1]Radboud University Medical Center, Department of Surgery, Nijmegen, Netherlands

[2]Radboud University Medical Center, Radboud REshape Innovation Center, Nijmegen, Netherlands

[3]Radboud University Medical Center, Department of Internal Medicine, Nijmegen, Netherlands

Corresponding Author:
Mariska Weenk, MD
Radboud University Medical Center
Department of Surgery
P.O. Box 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 24 361 1111
Fax: 31 24 363 5115
Email: mariska.weenk@radboudumc.nl

Related Article:

Correction of: https://mhealth.jmir.org/2017/7/e91/

(JMIR Mhealth Uhealth 2021;9(7):e31899) doi:10.2196/31899

In "Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study" (JMIR Mhealth Uhealth 2017;5(7):e91), one error was noted.

In the originally published manuscript, an incorrect ORCID number was listed for author Jan Smit. This ORCID number has now been removed.

The correction will appear in the online version of the paper on the JMIR Publications website on July 16, 2021, 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 08.07.21; this is a non–peer-reviewed article;accepted 08.07.21; published 16.07.21.

Please cite as:
Weenk M, van Goor H, Frietman B, Engelen LJLPG, van Laarhoven CJHM, Smit J, Bredie SJH, van de Belt TH
Correction: Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study
JMIR Mhealth Uhealth 2021;9(7):e31899
URL: https://mhealth.jmir.org/2021/7/e31899
doi:10.2196/31899
PMID:34270443

Original Paper

Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study

Kennedy Opoku Asare[1], MSc; Yannik Terhorst[2], MSc; Julio Vega[3], PhD; Ella Peltonen[1], PhD; Eemil Lagerspetz[4], PhD; Denzil Ferreira[1], PhD

[1]Center for Ubiquitous Computing, University of Oulu, Oulu, Finland

[2]Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

[3]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[4]Department of Computer Science, University of Helsinki, Helsinki, Finland

Corresponding Author:
Kennedy Opoku Asare, MSc
Center for Ubiquitous Computing
University of Oulu
Erkki Koiso-Kanttilan katu 3
PO Box 4500
Oulu, FI-90014
Finland
Phone: 358 294 482807
Fax: 358 8 553 2612
Email: kennedy.opokuasare@oulu.fi

Abstract

Background: Depression is a prevalent mental health challenge. Current depression assessment methods using self-reported and clinician-administered questionnaires have limitations. Instrumenting smartphones to passively and continuously collect moment-by-moment data sets to quantify human behaviors has the potential to augment current depression assessment methods for early diagnosis, scalable, and longitudinal monitoring of depression.

Objective: The objective of this study was to investigate the feasibility of predicting depression with human behaviors quantified from smartphone data sets, and to identify behaviors that can influence depression.

Methods: Smartphone data sets and self-reported 8-item Patient Health Questionnaire (PHQ-8) depression assessments were collected from 629 participants in an exploratory longitudinal study over an average of 22.1 days (SD 17.90; range 8-86). We quantified 22 regularity, entropy, and SD behavioral markers from the smartphone data. We explored the relationship between the behavioral features and depression using correlation and bivariate linear mixed models (LMMs). We leveraged 5 supervised machine learning (ML) algorithms with hyperparameter optimization, nested cross-validation, and imbalanced data handling to predict depression. Finally, with the permutation importance method, we identified influential behavioral markers in predicting depression.

Results: Of the 629 participants from at least 56 countries, 69 (10.97%) were females, 546 (86.8%) were males, and 14 (2.2%) were nonbinary. Participants' age distribution is as follows: 73/629 (11.6%) were aged between 18 and 24, 204/629 (32.4%) were aged between 25 and 34, 156/629 (24.8%) were aged between 35 and 44, 166/629 (26.4%) were aged between 45 and 64, and 30/629 (4.8%) were aged 65 years and over. Of the 1374 PHQ-8 assessments, 1143 (83.19%) responses were nondepressed scores (PHQ-8 score <10), while 231 (16.81%) were depressed scores (PHQ-8 score ≥10), as identified based on PHQ-8 cut-off. A significant positive Pearson correlation was found between screen status–normalized entropy and depression (r=0.14, P<.001). LMM demonstrates an intraclass correlation of 0.7584 and a significant positive association between screen status–normalized entropy and depression (β=.48, P=.03). The best ML algorithms achieved the following metrics: precision, 85.55%-92.51%; recall, 92.19%-95.56%; F1, 88.73%-94.00%; area under the curve receiver operating characteristic, 94.69%-99.06%; Cohen κ, 86.61%-92.90%; and accuracy, 96.44%-98.14%. Including age group and gender as predictors improved the ML performances. Screen and internet connectivity features were the most influential in predicting depression.

Conclusions: Our findings demonstrate that behavioral markers indicative of depression can be unobtrusively identified from smartphone sensors' data. Traditional assessment of depression can be augmented with behavioral markers from smartphones for depression diagnosis and monitoring.

(JMIR Mhealth Uhealth 2021;9(7):e26540) doi:10.2196/26540

KEYWORDS

mHealth; mental health; mobile phone; digital biomarkers; digital phenotyping; smartphone; supervised machine learning; depression

Introduction

Background

Depression is one of the most prevalent, complex, and heterogeneous mental health challenges of our time. In 2020, the World Health Organization (WHO) estimated that depression has impacted 264 million people worldwide [1], and it is projected to be the leading contributing factor to global disease burden by 2030 [2]. In these individuals, depression inflicts recurrent episodes of guilt, sadness, cognitive impairments, suicidal ideation, and sleep disturbances [1,3-5]. Depression increases the risk and medical costs of many medical disorders such as stroke, Parkinson, or Alzheimer [6-11]. Depression is treatable with psychotherapy and medication. Yet, in many individuals with depression, it remains undiagnosed and untreated due to barriers such as social stigma and inaccurate assessment methods [1,3,12,13]. The ability to detect early warning signs of depression, continuously and as effortlessly as possible, by extending current assessment methods could have a significant impact in mitigating or addressing depression and its related negative consequences [3,10,11,14].

For the past 30 years, clinician-administered and self-reported questionnaires remain the gold standard in the assessment and diagnosis of depression [3,13]. However, the limitations of these traditional depression assessment methods have been debated. Such methods are applied sparingly (eg, a couple of times within a year), thus missing out on the moment-by-moment behavioral patterns of individuals between health assessments. Lastly, self-reported appraisals are affected by memory and recall biases in reconstructing past events and may be prone to socially desirable reporting from individuals [12,15-17].

Today, smartphones and wearables offer a unique opportunity to overcome limitations in traditional depression assessment methods. Smartphones and wearables (eg, Fitbit, Oura Rings, and smartwatches) have become ubiquitous in the global population, they are inherently personal, and people are continuously monitored through their embedded sensors (eg, camera, accelerometer, global positioning system [GPS], Bluetooth, and many more) [18,19]. Instrumenting smartphones and wearables to capture in situ, fine-grained, and moment-by-moment data sets with sensing apps [20-24] has made it possible to passively collect data sets in naturalistic settings. Inherent in these data sets are behavioral patterns: routines, rhythms, activities, and interactions that are useful in complementing traditional depression assessment methods, in studying the mental health of individuals, and in developing timely mental health interventions [14,22,25-28].

Related Work

A growing body of research in smartphone and wearable sensing, human behavior modeling has improved our understanding of the relationship between mental health and biomarkers [3,20-22,26,27,29-31]. In medicine, biomarkers are pathological, anatomical, or physiological characteristics that are quantified and evaluated as indicators of a biological process or a response to medical interventions [23]. Here, we define biomarkers (or digital biomarkers) of mental health as quantifiable behaviors (or features) extracted from smartphone or wearable data that can be monitored and collected over time to objectively assess mental health and effectiveness of interventions. Monitoring these biomarkers' fluctuations is essential in the early detection and treatment of mental health disorders [3,32]. For example, in Alzheimer disease, biomarkers such as cognitive, sensory, and motor degeneration precedes clinical diagnosis for about 10 or 15 years [32].

In the StudentLife study [22], for example, geolocation, sleep, and activity-based biomarkers were extracted from a data set collected from 48 students over 10 weeks. Significant correlations were found between the following digital biomarkers: sleep duration (r=–0.360), activity duration (r=–0.388), traveled distance (r=0.338), and various mental health symptoms. Similarly, Wang et al [3] collected data sets with the StudentLife sensing app from 83 college students across two 9-week terms. Significant correlations were found between depression and accelerometer-based biomarkers (mean stationary time, r=0.256) and screen usage–based biomarkers (mean unlock duration, r=0.283). With an analysis of variance, a significant difference in unlock duration (F=5.733) was found between depressed and nondepressed groups. Saeb et al [33] in their study of 40 participants for 2 weeks also found a statistically significant correlation between depression and GPS location–based biomarkers (ie, variance in locations visited, [r=0.58] and regularity in 24-hour movement rhythm [r=0.63]) and phone usage–based biomarkers (ie, phone usage frequency [r=0.54]).

Another promising source of biomarkers is wearable devices [30].

Actigraphy-based biomarkers that quantify time sequences of rest and active behaviors with accelerometer sensors are known to be useful in predicting mood disorders such as depression and bipolar disorder [34,35]. In a 2-week study on 23 participants, Jacobson et al [34] extracted biomarkers from data sets collected with wrist-worn actigraph watches. A machine learning (ML) model trained with these digital biomarkers could predict the depression status of participants with high accuracy

(ie, accuracy 89%, Cohen κ=0.773). In another 2-week study [35] on 40 geriatric participants, accelerometer-based digital biomarkers were extracted from wrist-worn actigraph watches. With 4 ML models, the study found that these biomarkers could predict depression with a high accuracy (ie, accuracy 0.910, precision 0.929, and specificity 0.940). Other promising biomarkers from wearable devices are heart rate variability, which has been found to be consistently lower in patients with psychiatric disorders, electrodermal activity, and skin conductance [36,37].

Taken together, previous research has shown the potential of quantifying human behavior from smartphones and wearables data set as biomarkers. These biomarkers are insightful in understanding depression.

Objectives

In this study, we aim to investigate the feasibility of predicting depression using digital biomarkers quantified from a smartphone data set. To this end, we explore the relationship between digital biomarkers and depression severity with statistical methods. We investigate whether depression can be predicted with digital biomarkers using supervised ML algorithms.

Methods

The Data Set

We utilized an existing data set collected in a longitudinal observational study with the Carat app [21], derived from a cohort of anonymous Android participants worldwide [38]. The data set was collected from 843 participants between March and August 2018 (~6 months).

The Carat app is a mobile sensing app, originally developed by a team of researchers from the University of Helsinki and the University of California, Berkeley, for smartphone energy consumption research [21,39]. The Carat app is freely available on mobile app stores and gives users personalized smartphone battery consumption reports. Anonymous users worldwide who install the Carat app may voluntarily be recruited to contribute their data set to research.

The data set used in this study was a subset of the large-scale crowdsourced Carat app data set from anonymous volunteers. The study data set was collected for a multifaceted purpose, which includes studying the relationship between smartphone app usage and Big 5 personality traits [40]; studying the similarities and differences in demographic, geographic, and cultural factors of smartphone usage [38]; and mental health research. The advertisement for the recruitment of participants was sent as push notifications through the Carat app to 25,323 verified users (ie, users with matching time zone and mobile country code) [38].

All participants in this data set are Android-based smartphone users, who explicitly and voluntarily gave their consent from their mobile devices after they were informed about the purpose of the data collection, the data collection procedures, and management of the data set. The data set does not contain personally identifiable information, and was collected under the institutional review board license from the University of California, Berkeley and the University of Helsinki [21,40].

Mobile Sensing Variables Collected by Carat

Besides battery consumption data, the Carat Android app unobtrusively collected participants' time zone and time-stamped data, including foreground app usage (ie, app the participant has interacted with), internet connectivity (ie, connected and disconnected states), and screen lock and unlock logs. This data set was sampled at each 1% battery change (ie, while charging or discharging) on the participants' smartphone. The Carat app also collected participant's demographic information, including age, gender, education, and occupation via a self-report.

Mental Health Assessment

In addition to the mobile sensing and demographic variables, depression severity was assessed by a self-report instrument. Participants answered the 8-item Patient Health Questionnaire (PHQ-8) [41] at 2-week intervals as specified by the PHQ-8 protocol. Although the depression assessments were self-reported, the PHQ-8 is clinically validated for the assessment of depression severity, has high internal consistency (Cronbach α=.82), and has been used in several previous studies [3,7,41]. PHQ-8 measures depression severity for the past 2 weeks with items such as "Little interest or pleasure in doing things," "Feeling down, depressed, or hopeless," "Trouble falling or staying asleep, or sleeping too much." Each item of the PHQ-8 is scored on a scale from 0 (Not at all) to 3 (Nearly every day). The total PHQ-8 score ranges from 0 to 24, with a score of 10 or more indicating major depression or other severe forms of depression [41].

Data Inclusion and Exclusion

For each participant's data set, we excluded days with at least 10 missing log intervals (ie, days where no data were logged by the Carat app for at least 10% battery charging or discharging periods). Next, we only included PHQ-8 responses from participants with at least 8 days of data within the preceding 2 weeks of PHQ-8 response. Consequently, the final data contained 629 participants, 1374 PHQ-8 responses with 13,898 days of participants' data set.

Feature Engineering

Characterization of Data Set

Our data set is primarily categorized into screen status, internet connectivity, and foreground app usage logs. For data preprocessing, we converted the time stamps of the data set to local date and time, using the participants' time zone. We computed digital biomarkers (herein features) by quantifying the per-participant hourly and daily behavioral patterns (ie, routines, irregularity, variability) from these data sets with simple counts, SDs, entropy [6,14,25,42-45], and regularity index [27,44] measures. All computed features were merged per participant at the day level.

Entropy

We computed entropy to capture the degree of variability, complexity, disorder, and randomness in the participant behavior

states from screen status (ie, on and off states), internet connectivity (ie, disconnected and connected states), and foreground app (ie, the frequency of use per app) over a 24-hour period of each day. Entropy was calculated using the Shannon entropy [45] formula:

where N is the number of states and p_i is the percentage of the state i in the time series data. For example, a higher screen status entropy reflects the fact that the participant's screen on and off pattern is more distributed between on and off states, albeit with a high degree of uncertainty and complexity in the transition between the screen on and off states in a 24-hour period. Conversely, a lower screen status entropy reflects that fact that the participant's screen is much often in one state (on or off) over a 24-hour period. In addition to entropy, we computed normalized entropy as the entropy divided by log(N).

Regularity Index

Regularity index quantifies routines in participant behaviors by capturing the similarity (or difference) in participant behaviors between the same hours across different days. For internet connectivity, for instance, the regularity index quantifies the routineness of the participant's internet connectivity behavior at the same hours (eg, every 9 am) for all days. We determined the hourly values as follows: for screen status, the modal screen status for each hour; for internet connectivity, the modal connectivity state for each hour; and for foreground app usage, the number of distinct apps usage for each hour.

Following the regularity index computation method of Wang et al [44], we computed the regularity index of the screen, internet connectivity, and foreground app usage for days a and b using the formula

where a and b are 2-day pairs, T=24 hours, and  is the rescaled (ie, between –0.5 and 0.5) value of hour t of day b. For each day, we computed the average regularity index for all combinations of that day and other days of the week.

Standard Deviation and Counts

The SD features capture the variance of daily behavior between 4-day epochs based on the hour of the day. We defined morning as the 6-11th hour, afternoon as the 12-17th hour, evening as the 18-23rd hour, and night as the 0-5th hour of the day. We computed the count of each screen status, the count of each internet connectivity status, and the count of foreground app usage per day epoch. With these counts per day epoch, we computed the SD per day.

We also computed the day level count of each screen status, the count of each internet connectivity status, and the count of foreground app usage. Additionally, we computed the count of minutes until the first and last use of foreground app per day.

Correlation and Association Analysis

Before beginning the statistical analysis, we pooled (ie, aggregated) the extracted features within the preceding 2 weeks (ie, assessment window) of each PHQ-8 response from a participant. The pooling is to ensure that the timelines of the feature variables in the analysis are aligned with those of the PHQ-8 assessment window. The pooling was done as follows: for each PHQ-8 response, we pooled all entropy and regularity index features by computing the average feature values for all days within the PHQ-8 assessment window. Instead of average values for SD, we took a different approach due to the additive properties of SD measures. For SD features, we computed the pooled SD [46,47].

For correlation analysis, we used the pooled data to quantify the linear relationship between the features and depression severity (ie, PHQ-8 responses). The correlations were computed using the Pearson correlation coefficient. Full information maximum likelihood [48-50] was used in the correlation analysis to avoid biases introduced by missing data. We used the Holm–Bonferroni [51] method to adjust the P values for multiple testing, with a false discovery rate of 0.05.

For association analysis, we used the bivariate linear mixed model (LMM) [8,44,52-55] to study the association between the pooled features and the PHQ-8 response. The data set in this study is a longitudinal data set with repeated measures from the same individuals. Given this nested structure, the assumption of normally and independently distributed residuals would be violated in linear regression models. Hence, we opted for LMM, which takes into account fixed and random variations in the data set in respect of a grouping variable, the participant in this case. LMM also reduces the likelihood of Type I error [56]. To verify our decision for LMM, we computed the intraclass correlation (ICC). ICC > 0.05 necessitates LMM.

In the LMM, we used multiple imputation to handle missing data, taking into account the nested structure of the data set [57]. Using predictive mean matching for multilevel data [58], a total of 20 imputed data sets were generated. We used Robin's rule [59] to pool the results of the LMM run with each imputed data set. In the LMM analysis, all features were normalized to have a 0 mean and a unit SD. To address multiple comparison problems, we adjusted the P values in the LMM to control the false discovery rate with the Benjamini–Hochberg procedure [60,61]. Adjusted P value <.05 was considered to be significant.

Predictive Analysis With Machine Learning

Machine Learning Setup

We developed population-based supervised ML classifiers to explore how the digital biomarkers/features predict the depression state of an individual. We also explored whether including participants' self-reported demographics as features would improve the ML classifier performance.

To this end, we used 5 supervised ML models: random forest (RF), support vector machine (SVM) with radial basis function (RBF) kernel, XGBoost (XGB) [62], K-nearest neighbor (KNN), and logistic regression (LR). With the RBF kernel, an SVM classifier, a linear classifier, can classify nonlinear data sets [63,64]. These algorithms have been used in previous work [65-69] on mental health studies. We used the same pooled and imputed data set from statistical analysis, but ensured that all records are distinct. We labeled our data set with 2 classes,

based on PHQ-8 scoring guidelines, where PHQ-8 score ≥ 10 is depressed (label 1) and PHQ-8 score <10 is nondepressed (label 0). We created 2 data sets for the ML modeling: (1) a data set with PHQ-8 scores as labels, and digital biomarkers/features as predictors, and (2) a second data set with PHQ-8 scores as labels and digital biomarkers/features, age group, and gender as predictors. The age group and gender demographics were converted from categorical to numerical data using one-hot encoding [63,70].

All the ML modeling was performed using stratified and nested cross-validation [71,72]. The nested cross-validation is a state-of-the-art procedure to prevent overfitting and overestimation of the hyperparameters of the ML classifiers. Stratified 10 folds in the outer cross-validation and stratified 3 folds in the inner cross-validation were used. With this approach, for each iteration of the outer cross-validation, 1 stratified fold was used as a testing data set. The remaining 9 stratified folds were used for hyperparameter optimization in the inner cross-validation.

The hyperparameter optimization in the inner cross-validation was done with grid search over a grid of parameters, where all combinations of hyperparameters are exhaustively considered. We use the F1 (macro averaged) score to select the most optimized hyperparameters.

Imbalanced Data Handling

It is worth noting that in the ML setup, we used stratified sampling in the nested cross-validation. The stratified sampling ensures the splitting of the data set into folds that have an equal proportion of each class (ie, labels 1 and 0). However, the proportion of each class is still dependent on its availability in the data set. We handled class imbalance in the training data set with the synthetic minority over-sampling technique (SMOTE) [65,73,74], which generates synthetic data for the minority class, resulting in a balanced training data set.

Feature Analysis

We used the permutation importance method [75,76] to compute the importance of features. The permutation importance method is model agnostic and computes the proportional decrease in a model score when features are randomly shuffled. We used the area under the curve (AUC) receiver operating characteristic as the model in the permutation importance computation. We computed the feature importance using the test set of the outer cross-validation. For each ML classifier, we ranked the features by the average feature importance computed across all 10 folds of the outer cross-validation.

Model Evaluation

We created 2 baseline classifiers to benchmark the performance of the ML classifiers. The first baseline is a random weighted classifier (RWC) with 10,000 randomly generated predictions based on a multinomial distribution of the nondepressed and depressed classes. The second baseline is a decision tree (DT) classifier trained using the same approach as the ML classifiers, but with age group and gender as features. The performance of the ML classifiers and baseline classifiers was measured using the following performance metrics: accuracy, precision, recall, AUC, F1 score, and Cohen κ. The precision, recall, and F1 scores were computed with an emphasis on predicting the depressed score (ie, label 1).

Software

Data preprocessing and feature extraction pipeline were created with Python (version 3.7.6) and R (version 4.0.2) programming languages, using Snakemake [77] and RAPIDS [78] for workflow management. All statistical analysis was performed in R, with mice [79] package for multiple imputation, lmer4 [80] and lmerTest [81] packages for LMM, and psych [50] package for computing correlation. All the ML was done in Python, with scikit-learn [63], imbalanced-learn [33], and XGB library [62,82].

Results

Participants' Demographics

Self-reported demographic data from the 629 participants included in our analyses show that 69/629 (10.97%) were females, 546/629 (86.8%) were males, and 14/629 (2.2%) were nonbinary or preferred not to disclose their gender.

For the participants' age distribution, 73/629 (11.6%) were aged between 18 and 24, 204/629 (32.4%) were aged between 25 and 34, 156/629 (24.8%) were aged between 35 and 44, 166/629 (26.4%) were aged between 45 and 64, and 30/629 (4.8%) were aged 65 years and over. The participants were distributed across at least 56 different countries, including 91/629 (14.5%) from unknown countries, 199/629 (31.6%) from the USA, 66/629 (10.5%) from Finland, 32/629 (5.1%) from Great Britain, 42/629 (6.7%) from Germany, and 29/629 (4.6%) from India. The data set also has participants from varied educational and occupational backgrounds. Table 1 provides summary statistics of the 629 participants in this study.

Table 1. Summary statistics of participants who were included in the data analysis (N=629).

Variable	Value, n (%)
Age (years)	
18-24	73 (11.6)
25-34	204 (32.4)
35-44	156 (24.8)
45-64	166 (26.4)
≥65	30 (4.8)
Gender	
Female	69 (11.0)
Male	546 (86.8)
Other or Rather not tell	14 (2.2)
Education	
Elementary school/basic education	9 (1.4)
High school/sixth form/other upper secondary level	98 (15.6)
No education or rather not to tell	5 (0.8)
Professional graduate degree/higher university degree (master's or equivalent)	193 (30.7)
Research graduate degree (PhD or equivalent)	34 (5.4)
Undergraduate degree/lower university degree (bachelor's or equivalent)	228 (36.2)
Vocational school/trade school/other education leading to a profession	62 (9.9)
Occupation	
Agricultural forestry or fishery	1 (0.2)
Clerical support	14 (2.2)
Craft and trade or plant and machine operations	8 (1.3)
Entrepreneur or freelancer	30 (4.8)
Manager	59 (9.4)
No suitable option or rather not to tell	34 (5.4)
Professional	227 (36.1)
Retired	39 (6.2)
Sales or services	29 (4.6)
Staying at home (eg, with kids)	5 (0.8)
Student	74 (11.8)
Technician or associate professional	90 (14.3)
Unemployed or between jobs	19 (3.0)
Country	
Unknown	91 (14.5)
USA	199 (31.6)
Finland	66 (10.5)
Great Britain	32 (5.1)
Germany	42 (6.7)
Canada	16 (2.5)
India	29 (4.6)
Other[a]	154 (24.5)

[a]Comprising 49 different countries with less than 15 participants, including South Africa, Morocco, Brazil, Philippines, Qatar, Japan, Russia, and Denmark.

Smartphone Data and PHQ-8 Distribution

We had 1374 PHQ-8 responses. Table 2 presents the distribution of participants and their corresponding number of responses in the PHQ-8 data set. All PHQ-8 responses were collected every 2 weeks. At the minimum, 316/629 (50.2%) participants responded 1 time to the PHQ-8 depression assessment, and at the maximum, 1/629 (0.2%) participants responded 7 times. The mean number of responses per participant is 2.18 (SD 1.57).

For the distribution of the PHQ-8 scores, 1143/1374 (83.19%) responses were nondepressed scores (PHQ-8 score <10), while 231/1374 (16.81%) were depressed scores (PHQ-8 score ≥10). The mean PHQ-8 score is 5.19 (SD 5.22; range 0-24).

The number of smartphone data set days was 13,898 for all 629 participants. Table 3 shows the distribution of participants and their corresponding number of days in the smartphone data set. The mean number of days per participant is 22.1 (SD 17.90; range 8-86).

Table 2. Distribution of participants' contribution to the PHQ-8[a] responses (N=629).

Participants, n (%)	PHQ-8 responses, n
316 (50.2)	1
129 (20.5)	2
57 (9.1)	3
47 (7.5)	4
40 (6.4)	5
39 (6.2)	6
1 (0.2)	7

[a]PHQ-8: 8-Item Patient Health Questionnaire.

Table 3. Distribution of participants' smartphone data set days (N=629).

Days, n	Participants, n (%)
8-14	364 (57.9)
15-28	126 (20.0)
29-42	53 (8.4)
43-56	34 (5.4)
57-70	29 (4.6)
71-84	22 (3.5)
85-98	1 (0.2)

Features Engineered From Smartphone Data Set

In all, we computed 22 features from the smartphone data set. All features were aggregated at the day level. For example, the *screen_offCount* feature is the count of all screen off states during the day. We summarize the engineered features in Multimedia Appendix 1.

Correlation and Association Between Features and Depression

We found a significant positive correlation between screen status–normalized entropy and depression (r=0.14, P<.001). We found no significant correlation between other screen, app, and internet connectivity features and PHQ-8 depression score. Multimedia Appendix 2 presents the full Pearson correlation coefficients and adjusted P values, with the Holm–Bonferroni method, between features and PHQ-8 depression score.

Regarding the association analysis, we found an ICC of 0.7584; thus, 75.84% of the variations in the features are explainable by the interindividual differences. We found a significant positive association between screen status–normalized entropy and depression (β=.48, P=.03). We found no significant association between other screen, app, and internet connectivity features and depression. Multimedia Appendix 2 presents the results of the LMM analysis showing the estimates (β) and adjusted P values calculated using the Benjamini–Hochberg method.

Predicting Depression From Features

The overall performance of the ML classifiers trained with features only (ie, with no demographics data set) is listed in Table 4. Table 5 shows the ML classifier performance with the features plus age group and gender data set as predictors. The tuned hyperparameters of all ML classifiers are detailed in Multimedia Appendix 3. The performance of all 10 cross-validation folds of the features-only data set and that of the feature plus demographics data set are detailed in Multimedia

Appendix 3. The performance of the DT and RWC baselines is shown in Table 6.

As shown in Table 4, it is evident that nonlinear classifiers such as XGB, RF, and KNN had superior performance in all metrics than LR. SVM (ie, SVM with RBF kernel) also performed better than LR.

In terms of precision, recall, and F1 scores, which were computed with an emphasis on the predictive performance of the positive label (ie, depressed score, PHQ-8 ≥ 10), XGB was the best performing classifier, followed by RF and KNN. XGB, RF, and KNN performed better than the RWC and DT baselines, as shown in Table 6.

Likewise, with AUC and Cohen κ performance metrics, which take into consideration both positive and negative labels (ie, nondepressed score, PHQ-8 <10), XGB, RF, and KNN classifiers had the best performance, as shown in Table 4. The AUC and Cohen κ are not biased by imbalance labels.

As shown in Table 4, the worst performing classifier is LR. Compared with the baseline classifiers in Table 6, the RWC and DT baselines classifiers outperform the LR classifier in terms of recall. The LR could predict the PHQ-8 depression score barely better than the RWC and DT baselines in all other performance metrics. The RWC baseline classifier also outperformed the SVM classifier on the recall metrics.

When age group and gender were included with features as predictors, we observed a general improvement in all performance metrics for all classifiers, as shown in Table 5. The SVM classifier had the most substantial improvement with precision increasing by 18.48% and Cohen κ increasing by 19.3%. KNN had a 6.58% improvement in precision and a 5.1% improvement in F1 score. RF and XGB classifiers had marginal improvements in all performance metrics. The worst performing classifier (ie, LR) had some gains in performance, but it could still barely outperform the DT baseline classifier in Table 6 in all performance metrics.

Table 4. Average and SDs of accuracy, precision, recall, F1, area under the curve, and Cohen κ metrics for 10-fold cross-validation, with features-only data set as predictors.

Metric	RF[a], mean (SD)	XGB[b], mean (SD)	SVM[c], mean (SD)	LR[d], mean (SD)	KNN[e], mean (SD)
Accuracy	97.97 (0.37)	98.14 (0.37)	85.68 (1.16)	59.27 (1.45)	96.44 (0.52)
Precision	92.50 (1.78)	92.51 (1.25)	51.98 (2.58)	20.29 (1.25)	85.55 (1.97)
Recall	94.38 (1.86)	95.56 (1.99)	80.67 (2.36)	57.25 (4.14)	92.19 (2.24)
F1	93.41 (1.19)	94.00 (1.21)	63.20 (2.29)	29.95 (1.87)	88.73 (1.63)
Area under the curve	98.83 (0.67)	99.06 (0.54)	89.47 (1.06)	62.43 (2.22)	94.69 (1.15)
Cohen κ	92.21 (1.41)	92.90 (1.43)	54.83 (2.92)	9.66 (2.38)	86.61 (1.93)

[a]RF: random forest.

[b]XGB: XGBoost.

[c]SVM: support vector machine.

[d]LR: logistic regression.

[e]KNN: K-nearest neighbor.

Table 5. Average and SDs of accuracy, precision, recall, F1, area under the curve, and Cohen κ metrics for 10-fold cross-validation, with features, age group, and gender data set as predictors.

Metric	RF[a], mean (SD)	XGB[b], mean (SD)	SVM[c], mean (SD)	LR[d], mean (SD)	KNN[e], mean (SD)
Accuracy	98.55 (0.40)	98.56 (0.31)	92.61 (0.46)	60.37 (1.39)	98.09 (0.26)
Precision	95.65 (1.59)	94.93 (1.08)	70.46 (1.63)	21.40 (1.20)	92.13 (1.41)
Recall	94.78 (1.59)	95.62 (1.52)	88.76 (3.19)	60.00 (3.94)	95.62 (1.56)
F1	95.20 (1.31)	95.27 (1.03)	78.52 (1.41)	31.54 (1.78)	93.83 (0.85)
Area under the curve	99.01 (0.51)	99.36 (0.33)	95.45 (1.00)	66.62 (3.06)	97.07 (0.73)
Cohen κ	94.34 (1.55)	94.42 (1.21)	74.13 (1.66)	11.74 (2.28)	92.69 (1.00)

[a]RF: random forest.

[b]XGB: XGBoost.

[c]SVM: support vector machine.

[d]LR: logistic regression.

[e]KNN: K-nearest neighbor.

Table 6. Average and SDs of accuracy, precision, recall, F1, area under the curve, and Cohen κ metrics for the RWC and DT baselines.

Metric	RWC[a], mean (SD)	DT[b], mean (SD)
Accuracy	25.80 (0.33)	46.80 (3.77)
Precision	15.21 (0.36)	18.70 (0.66)
Recall	84.79 (0.86)	74.33 (4.71)
F1	25.80 (0.24)	29.85 (0.75)
Area under the curve	50.00 (0.47)	62.94 (1.38)
Cohen κ	0.00 (0.32)	7.33 (1.26)

[a]RWC: random weighted classifier; RWC metrics is the average and SD of 10,000 random predictions.

[b]DT: decision tree; DT metrics is the average for 10-fold cross-validation, with age group and gender only as features.

Feature Importance Analysis

We present the mean permutation feature importance in predicting PHQ-8 depression score across the 10-fold cross-validation with the top 3 performing ML classifiers (ie, XGB, RF, and KNN) in Figures 1-3.

For the XGB classifier in Figure 1, the top 5 most important features were the internet regularity index, screen on count, screen regularity index, screen status entropy, and the screen off count.

For the RF classifier in Figure 2, the top 5 most important features were screen status–normalized entropy, screen regularity index, screen off count SD, screen off count, and internet regularity index.

Likewise, for the KNN classifier in Figure 3, the top 5 most important features are the internet regularity index, screen status–normalized entropy, screen regularity index, internet status–normalized entropy, and the internet status entropy. As shown in Figure 3 app entropy, count, distinct count, regularity index, and count SD were less important to the KNN classifier. Removing these less important features could further improve the performance of the KNN classifier.

By ranking all important features for KNN, XGB, and RF classifiers, the top 5 most were the screen regularity index, screen status entropy, internet regularity index, screen status–normalized entropy, and the screen off count SD. App count SD is the least important feature for all classifiers (Figures 1-3), and could be removed to improve ML classifier performance in the case of RF and KNN.

Figure 1. Mean permutation feature importance across 10-fold cross-validation with the XGBoost machine learning classifier.

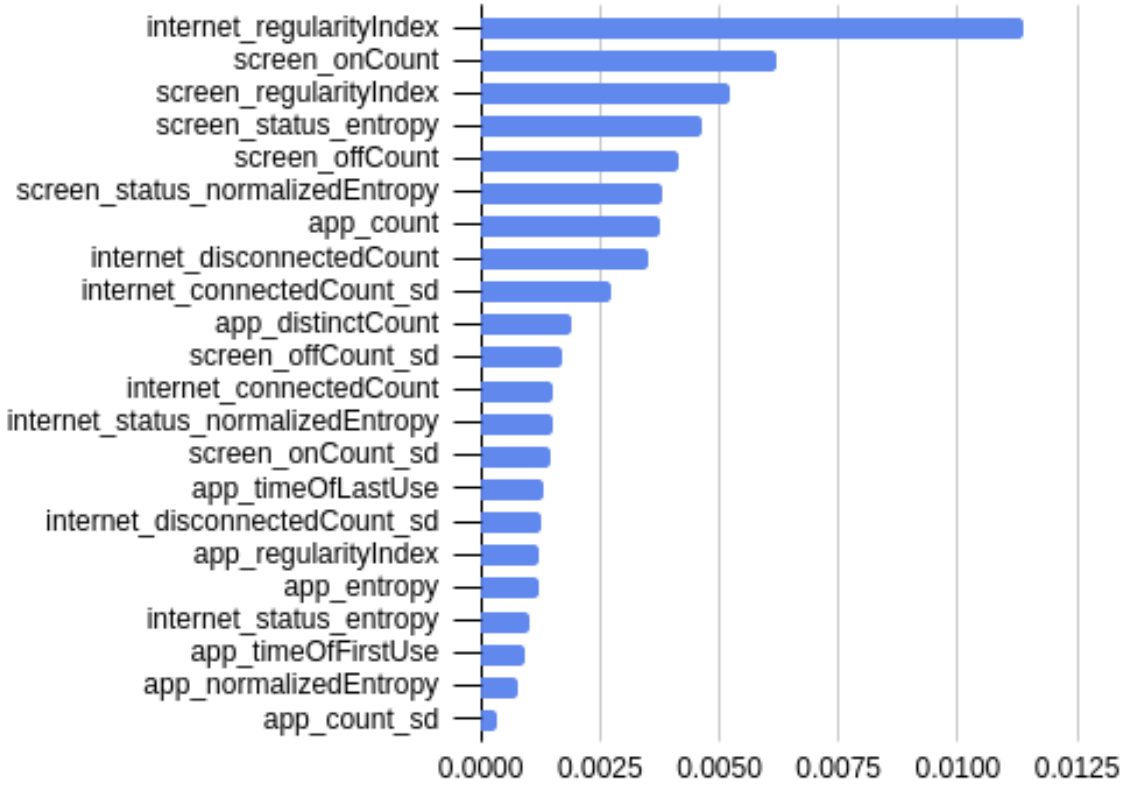


Figure 2. Mean permutation feature importance across 10-fold cross-validation with the random forest machine learning classifier.

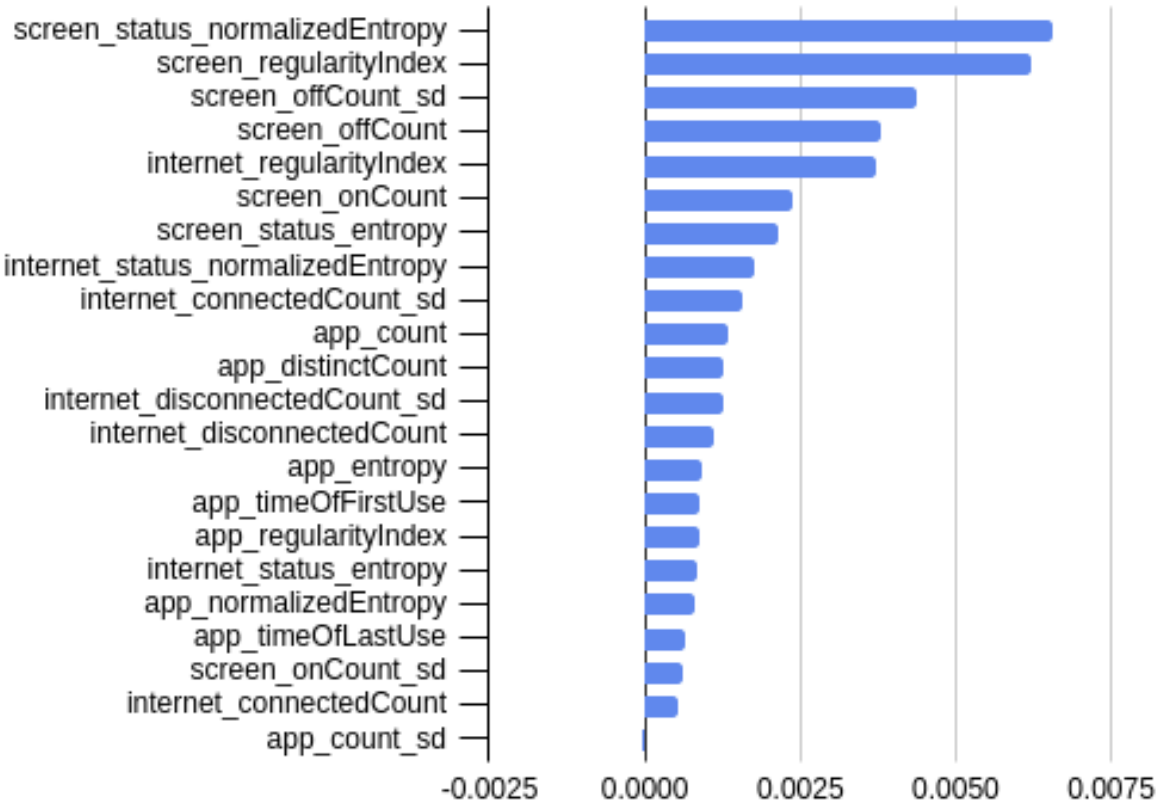
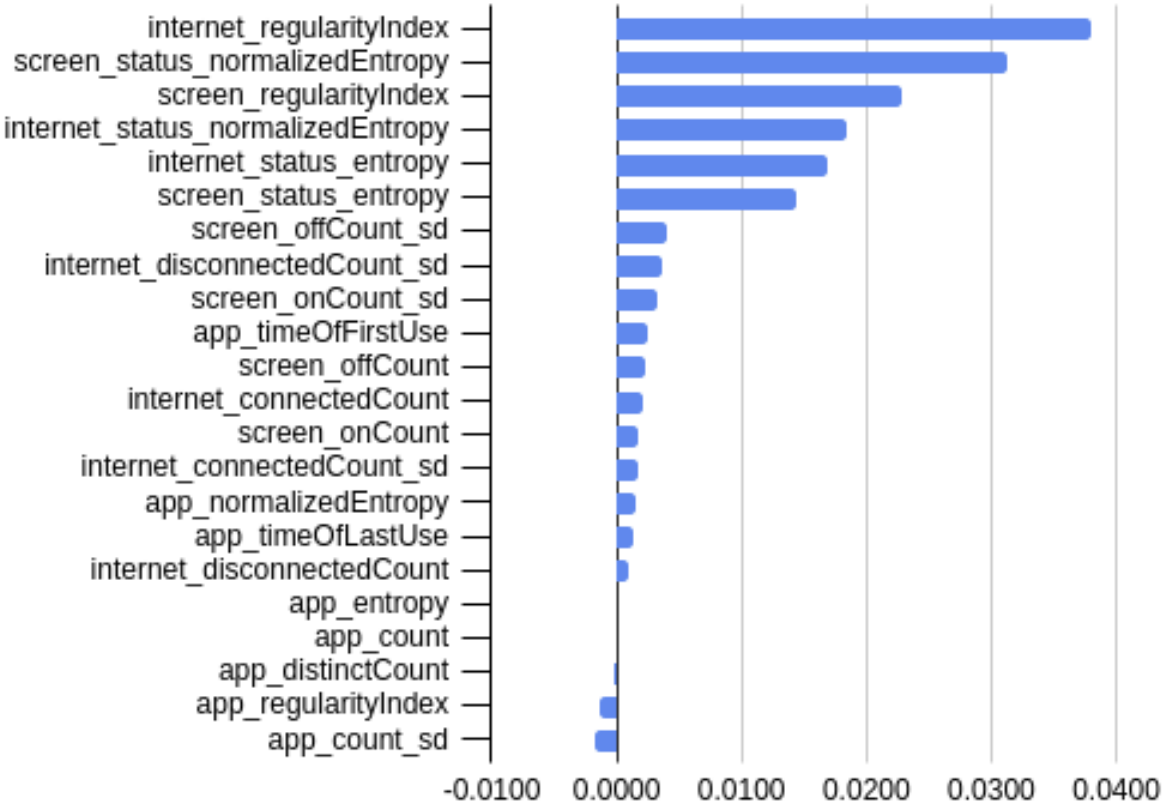


Figure 3. Mean permutation feature importance across 10-fold cross-validation with the K-nearest neighbor machine learning classifier.



Discussion

Overview of Data Set Employed

Our objective was to investigate the feasibility of predicting depression using multivariate digital biomarkers quantified from smartphone data sets collected in a real-world study. In this study, we used 13,898 days of smartphone data set, and 1374 PHQ-8 depression assessments from 629 participants to explore the feasibility of detecting depression from participants' behavioral markers (ie, digital biomarkers) quantified from their smartphones. We focused on finding the relationship between repeated measures of depression scores and participant's digital biomarkers and developing predictive models to classify depressed and nondepressed symptom severity scores.

Principal Results

This data set was collected from a heterogeneous geographic (ie, from at least 56 different countries), occupational, and educational population, with high interindividual differences (ie, 75.84% interclass correlation).

Despite this heterogeneity, digital biomarkers extracted from participants' smartphone data set were able to predict participants' depression state (ie, depressed or nondepressed) with high predictive performance using ML models. The ML models achieved the following: precision, 85.55%-92.51%; recall, 92.19%-95.56%; F1, 88.73%-94.00%; AUC, 94.69%-99.06%; Cohen κ, 86.61%-92.90%; and accuracy, 96.44%-98.14%. These findings show that predictive modeling of mental health using digital biomarkers is not only possible in small homogenous populations [83,84], but also in a more general population, which further supports the scalability of this approach and its potential positive impact on health care if implemented (eg, early detection of mental disorders, RED-flag systems after treatment).

Moreover, we found that the predictive performances of ML classifiers improved when demographic characteristics were included among predictors, indicating that such variables should also be included in clinical applications. Previous studies suggest a relationship between demographic factors, smartphone usage behavior, and depression [38,85,86]. Thus, encoded in the demographic data of this study's population is additional information that is useful in predicting the depression state of the participants. Therefore, the inclusion of additional data from clinical information systems (eg, blood parameters, previous clinical diagnosis) might be a valuable way to further increase the performance of prediction models.

Interestingly, tree-based, nearest neighbor–based classifiers had superior performance over linear classifiers, including SVM with RBF kernel, corroborating the existence of nonlinear relationships between digital biomarkers and depression. This finding further supported the correlation finding, which failed to replicate previous results reported in Saeb et al [33]. We could only identify that participants with depression symptoms were more likely to lock and unlock their phone's screen in a random and uncertain manner (ie, significant positive correlation between screen status–normalized entropy and depression, r=0.14, P<.001). The screen status–normalized entropy biomarker quantified the frequency and distribution (ie, complexity and uncertainty) in the transition of the participants' phone screen on and off states. All other indicators were nonsignificant. Moreover, correlation coefficients only measure the extent of the linear relationship between variables [87,88]. Instead of correlations, previous studies on mental health relied on the mutual information (MI) method from *Information* theory [87-89]. The advantage of using the MI method is that the MI measures both linear and nonlinear statistical dependencies between variables.

Given the high ICC, we tested whether LMM, a much robust method for finding linear relationships, can identify additional linear relationships. However, the results from the association analyses further showed that a unit increase in the screen status–normalized entropy positively increases the average depression score (β=.48, P=.03), but all other variables remained nonsignificant. This suggests that conventional statistical methods (eg, correlation or LMM) may not depict the complex nonlinear relationship between digital biomarkers and depression, and indeed more powerful methods such as ML models (eg, XGB) are needed to make better predictions [90].

The heterogeneity and inconsistency in correlation findings (ie, linear relationships) in the field are common issues [26]. Until now, it is unclear whether this is due to differences in sociodemographic characteristics in samples, in used sensors, in the method to calculate features, small sample sizes and lack of power, or even due to other between-study factors. Meta-analysis on digital markers and health outcomes (eg, depression) would be highly valuable to clearly show whether and to which extent linear relationships exist. Using meta-regression, the factors causing the differences in correlation findings may also be identified.

Nevertheless, both the correlation findings and the feature importance analysis in the prediction models clearly showed that participants' phone screen (lock and unlock) behaviors, such as routinely and randomly locking and unlocking phone screen, and internet connectivity behaviors played the most important role in predicting their depression state. The findings in this study are also supported by prior research that investigated the relationship between screen interactions and mental health [3,27,28,33]. Passively sensed participants' smartphone screen interaction (ie, on and off states) behavior was demonstrated to be an important predictor of mental health [27]. Similar findings have been reported previously [3,33], where the number of times a participant interacts with their phone, including screen lock and unlocks, was found to correlate with participants' mental health state. In a neuroscience study [91], screen unlocks were found to be important behavioral markers that correlate and predict resting state brain functional connectivity, which is known to be associated with depression [92]. On internet usage behaviors, research has demonstrated an association between internet usage patterns and depression [93,94], which was also a key feature in our analysis. Thus, including these features in future studies is highly recommended.

Limitations and Future work

Given the crowdsourced nature of the deployment of the Carat app, the sample size in the data set is small (N=629) and may

not be representative of the general population. Despite the data set having a fair distribution of age groups, with a spread over several countries, it is biased toward highly educated and professional occupations. The data set is also biased toward males in gender distribution. Future research with a larger sample size and a balanced gender distribution could explore correlations, associations, and prediction performance for population subgroups.

Clinical diagnosis of depression was not an inclusion criterion for our sample population. The data set also does not contain a clinical or self-reported baseline assessment of depression and has scarce high depression scores. Because of the crowdsourced rolling recruitment nature of participants, the data set contained an unequal number of repeated depression assessments for all participants. Future research should benefit from replicating the experiment in a clinical population and a more controlled experimental design. With a clinical baseline data set, future research could study the differences in features between depressed and nondepressed groups.

The correlation and association between behavioral patterns extracted from the data set in this study and depression do not necessarily imply causal relationships. For example, the correlation between screen status–normalized entropy and depression may be caused by other confounding variables. In addition, the correlation and association between screen status–normalized entropy and depression are not strong and may not generalize in other populations. Further research is needed to establish the extent to which such behaviors cause or are a consequence of depression.

Lastly, the data set was collected from Android participants only, and the long-term use of the Carat app could influence participants' behavior [38,39]. Research has shown that participant's behavior and sociodemographics may differ between Android platforms and other mobile platforms such as iOS [44,95]. Future research could replicate this study to explore the extent of the differences in the participants' behaviors (ie, digital biomarkers of participants using Android, iOS, and other mobile platforms).

Replicating the findings from this study with additional biomarkers from GPS and wearable sensor data sets and comparing their correlation and biomarker predictive importance will be interesting in future work. There would be a major design implication for depression intervention development if the behavioral markers from screen and internet connectivity achieve similar promising results as biomarkers from GPS and wearable devices. We hypothesize that screen interaction and internet connectivity data sets alone are less privacy intrusive, could better capture behaviors of immobile persons, and people may be more willing to donate such data sets to science. For example, Apple's Screen Time and Google's Digital Wellbeing app are processing and presenting such data to users to inform where and how users spent their time on smartphones.

Conclusions

In summary, this study sought to find whether we can detect changes in human behavior that would be indicative of depression using smartphones. In addition, we sought to find what objective measures of human behavior from smartphones are insightful in understanding depression. Our results established a positive statistically significant linear correlation and association between depression and screen status–normalized entropy behavior quantified from smartphone data sets. Our findings also establish that behavioral markers extracted from smartphone data sets can predict whether or not a participant is depressed based on the PHQ-8 depression score, and that phone screen and internet connectivity behaviors were the most insightful behaviors that influence depression in participants. The findings in this study are supported by previous research findings and contribute to compelling evidence on the utility of digital biomarkers in augmenting traditional assessment of depression, thus enabling continuous and passive monitoring of the complex vectors of depression.

Acknowledgments

This research is supported by the Academy of Finland 6Genesis Flagship (Grant No. 318927), SENSATE (Grant Nos 316253, 320089), Infotech Institute University of Oulu Emerging Project, and Nokia Foundation (Jorma Ollila Grant for EP). We thank the Carat Project for making their data set available for this study, and all participants who contributed to the Carat Project.

Authors' Contributions

KOA, DF, and EP contributed to the conceptualization of the study and bulk data transfers from Carat project archive servers. EP and EL were principally involved in the Carat Project. KOA and JV contributed to data preparation, feature extraction, and machine learning analysis. KOA and YT contributed to the statistical analysis. KOA and DF prepared the original draft. All authors critically reviewed and edited the draft. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of Features Extracted from the Smartphone Dataset.
[PDF File (Adobe PDF File), 36 KB - mhealth_v9i7e26540_app1.pdf]

Multimedia Appendix 2
Extended Results for Statistical Analysis.
[PDF File (Adobe PDF File), 45 KB - mhealth_v9i7e26540_app2.pdf]

Multimedia Appendix 3
Extended Results for Machine Learning Analysis.
[PDF File (Adobe PDF File), 46 KB - mhealth_v9i7e26540_app3.pdf]

References

1. Depression. World Health Organisation (2020). URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2021-06-19]
2. Lépine J, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat 2011 May;7(Suppl 1):3-7 [FREE Full text] [doi: 10.2147/NDT.S19617] [Medline: 21750622]
3. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26 [FREE Full text] [doi: 10.1145/3191775]
4. Perlis ML, Giles DE, Buysse DJ, Tu X, Kupfer DJ. Self-reported sleep disturbance as a prodromal symptom in recurrent depression. J Affect Disord 1997 Feb;42(2-3):209-212. [doi: 10.1016/s0165-0327(96)01411-5] [Medline: 9105962]
5. Hamilton J, Ladouceur C, Silk J, Franzen P, Bylsma L. Higher Rates of Sleep Disturbance Among Offspring of Parents With Recurrent Depression Compared to Offspring of Nondepressed Parents. J Pediatr Psychol 2020 Jan 01;45(1):1-11 [FREE Full text] [doi: 10.1093/jpepsy/jsz079] [Medline: 31633781]
6. Saeb S, Zhang M, Kwasny M, Karr C, Kording K, Mohr D. The relationship between clinical, momentary, and sensor-based assessment of depression. 2015 Presented at: 9th Int Conf Pervasive Comput Technol Healthc PervasiveHealth IEEE; .?232; 2015; Istanbul p. 229. [doi: 10.4108/icst.pervasivehealth.2015.259034]
7. Pressler SJ, Subramanian U, Perkins SM, Gradus-Pizlo I, Kareken D, Kim J, et al. Measuring depressive symptoms in heart failure: validity and reliability of the patient health questionnaire-8. Am J Crit Care 2011 Mar;20(2):146-152 [FREE Full text] [doi: 10.4037/ajcc2010931] [Medline: 20378777]
8. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
9. Hemmerle A, Herman J, Seroogy K. Stress, depression and Parkinson's disease. Exp Neurol 2012 Jan;233(1):79-86 [FREE Full text] [doi: 10.1016/j.expneurol.2011.09.035] [Medline: 22001159]
10. Macchia A, Monte S, Pellegrini F, Romero M, D'Ettorre A, Tavazzi L, et al. Depression worsens outcomes in elderly patients with heart failure: an analysis of 48,117 patients in a community setting. Eur J Heart Fail 2008 Jul 13;10(7):714-721 [FREE Full text] [doi: 10.1016/j.ejheart.2008.05.011] [Medline: 18565789]
11. Depression costs employers $210 billion Internet. Behavioral Health Executive. 2015. URL: https://www.psychcongress.com/news-item/depression-costs-employers-210-billion [accessed 2021-06-28]
12. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. J Affect Disord 2015 Mar 01;172:96-102 [FREE Full text] [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
13. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med 2015 Apr 06;13(1):72 [FREE Full text] [doi: 10.1186/s12916-015-0325-4] [Medline: 25879936]
14. Opoku Asare K, Visuri A, Ferreira D. Towards Early Detection of Depression Through Smartphone Sensing. In: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. New York, NY: ACM; 2019 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; September 09, 2019; London p. 1158-1161. [doi: 10.1145/3341162.3347075]
15. van Berkel N, Ferreira D, Kostakos V. The Experience Sampling Method on Mobile Devices. ACM Comput. Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]
16. Paulhus D. Socially Desirable Responding on Self-Reports. In: Encyclopedia of Personality and Individual Differences. Cham, Switzerland: Springer International Publishing; 2017:1-5.
17. Gorzelitz J, Peppard PE, Malecki K, Gennuso K, Nieto FJ, Cadmus-Bertram L. Predictors of discordance in self-report versus device-measured physical activity measurement. Ann Epidemiol 2018 Jul;28(7):427-431 [FREE Full text] [doi: 10.1016/j.annepidem.2018.03.016] [Medline: 29681429]
18. Spathis D, Servia-Rodriguez S, Farrahi K, Mascolo C, Rentfrow J. Passive mobile sensing and psychological traits for large scale mood prediction. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: ACM; 2019 Presented at: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20, 2019; Trento, Italy p. 272-281. [doi: 10.1145/3329189.3329213]

19. Wagner DT, Rice A, Beresford AR. Device analyzer. SIGMETRICS Perform. Eval. Rev 2014 Apr 17;41(4):53-56. [doi: 10.1145/2627534.2627553]
20. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile Context Instrumentation Framework. Front. ICT 2015 Apr 20;2:6. [doi: 10.3389/fict.2015.00006]
21. Oliner A, Iyer A, Stoica I, Lagerspetz E, Tarkoma S. Carat: Collaborative energy diagnosis for mobile devices. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Nov 11 Presented at: 11th ACM Conference on Embedded Networked Sensor Systems; November 11-15, 2013; Roma, Italy p. 1-14. [doi: 10.1145/2517351.2517354]
22. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY: Association for Computing Machinery; 2014 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13, 2014; Seattle, WA p. 3-14. [doi: 10.1145/2632048.2632054]
23. Coravos A, Khozin S, Mandl KD. Erratum: Author Correction: Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019 May 10;2(1):40 [FREE Full text] [doi: 10.1038/s41746-019-0119-8] [Medline: 31304386]
24. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
25. Rohani D, Tuxen N, Lopategui A, Faurholt-Jepsen M, Kessing L, Bardram J. Personalizing Mental Health: A Feasibility Study of a Mobile Behavioral Activation Tool for Depressed Patients. 2019 Presented at: PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 282-291. [doi: 10.1145/3329189.3329214]
26. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
27. Tseng VW, Sano A, Ben-Zeev D, Brian R, Campbell AT, Hauser M, et al. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. Sci Rep 2020 Sep 15;10(1):15100 [FREE Full text] [doi: 10.1038/s41598-020-71689-1] [Medline: 32934246]
28. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. Front Psychiatry 2021 Jan 28;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
29. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
30. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
31. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile Phone and Wearable Sensor-Based mHealth Approaches for Psychiatric Disorders and Symptoms: Systematic Review. JMIR Ment Health 2019 Mar 20;6(2):e9819 [FREE Full text] [doi: 10.2196/mental.9819] [Medline: 30785404]
32. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2(1):1-9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
33. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The Relationship between Clinical, Momentary, and Sensor-based Assessment of Depression. Int Conf Pervasive Comput Technol Healthc 2015 Aug;17(7):e175 [FREE Full text] [doi: 10.4108/icst.pervasivehealth.2015.259034] [Medline: 26640739]
34. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019 Feb 1;2(1):3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
35. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
36. Malhi GS, Hamilton A, Morris G, Mannie Z, Das P, Outhred T. The promise of digital mood tracking technologies: are we heading on the right track? Evid Based Ment Health 2017 Nov;20(4):102-107. [doi: 10.1136/eb-2017-102757] [Medline: 28855245]
37. Ward NG, Doerr HO, Storrie MC. Skin conductance: a potentially sensitive test for depression. Psychiatry Res 1983 Dec;10(4):295-302. [doi: 10.1016/0165-1781(83)90076-8] [Medline: 6583718]
38. Peltonen E, Lagerspetz E, Hamberg J, Mehrotra A, Musolesi M, Nurmi P, et al. The hidden image of mobile apps: geographic, demographic, and cultural factors in mobile usage. Proc 20th Int Conf Hum-Comput Interact Mob Devices Serv 2018 Sep 03:1-12. [doi: 10.1145/3229434.3229474]

39. Athukorala K, Lagerspetz E, von KM, Jylhä A, Oliner A, Tarkoma S, et al. How carat affects user behavior: implications for mobile battery awareness applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing; 26 April, 2014-1 May, 2014; Toronto, ON p. 1029-1038. [doi: 10.1145/2556288.2557271]
40. Peltonen E, Sharmila P, Opoku Asare K, Visuri A, Lagerspetz E, Ferreira D. When phones get personal: Predicting Big Five personality traits from application usage. Pervasive and Mobile Computing 2020 Nov;69:101269. [doi: 10.1016/j.pmcj.2020.101269]
41. Kroenke K, Strine TW, Spitzer RL, Williams JBW, 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]
42. Hausser J, Strimmer K. entropy: Estimation of Entropy, Mutual Information and Related Quantities. CRAN. 2021 Apr 25. URL: https://CRAN.R-project.org/package=entropy [accessed 2021-05-28]
43. Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, et al. Predicting personality from patterns of behavior collected with smartphones. Proc Natl Acad Sci U S A 2020 Jul 28;117(30):17680-17687 [FREE Full text] [doi: 10.1073/pnas.1920484117] [Medline: 32665436]
44. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, et al. Sensing Behavioral Change over Time. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-21. [doi: 10.1145/3264951]
45. Shannon C. A mathematical theory of communication. Bell System Technical Journal 1948;27(3):379-423. [doi: 10.1002/j.1538-7305.1948.tb01338.x]
46. Gaskin CJ, Happell B. Power of mental health nursing research: a statistical analysis of studies in the International Journal of Mental Health Nursing. Int J Ment Health Nurs 2013 Mar 27;22(1):69-75. [doi: 10.1111/j.1447-0349.2012.00845.x] [Medline: 22738423]
47. Cohen J. Statistical Power Analysis for the Behavioral Sciences (Second Edition). New York, NY: Academic Press; 1988.
48. Enders C. Applied Missing Data Analysis. New York, NY: The Guilford Press; 2010.
49. Larsen R. Missing Data Imputation versus Full Information Maximum Likelihood with Second-Level Dependencies. Structural Equation Modeling: A Multidisciplinary Journal 2011 Oct 05;18(4):649-662. [doi: 10.1080/10705511.2011.607721]
50. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. 2020 Jun 18. URL: https://CRAN.R-project.org/package=psych [accessed 2021-06-28]
51. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 1979;6(2):65-70.
52. Goldstein H. Hierarchical Data Modeling in the Social Sciences. Journal of Educational and Behavioral Statistics 1995;20(2):201. [doi: 10.2307/1165357]
53. Nezlek J. Multilevel modeling for psychologists. In: APA Handbook of Research Methods in Psychology. Washington, D.C: American Psychological Association; 2012:219-241.
54. Kendzor DE, Shuval K, Gabriel KP, Businelle MS, Ma P, High RR, et al. Impact of a Mobile Phone Intervention to Reduce Sedentary Behavior in a Community Sample of Adults: A Quasi-Experimental Evaluation. J Med Internet Res 2016 Jan 25;18(1):e19 [FREE Full text] [doi: 10.2196/jmir.5137] [Medline: 26810027]
55. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
56. Musca SC, Kamiejski R, Nugier A, Méot A, Er-Rafiy A, Brauer M. Data with hierarchical structure: impact of intraclass correlation and sample size on type-I error. Front Psychol 2011;2:74 [FREE Full text] [doi: 10.3389/fpsyg.2011.00074] [Medline: 21687445]
57. Buuren SV, Groothuis-Oudshoorn K. : Multivariate Imputation by Chained Equations in. J. Stat. Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
58. Snijders T, Bosker R. An Introduction to Basic and Advanced Multilevel Modeling. London, UK: SAGE; 2011.
59. Rubin D. Multiple Imputation after 18+ Years. Journal of the American Statistical Association 1996 Jun;91(434):473-489 [FREE Full text] [doi: 10.1080/01621459.1996.10476908]
60. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
61. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann. Statist 2001 Aug 1;29(4):1165-1188. [doi: 10.1214/aos/1013699998]
62. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery; 2016 Presented at: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 13-17 August, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
63. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp. and Comm 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]
64. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Newton, MA: O'Reilly Media, Inc; 2019.

65. Sultana M, Al-Jefri M, Lee J. Using Machine Learning and Smartphone and Smartwatch Data to Detect Emotional States and Transitions: Exploratory Study. JMIR Mhealth Uhealth 2020 Sep 29;8(9):e17818 [FREE Full text] [doi: 10.2196/17818] [Medline: 32990638]
66. Birnbaum ML, Kulkarni P, Van Meter A, Chen V, Rizvi AF, Arenare E, et al. Utilizing Machine Learning on Internet Search Activity to Support the Diagnostic Process and Relapse Detection in Young Individuals With Early Psychosis: Feasibility Study. JMIR Ment Health 2020 Sep 01;7(9):e19348 [FREE Full text] [doi: 10.2196/19348] [Medline: 32870161]
67. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
68. Wshah S, Skalka C, Price M. Predicting Posttraumatic Stress Disorder Risk: A Machine Learning Approach. JMIR Ment Health 2019 Jul 22;6(7):e13946 [FREE Full text] [doi: 10.2196/13946] [Medline: 31333201]
69. Benoit J, Onyeaka H, Keshavan M, Torous J. Systematic Review of Digital Phenotyping and Machine Learning in Psychosis Spectrum Illnesses. Harv Rev Psychiatry 2020;28(5):296-304. [doi: 10.1097/HRP.0000000000000268] [Medline: 32796192]
70. One-Hot Encoding - An Overview. URL: https://www.sciencedirect.com/topics/computer-science/one-hot-encoding [accessed 2021-06-28]
71. Dallora AL, Kvist O, Berglund JS, Ruiz SD, Boldt M, Flodmark C, et al. Chronological Age Assessment in Young Individuals Using Bone Age Assessment Staging and Nonradiological Aspects: Machine Learning Multifactorial Approach. JMIR Med Inform 2020 Sep 21;8(9):e18846 [FREE Full text] [doi: 10.2196/18846] [Medline: 32955457]
72. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014 Mar 29;6(1):10 [FREE Full text] [doi: 10.1186/1758-2946-6-10] [Medline: 24678909]
73. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
74. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced data sets in machine learning. Journal of Machine Learning Research 2017;18(1):559-563. [doi: 10.5555/3122009.3122026]
75. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics 2010 May 15;26(10):1340-1347. [doi: 10.1093/bioinformatics/btq134] [Medline: 20385727]
76. Beware Default Random Forest Importances. URL: https://explained.ai/rf-importance/#6.2 [accessed 2021-06-28]
77. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics 2012 Oct 01;28(19):2520-2522. [doi: 10.1093/bioinformatics/bts480] [Medline: 22908215]
78. Vega J, Li M, Aguillera K, Goel N, Joshl E, Durica K, et al. Reproducible Analysis Pipeline for Data Streams (RAPIDS): Open Source Software to Process Data Collected with Mobile Devices Internet. JMIR Preprints. 2021. URL: https://preprints.jmir.org/preprint/23246 [accessed 2021-06-28]
79. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 2016 Jan;4(2):30 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2015.12.63] [Medline: 26889483]
80. Bates D, Maechler M, Bolker AB, cre, Walker S, Christensen R, et al. lme4: Linear Mixed-Effects Models using 'Eigen' and S4. URL: https://CRAN.R-project.org/package=lme4 [accessed 2021-06-28]
81. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1):e1. [doi: 10.18637/jss.v067.i01]
82. XGBoost Documentation. URL: https://xgboost.readthedocs.io/en/latest/ [accessed 2021-06-28]
83. Gruenerbl A, Osmani V, Bahle G, Carrasco J, Oehler S, Mayora O, et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. In: Proceedings of the 5th Augmented Human International Conference. USA: Association for Computing Machinery; 2014 Presented at: 5th Augmented Human International Conference; 7 March, 2014- 8 March, 2014; Kobe, Japan. [doi: 10.1145/2582051.2582089]
84. Farhan A, Lu J, Bi J, Russell A, Wang B, Bamis A. Multi-view Bi-clustering to Identify Smartphone Sensing Features Indicative of Depression. In: Proceedings of the First IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies. New York, NY: IEEE; 2016 Presented at: First IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies; June 27-29, 2016; Washington, DC. [doi: 10.1109/chase.2016.27]
85. Hovy D. Demographic factors improve classification performance. 2015 Jul Presented at: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (V; July 2015; Beijing, China p. 752-762. [doi: 10.3115/v1/p15-1073]
86. Bonful HA, Anum A. Sociodemographic correlates of depressive symptoms: a cross-sectional analytic study among healthy urban Ghanaian women. BMC Public Health 2019 Jan 10;19(1):50 [FREE Full text] [doi: 10.1186/s12889-018-6322-8] [Medline: 30630448]
87. Gershon A, Thompson WK, Eidelman P, McGlinchey EL, Kaplan KA, Harvey AG. Restless pillow, ruffled mind: sleep and affect coupling in interepisode bipolar disorder. J Abnorm Psychol 2012 Nov;121(4):863-873 [FREE Full text] [doi: 10.1037/a0028233] [Medline: 22845651]

88. Dionisio A, Menezes R, Mendes DA. Mutual information: a measure of dependency for nonlinear time series. Physica A: Statistical Mechanics and its Applications 2004 Dec;344(1-2):326-329. [doi: 10.1016/j.physa.2004.06.144]
89. Jeong J, Gore JC, Peterson BS. Mutual information analysis of the EEG in patients with Alzheimer's disease. Clin Neurophysiol 2001 May;112(5):827-835. [doi: 10.1016/s1388-2457(01)00513-2] [Medline: 11336898]
90. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. Medicina (Kaunas) 2020 Sep 08;56(9):455 [FREE Full text] [doi: 10.3390/medicina56090455] [Medline: 32911665]
91. Obuchi M, Huckins JF, Wang W, daSilva A, Rogers C, Murphy E, et al. Predicting Brain Functional Connectivity Using Mobile Sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-22. [doi: 10.1145/3381001]
92. Connolly CG, Ho TC, Blom EH, LeWinn KZ, Sacchet MD, Tymofiyeva O, et al. Resting-state functional connectivity of the amygdala and longitudinal changes in depression severity in adolescent depression. J Affect Disord 2017 Jan 01;207:86-94 [FREE Full text] [doi: 10.1016/j.jad.2016.09.026] [Medline: 27716542]
93. Alavi SS, Maracy MR, Jannatifard F, Eslami M. The effect of psychiatric symptoms on the internet addiction disorder in Isfahan's University students. J Res Med Sci 2011 Jun;16(6):793-800 [FREE Full text] [Medline: 22091309]
94. Yang S, Fu S, Chen K, Hsieh P, Lin P. Relationships between depression, health-related behaviors, and internet addiction in female junior college students. PLoS One 2019 Aug 9;14(8):e0220784 [FREE Full text] [doi: 10.1371/journal.pone.0220784] [Medline: 31398212]
95. Pryss R, Schobel J, Hoppenstedt B, Spiliopoulou M, Langguth B, Probst T, et al. Ecological Momentary Assessment based Differences between Android and iOS Users of the TrackYourHearing mHealth Crowdsensing Platform. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3951-3955. [doi: 10.1109/EMBC.2019.8857854] [Medline: 31946737]

Abbreviations

AUC: area under the curve
DT: decision tree
ICC: intraclass correlation
KNN: K-nearest neighbor
LMM: linear mixed model
LR: logistic regression
ML: machine learning
PHQ-8: 8-Item Patient Health Questionnaire
RF: random forest
RWC: random weighted classifier
SVM: support vector machine
XGB: XGBoost

Edited by L Buis; submitted 16.12.20; peer-reviewed by A McLean, Y Mao; comments to author 06.02.21; revised version received 15.03.21; accepted 14.05.21; published 12.07.21.

Please cite as:
Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D
Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study
JMIR Mhealth Uhealth 2021;9(7):e26540
URL: https://mhealth.jmir.org/2021/7/e26540
doi:10.2196/26540
PMID:34255713

Original Paper

Ethical Development of Digital Phenotyping Tools for Mental Health Applications: Delphi Study

Nicole Martinez-Martin[1], JD, PhD; Henry T Greely[2], JD; Mildred K Cho[1], PhD

[1]Center for Biomedical Ethics, School of Medicine, Stanford University, Stanford, CA, United States

[2]Stanford Law School, Stanford, CA, United States

Corresponding Author:
Nicole Martinez-Martin, JD, PhD
Center for Biomedical Ethics
School of Medicine
Stanford University
1215 Welch Road
Modular A
Stanford, CA, 94305
United States
Phone: 1 7734585750
Email: nicolemz@stanford.edu

Abstract

Background: Digital phenotyping (also known as *personal sensing*, *intelligent sensing*, or *body computing*) involves the collection of biometric and personal data *in situ* from digital devices, such as smartphones, wearables, or social media, to measure behavior or other health indicators. The collected data are analyzed to generate moment-by-moment quantification of a person's mental state and potentially predict future mental states. Digital phenotyping projects incorporate data from multiple sources, such as electronic health records, biometric scans, or genetic testing. As digital phenotyping tools can be used to study and predict behavior, they are of increasing interest for a range of consumer, government, and health care applications. In clinical care, digital phenotyping is expected to improve mental health diagnoses and treatment. At the same time, mental health applications of digital phenotyping present significant areas of ethical concern, particularly in terms of privacy and data protection, consent, bias, and accountability.

Objective: This study aims to develop consensus statements regarding key areas of ethical guidance for mental health applications of digital phenotyping in the United States.

Methods: We used a modified Delphi technique to identify the emerging ethical challenges posed by digital phenotyping for mental health applications and to formulate guidance for addressing these challenges. Experts in digital phenotyping, data science, mental health, law, and ethics participated as panelists in the study. The panel arrived at consensus recommendations through an iterative process involving interviews and surveys. The panelists focused primarily on clinical applications for digital phenotyping for mental health but also included recommendations regarding transparency and data protection to address potential areas of misuse of digital phenotyping data outside of the health care domain.

Results: The findings of this study showed strong agreement related to these ethical issues in the development of mental health applications of digital phenotyping: privacy, transparency, consent, accountability, and fairness. Consensus regarding the recommendation statements was strongest when the guidance was stated broadly enough to accommodate a range of potential applications. The privacy and data protection issues that the Delphi participants found particularly critical to address related to the perceived inadequacies of current regulations and frameworks for protecting sensitive personal information and the potential for sale and analysis of personal data outside of health systems.

Conclusions: The Delphi study found agreement on a number of ethical issues to prioritize in the development of digital phenotyping for mental health applications. The Delphi consensus statements identified general recommendations and principles regarding the ethical application of digital phenotyping to mental health. As digital phenotyping for mental health is implemented in clinical care, there remains a need for empirical research and consultation with relevant stakeholders to further understand and address relevant issues.

(JMIR Mhealth Uhealth 2021;9(7):e27343) doi:10.2196/27343

KEYWORDS

ethics; neuroethics; digital phenotyping; digital mental health; Delphi study; mental health; machine learning; artificial intelligence; mobile phone

Introduction

Background

Digital phenotyping tools are expected to improve mental health diagnosis and treatment when integrated into clinical care [1-3]. Digital phenotyping presents significant areas of ethical concern, particularly in terms of privacy and data protection, consent, bias, and accountability [4]. For this study, a modified Delphi approach was used to identify recommendations from panelists with relevant expertise (eg, computer science, mental health care, health law, and ethics) for the ethical application of this emerging technology to mental health.

Digital phenotyping refers to new approaches to measure behavior through the collection of biometric and personal data in situ from digital devices, such as smartphones, wearables, or social media. The data are analyzed to generate moment-by-moment quantification of a person's mental state or prediction of their future behavior [5]. For example, data on pulse rate, finger taps, or voice features can be tracked using an individual's smartphone and then analyzed to measure behavior, physiological states, and cognitive functioning [6-9]. As the field of digital phenotyping has evolved, projects increasingly include multiple data streams in the analyses, such as data from electronic health records (EHRs), facial recognition technology, ambient sensors, biological scans, or genomic information [10-13]. The proper terminology for these techniques is still under debate, with terms such as *computational behavioral analysis* [14], *continuous measurement* [15], or *personal sensing* also being applied to similar research approaches that involve continuous monitoring of behavioral data gathered from sensors or digital sources [16,17]. Liang et al [18] suggest a broadened definition of digital phenotyping to incorporate the trends toward using multiple data streams, encompassing intelligent systems that sense and mine information related to mental health states "based on the ubiquitous 'digital footprints' from multiple data sources, e.g., ubiquitous sensors, social media and healthcare systems." The term *digital phenotyping* is used in this paper, in part because it was the term used in the Delphi study. Furthermore, the expanded definition of *digital phenotyping* by Liang et al [18] captures the range of ethical concerns regarding the collection and use of data for digital phenotyping projects addressed by the Delphi participants.

Digital phenotyping has a range of health applications, such as the identification of cardiovascular disease risk [19] or suicidal ideation [20]. However, mental health has been a primary area for the investment and development of this technology. Mental health applications of digital phenotyping include analysis of sleep patterns to predict episodes of relapse in schizophrenia [21], early identification of postpartum depression [22], use of keystroke patterns to predict episodes of mania [23], movement or linguistic analysis to predict episodes of depression [24,25], and social media data used to identify drinking and tobacco abstinence behaviors [26]. Mental health applications have been a primary focus of digital phenotyping projects in part because of the ease with which mobile technology can be used to gather massive amounts of fine-grained behavioral data from the user at any time and in any location [27-29]. These types of data are seen as having great potential to address one of the long-standing difficulties in psychiatric research, namely, the lack of definitive biomarkers or objective physiological measures for reliable psychiatric diagnosis [30,31]. Moreover, the collection of psychiatric data had previously been limited to clinical encounters, making it difficult to gather a complete picture of the day-to-day course of behavioral disorders [32]. The advances in technology for collecting and analyzing behavioral data have been applied toward filling this need for better psychiatric research tools.

The consumer domain and institutions such as the military, employers, insurance organizations, and the criminal justice system have also demonstrated a strong interest in the type of behavioral analyses and predictions offered by digital phenotyping [33-35]. The recommendations of the panel focused primarily on clinical applications because it is the domain in which there is primary investment and publications related to behavioral digital phenotyping [36-39]. We also focused on applications in one country (the United States) in order to facilitate the analysis of regulatory implications based on a limited set of regulations and regulatory frameworks. In the United States, clinical applications are *regulated* technologies [40], in that they are subject to government regulation, such as by the US Food and Drug Administration (FDA) or the privacy rule under the Health Information Portability and Accountability Act (HIPAA). Nonetheless, clinical applications of digital phenotyping and associated data collection practices, as explained in detail below, present challenges for the traditional frameworks used for the regulation of data or medical devices. Furthermore, data collection for clinical digital phenotyping may use consumer devices or take place outside of the regulatory frameworks. The Delphi panelists paid attention to ethical concerns relevant to both regulated and unregulated applications for digital phenotyping, because the traditional ethical and regulatory frameworks may inadequately account for issues such as data protection or oversight in digital phenotyping.

This Delphi study was used to address ethical issues raised by mental health applications of digital phenotyping, such as privacy and data protection, consent, transparency, potential for bias in outcomes, and accountability [4]. Digital phenotyping presents novel concerns because the types of data collection and analytics involved are not adequately addressed under current ethical and regulatory frameworks [41,42]. For example, in the health care domain, the FDA is still evolving in its approach to regulating digital software and algorithms [43]. The HIPAA Privacy Rule provides protection for health information collected in health care systems [44]. However, digital phenotyping has the potential to create sensitive health information outside of contexts covered by HIPAA, such as

information collected by consumer devices or in settings outside of health care, and the Federal Trade Commission can provide oversight regarding deceptive claims or transparency in relation to consumer uses of digital phenotyping. However, the Federal Trade Commission is limited to the scope of its authority to address broader concerns of safety and privacy in digital phenotyping [45].

Digital phenotyping projects may include many forms of data, from social media, location data, and EHRs to screen taps to genomic data and biometric scans, raising concerns regarding the massive volume of data and appropriately addressing the relevant data protection issues [46]. Under HIPAA, health data that contain personal identifiers can only be shared with third parties when it is used for the purposes of treatment, payment, and health care operations and when a business associate agreement is in place [47,48]. In practice, information in EHRs may be accessible to third parties in ways that patients are not expecting [49]. There have also been examples of third-party companies with whom health care data are shared under business associate agreements and inadequate patient records [50]. Deidentified data (data from which 18 specific identifiers, such as name and age, have been removed) may be shared without restriction under HIPAA [51]. At the same time, owing to advances in computing and the availability of large public databases, reidentification of personal data can be accomplished with increasing ease [52,53]. Thus, there is potential for deidentified patient data that are shared with third parties to later be reidentified and used in ways that the patient could not have foreseen or expected [54].

In the current data landscape, the brokerage of personal data and, more specifically, the sale of behavioral and health inferences that can be generated from those data, is a US $200 billion industry [55]. Outside of the health care domain, privacy protection for personal data varies widely according to jurisdiction and type of data. There has been a gradual movement for more jurisdictions to consider the regulation of personal and biometric data, such as the General Data Protection Regulation in the European Union or the California Consumer Privacy Act [56]. Although these regulations provide a useful model for personal data protection, they are not without shortcomings. For example, existing regulations do not address or sufficiently protect individuals from companies and institutions, drawing health inferences from personal data [57,58]. Furthermore, these data or health inferences may be used in ways that have negative ramifications for people, such as higher insurance rates or employment discrimination [59,60]. Adding further concern, some consumer digital mental health services have also been found to use misleading or false claims regarding their collection and use of sensitive personal information [61]. Against this backdrop, even clinical, *regulated* applications of digital phenotyping present significant concerns regarding transparency, consent, and the distribution of risks and benefits for patients and users regarding how their data may be shared and used.

The algorithms used for many digital phenotyping applications, particularly machine learning algorithms, present additional challenges in terms of the regulation and oversight of these tools. With machine learning algorithms, it can be difficult for those reviewing the machine learning tool to be able to evaluate why the data inputs led to a particular output or findings [62]. This *black box* problem, combined with industry concerns for protection of intellectual property, can make it more difficult to detect and address potential systematic problems in the outputs, such as biases in analyses that disproportionately impact different user populations [63,64]. For that reason, efforts have been made to better define and achieve adequate transparency in health algorithms, as well as calls for *explainability* in algorithms [65]. In terms of regulation, the FDA has been shifting its approach to the regulation of digital medical devices. The FDA's Digital Software Precertification program is a relatively recent approach in which companies that are certified as having *a robust culture of quality and organizational excellence* are given a streamlined process for product approval [66]. This type of approach has been criticized for needing more clearly defined standards for *excellence*, as well as insufficiently identifying a process for re-evaluation of products that are in use or accountability for maintaining standards [67]. Gerke et al [68] noted that the FDA and European and US regulations of medical devices have been product-based, and thus need to be further adapted to be able to more effectively address the safety and efficacy concerns that machine learning tools present when placed within a health delivery system. In other words, a systems approach is recommended for the appropriate regulation of algorithmic devices in health care settings.

Bias and fairness are concerns for a range of machine learning and digital health technologies [69,70]. Bias can take a number of forms, including a poor fit between the data collected and the research question being asked, data sets that do not adequately represent the target population, and digital tools that may produce disparate effects when applied to different groups [71,72]. Within digital phenotyping specifically, each of the different types of data streams potentially involved, from social media postings to EHR data, may not adequately include people of different racial, socioeconomic, or disability status [73,74]. Furthermore, data used to develop digital phenotyping tools may reflect social inequalities in ways that are difficult to fully account for and address technological fixes. For example, the data in EHRs may reflect physicians' perceptions and treatment of racialized minorities and associated differential outcomes. There is a need for further research to adequately assess how certain types of digital phenotyping data such as *digital exhaust* may differentially collect information from groups such as people with disabilities or from different racial or cultural groups or different socioeconomic status. Certain predictive uses and applications for digital phenotyping, such as efforts to predict aggression or violence, could be applied in contexts or toward purposes that disproportionately impact marginalized groups. When digital phenotyping tools are not designed or accessible to a range of populations, they can widen gaps in research data or impact mental health diagnosis and treatment in ways that exclude marginalized groups from benefits and even harm those groups [75].

There are a number of efforts underway to address bias in machine learning tools, such as technological fixes to address bias in data sets and algorithms or efforts to provide principles for fairness in algorithmic tools. These are important steps but

are unlikely to fully address the many ways in which social inequities may shape the development and results of digital phenotyping tools [76]. For clinical applications, it is important to note that the FDA does not require data regarding the diversity in training data for machine learning tools. A recent review of machine learning health care devices approved by the FDA found that of 130 tools, most did not report whether they had been evaluated at more than one site, and only 17 included demographic subgroup evaluations in their submissions [77]. The digital divide in digital phenotyping devices could further exacerbate inequities in the distribution of risks and benefits in mental health care.

In clinical and health research settings, consent procedures will need to adequately inform individuals of when and how their data are being gathered and used, as well as whether and how they may receive notice of the findings or repercussions of the digital phenotyping analyses. For digital phenotyping, consent challenges include the difficulty of adequately explaining the probabilistic nature of findings, as well as the potential ramifications from personal data or the inferences that may be drawn from seemingly mundane data such as screen taps or location. The complexity of digital phenotyping findings, as well as the potential ramifications from the data and health inferences generated, can be difficult to convey. Although these consent issues overlap with those applicable to genomic research, some differences are the shorter timeframe for digital phenotyping predictions (eg, risk of a psychotic episode in the next month), more direct responsibility placed on patients to modify their behavior immediately, and the potential for a person's results to be shared or used in domains outside of health care. In addition, there are considerations of appropriate transparency and informed consent for the use of digital phenotyping tools in vulnerable populations, such as children and older adults [78]. As an early intervention in psychiatric conditions generally improves treatment outcomes, mental health research often aims to identify indicators of severe mental illness in early childhood and adolescence [79,80]. Informed consent and transparency procedures for digital phenotyping in children will need to be sensitive to the potential negative impacts of returning predictive results to young people and take into account children's rights to autonomy and parental interest in being informed [81].

The clinical use of digital phenotyping tools is also thought to have the potential to disrupt the traditional patient-therapist relationship. Artificial intelligence tools are thought to have the potential to disrupt or even replace some of the roles traditionally held by therapists or clinicians [82-84]. The use of artificial intelligence methods, such as machine learning and natural language processing, is thought to raise issues of whether the device's findings will be viewed by physicians and patients as more objective than as physician judgment or patient's self-report, thus intruding upon the therapeutic relationship. In instances where a device's recommendations differ from the physician's judgment, there are concerns regarding liability and accountability for any errors in the tool's findings, as well as the nature of the fiduciary relationships involved [85].

Objectives

The Delphi technique is a widely used method for engaging a group of experts to identify and explore a range of approaches to a policy issue, potentially establish areas of convergence and consensus among the recommendations and reveal key assumptions or correlations for different judgments [86,87]. The purpose of this modified Delphi study is to identify priority issues of ethical concern in the development of mental health applications using digital phenotyping and areas of agreement regarding principles for approaching the ethics of digital phenotyping.

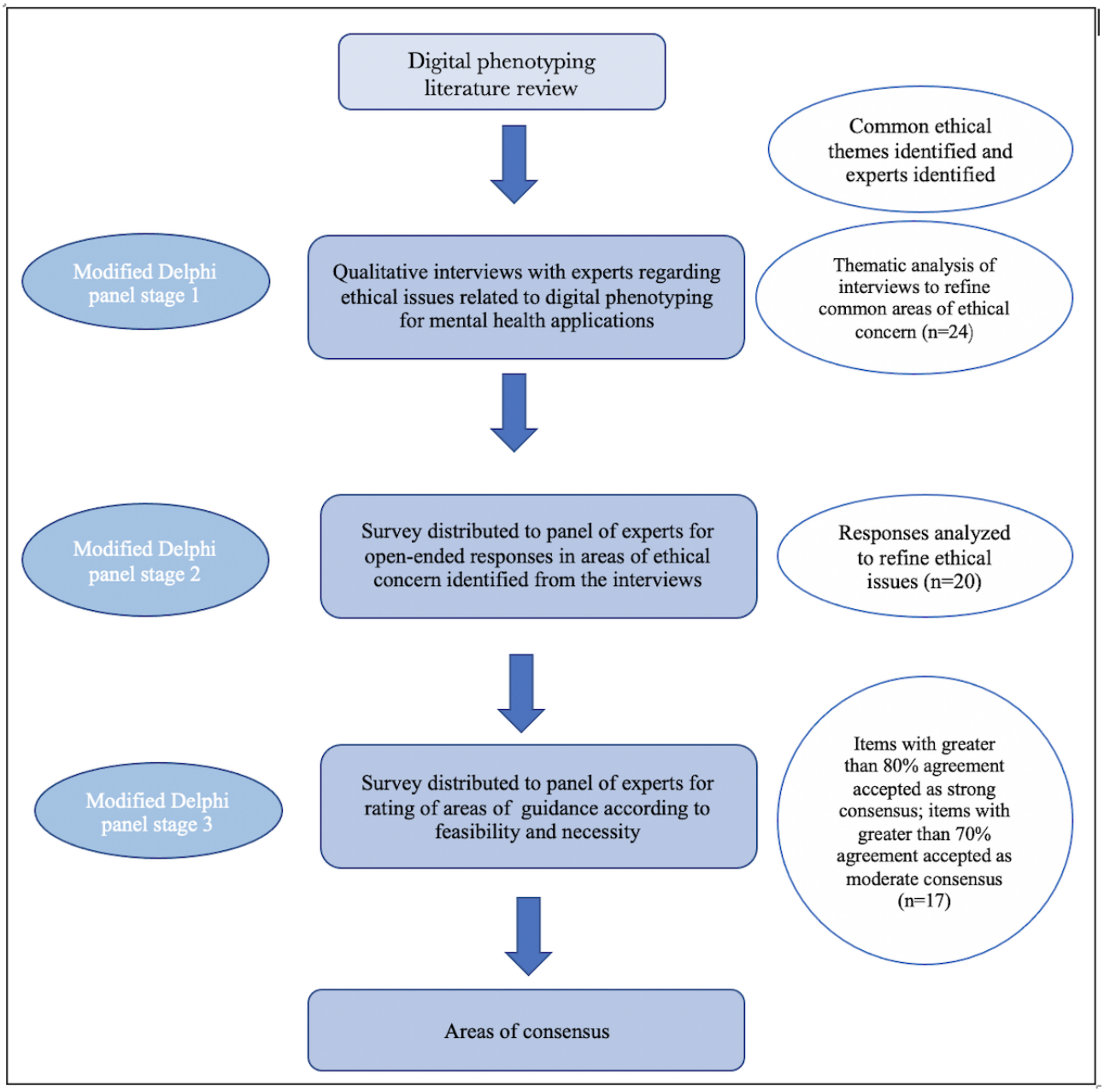
Methods

Overview

The Delphi technique is essentially a method of structuring communication among a group of people with relevant expertise to discuss resolutions to a complex problem [88,89]. Although many modifications to the Delphi technique have evolved over time, the main features of this method include (1) anonymity of the panelists, meaning they do not know of each other's identities or which panelist provided which answers, in order to avoid the influence of status or personality on the discussion; (2) controlled feedback, in which the panelists' answers are given to the study coordinator who then processes and disseminates the resulting information; and (3) an iterative process in which experts are consulted more than once to give them the opportunity to reconsider and refine their views [90]. For this study, the modified Delphi technique was used in the stages depicted in Figure 1. This study was designated as exempt by the local institutional review board.

We recruited experts to represent areas of stakeholder relevance in digital phenotyping: (1) computer science, (2) psychiatry and mental health therapy, (3) law, (4) ethics, and (5) lived mental health experience. The category of people with *lived experience* refers to people who have a diagnosis of mental illness. Inclusion of this area of expertise was meant to provide a fuller perspective on the potential ethical impacts of digital phenotyping [91]. For this category, we also looked for people with some experience in mental health advocacy or policy as a foundation for discussing potential ethical issues, such as privacy or consent, relevant to digital phenotyping.

Figure 1. Delphi study overview.



Composition of the Delphi Panel

We searched the PubMed, Google Scholar, and LexisNexis databases to identify people in industry and academia developing technology in the area of digital phenotyping. Search terms included *digital phenotyping*, *personal sensing*, *computational behavioral analysis*, and *behavioral analytics*. The literature review also yielded specific subareas of expertise relevant to computer science, ethics, and law, relating to emerging technologies, privacy, data protection, machine learning, and bias. Within areas of expertise 1-4, there are also these subareas of expertise represented. For example, in computer science, we included people who worked directly with digital phenotyping as well as those who had related expertise in machine learning, data science, or predictive analytics; within law, we identified people with subspecialties in health, data, and health technology law.

There is no established optimal number of experts for a Delphi panel [92]. Primary factors in deciding on the size of a Delphi panel are appropriate representations of variations in judgment among those with expertise and the drawbacks involved in managing multiple surveys, such as decreasing response rates and increased time needed by researchers in between rounds [93]. Most Delphi studies have used between 15 and 20 participants [94]. In this study, 28 people with relevant expertise were identified through the review process and invited to participate in the qualitative interviews for the first stage of the Delphi study. There were fewer people with lived experience represented on the panel than in the other categories. This reflects a smaller pool of potential panelists that we were able to identify through our search than for other categories. We identified 8 people who were in contact with their participation; 3 people responded to our invitations, and one of those 3 subsequently decided not to proceed with scheduling an interview for personal reasons unrelated to the study itself.

Interview and Survey Stages

The qualitative interviews were semistructured and explored questions regarding participants' views of the ethical issues presented by digital phenotyping for mental health applications. Interview transcripts were reviewed to identify the main ethical themes found in expert panelists' interviews as well as the main areas of their recommendations to address these ethical concerns [95]. The transcripts of the interviews were generated as Microsoft Word documents. An *identifying-category* strategy for reviewing the transcripts and determining the preliminary codes was used [96]. The content of the interviews involved the participants directly referencing ethical categories relevant to health technologies, such as consent or privacy, which facilitated the identification of relevant categories from the transcripts. Once the preliminary categories for the transcript themes were established, we reviewed the transcripts to confirm the categories and identify associated recommendations for addressing the areas of ethical concern. We then used the main ethical themes to generate an open-ended qualitative survey that we distributed among the panelists in which we asked them whether they thought that the identified ethical issue was relevant to digital phenotyping for mental health applications and whether the recommendations to address that issue were appropriate.

Narrative comments from the qualitative survey were used to assist in drafting the statements for the second survey relating to recommendations for ethical mental health applications of digital phenotyping. The second survey was conducted with the same panel of experts who responded to the first survey. For the second survey, we asked panelists to rate statements according to the necessity of a particular recommendation or guidance statement on a four-point scale (1=strong agreement, 2=moderate agreement, 3=neutral, and 4=disagreement). In the second survey, we used a cut-off of 80% rating agreement to indicate strong agreement among panelists, and we deemed 70% moderate agreement with respect to consensus, consistent with the methodology in the Delphi literature [85,97].

Panelists were also asked to rate statements according to feasibility on the same four-point scale. *Feasibility* refers to the likelihood that a particular recommendation could be effectively implemented. During the interview stage, some participants noted that there were some potential recommendations for addressing ethical issues in digital phenotyping that were infeasible. For example, a recommendation for data protection regulation might be identified as desirable but unlikely to be implemented. In some cases, a panelist's specific expertise provided them with additional insight into the feasibility of an option that is different. In Delphi studies applicable to health care, including ratings for both necessity and feasibility, were found to be more useful for identifying recommendations that could be effectively implemented [98]. For these reasons, we assessed both the necessity and feasibility.

Results

Of the 28 invitations, 24 (86%) participated in the qualitative interviews, 20 (71%) participated in the first survey, and 17 (61%) participated in the second survey (Table 1).

Table 1. Expertise represented at each stage[a].

Stage	Computer science, n (%)	Psychiatry or therapy, n (%)	Law or ethics, n (%)	Lived experience, n (%)
Interviews (n=24)	8 (33)	8 (33)	9 (38)	2 (8)
Survey 1 (n=20)	6 (30)	6 (30)	7 (35)	2 (10)
Survey 2 (n=17)	5 (29)	6 (35)	7 (41)	2 (12)

[a]Some panelists had expertise in more than one area.

The main ethical concerns that emerged from the qualitative interviews were (1) privacy and data protection, (2) transparency, (3) consent, (4) reporting of findings or return of results, (5) oversight and accountability, (6) fairness and bias, and (7) validation of digital phenotyping tools. Although panelists identified return of results as a potential area of concern, the specific issues identified overlapped heavily with the types of concerns and recommendations aimed at the consent process for digital phenotyping, such as the need to inform patients of the types of results to expect. The panelists also generally did not go further in providing specific recommendations for the return of results beyond what needed to be discussed in consent, as those particulars were seen to be more dependent on the context of the digital phenotyping application.

In the first survey, the panelists were presented with the ethical categories and then asked to provide additional feedback concerning priority areas of ethical concern within those categories and additional details for recommendations to address those concerns. Those areas of concern and associated recommendations were then presented as statements in the second survey for the panelists to rate. Table 2 presents the results of the Delphi method. The statements in the table present the ethical issues in digital phenotyping for mental health applications resulting from the interviews and first survey. The agreement rating listed in the table represents the level of consensus for statements that were determined through the second survey.

Table 2. Consensus statements on ethics of mental health applications of digital phenotyping.

Statement[a]	Agreement level[b]	
	Necessity	Feasibility
Evidence of validity for the intended use		
Algorithms incorporated into a digital phenotyping tool, especially at a large scale, have to be thoroughly evaluated in terms of performance and accuracy, including false positives and false negatives.	Strong	Strong
Implement processes for review of digital phenotyping tools' effectiveness after implementation, including review of updates, and monitoring and reporting of adverse events caused by an algorithm's findings.	Strong	Moderate
Digital phenotyping tools that are intended for use in health care should use relevant standards for data systems to support the goal of interoperability with existing health data systems.	Strong	Moderate
Digital phenotyping tools for mental health applications should respond to real-world needs and concerns of the intended users, such as clinicians, patients or consumers, in order to enhance user engagement and provide value.	Strong	Strong
Transparency		
Explanations of the processes, risks, limitations, and results that are relevant to different stakeholders should be provided to them in an appropriate format and reading level.	Strong	Strong
Processes involved in the collection, storage, and dissemination of raw data, as well as data processing and the architecture of the algorithms, should be explainable.	Strong	Moderate
Accountability		
Development and use of digital phenotyping tools (eg, plans for data collection or validation) should be reviewed for potential ethical issues by an independent interdisciplinary group with relevant expertise, starting early in the development process.	Strong	Moderate
Provision of appropriate educational and training materials for IRBs[c] handling review of digital phenotyping projects is also necessary.	Strong	Moderate
Consent		
Consent should be required from individuals when their personal data are collected for digital phenotyping tools.	Strong	Moderate
Consent for collection of digital phenotyping data should include information at a sixth-grade level regarding the types of data collected, the inferences that can be drawn from the data, the reports made from the data, who the data and reports would be shared with, the potential risks and benefits to the user, and the limitations that apply to the findings.	Strong	Moderate
Include relevant stakeholders in efforts to formulate and disseminate relevant information for disclosure (eg, data storage, utilizing appropriate languages and formats for relevant stakeholders, such as health care providers, government institutions, advocacy organizations, patients, consumers, or the public).	Strong	Strong
Data security and privacy		
Data and findings that are identifying should not be collected, used or shared with third parties without the informed consent of that individual.	Strong	Strong
Sharing of data to advance scientific research and the validity of the tools remains an important goal.	Strong	Moderate
If data will be shared with third-party researchers, clear information, written at sixth-grade reading level, must be given to the individual user about third-party researcher and how they plan to store, use and/or share the data.	Strong	Strong
The individual user also must have an option to opt out of sharing their data with third parties.	Strong	Moderate
Raw data that is nonidentifying, and nonidentifying summary statistics, may be shared without consent.	Strong	Strong
There should be periodic review to re-evaluate whether identifying information can be drawn from the raw data, particularly when combined with other available data.	Moderate	Moderate
Raw data should always be encrypted when stored or transmitted; potential identifiers in data (eg, phone numbers and IP addresses) should be replaced with surrogates (eg, hashed or encrypted).	Moderate	Moderate

Statement[a]	Agreement level[b]	
	Necessity	Feasibility
Standards and approaches to minimize risk of reidentification of individuals, such as differential privacy measures, should be implemented.	Strong	Moderate
The security standards for data storage, sharing, and use of the individual's data, as well as the process for monitoring compliance with these standards, should be clearly defined and communicated to users of digital phenotyping tools.	Strong	Moderate
Security reviews and audits of data practices should also be implemented.	Strong	Moderate
Fairness		
Encourage collaborative research and partnerships to develop ways to identify and minimize bias or discrimination in the development of digital phenotyping tools and to identify and minimize any potential bias that may occur because of how the tools may be used in different communities or local contexts.	Moderate	Moderate
Conduct research into and implement methods to mitigate bias in different levels of algorithm development, including in the training data, in the algorithmic process or focus, in the transfer of digital phenotyping tools to different contexts, and in the interpretation of digital phenotyping findings.	Strong	Strong
Identify the specific ways that mental health and clinical care may impact the potential for bias in these areas. Periodic review and re-evaluation of the methods for addressing and mitigating bias at the different levels of algorithmic development may be needed.	Strong	Strong

[a]The statements represent the ethical issues in digital phenotyping for mental health applications resulting from the interviews and the first survey.

[b]The agreement rating listed represents the level of consensus for statements that were determined through the second survey.

[c]IRB: institutional review board.

Discussion

Principal Findings

The results of this study showed strong agreement for several ethical issues in the development of digital phenotyping: privacy, transparency, consent, accountability, and fairness. Agreement was strongest when the guidance statements were broad enough to accommodate a range of applications. The panelist comments for the survey indicate that the consensus around broader principles reflects the need to allow flexibility for specific contexts and projects for which digital phenotyping might be used for mental health purposes.

The privacy and data protection issues that Delphi participants found particularly concerning generally related to the perceived inadequacies of current regulations and frameworks for protecting sensitive personal information and the potential for sale and analysis of personal data outside of health systems. Most of the participants noted in the interviews that additional data regulation would most likely be necessary to fully address the privacy concerns posed by digital phenotyping. However, advocating for specific technological standards or regulatory measures was seen as beyond the scope of what the panel could meaningfully address. The panelists focused on addressing general principles for privacy and data protection rather than on specific technological standards or regulatory measures.

Clinical digital phenotyping applications are subject to the security and privacy provisions of HIPAA. Nonetheless, panelists noted that digital phenotyping tools may involve data or be applied outside of contexts for which HIPAA or other personal data protections currently apply. As one panelist stated, "HIPAA criteria don't include new forms of identifiable data like keystroke kinematics - principles and practices need to be more sophisticated to address digital health tech."

Digital phenotyping poses specific concerns regarding privacy because much of the raw data that are collected, such as screen taps or location data, may not be information that patients or users consider sensitive personal information. Thus, patients and users may not be aware of or be able to foresee how that data may be analyzed to reveal information about their mental state that they would want to keep private.

Transparency and consent were seen as key areas for presenting patients and users with information about privacy and data protection. For the clinical use of digital phenotyping, informed consent would need to include careful consideration of how to communicate the risks and benefits and what, how, and when findings would need to be reported afterward. At the same time, as 2 of the panelists noted in the first round of surveys, providing information effectively can be difficult, especially as patients and users feel that there is too much consent information being given to them and feel overwhelmed or prefer to ignore it. Owing to the complexity involved in collecting data, generating results, and understanding downstream health and data implications, the achievability of complete informed consent is arguable. All panelists agreed that stakeholders should be included in collaborative processes to determine what information should be included in the consent and return of digital phenotyping results.

The study found strong agreement regarding the need for consent for the collection and use of raw data. Increasingly, owing to advances in data science and the availability of massive public databases, personal data can be reidentified [99]. Furthermore, health inferences can be generated from seemingly mundane

personal information that can have repercussions for individuals and groups [100]. Existing frameworks under HIPAA distinguish personal health information from deidentified data, with no consent required for the use of deidentified data [101]. However, one panelist stated, "Raw data is not able to be 'non-identifying' and consent should be a norm when using or sharing personal data that has potential health implications."

Another panelist noted, "consent is not needed for analysis of deidentified data by a trusted entity; but public information about the process, including return of aggregate results, is essential."

One panelist stated in feedback, "All information that is deidentified should not be 'fair game' for any uses and disclosures without consent. This is a flaw in the Common Rule. Sensitive and stigmatizing information may be attributable to socially vulnerable groups."

A different panelist noted that "with respect to the use of personal data in digital phenotyping, it will likely require extensive education of the public to increase data and technology literacy. Developing public trust should be a priority and engaging the public as partners in this endeavor is critical and, expensive."

Most panelists noted that digital phenotyping for mental health presented significant privacy challenges outside the clinical domain, especially in terms of consumer applications. Given the lack of sufficient relevant privacy regulation or consent requirements in the consumer domain, the panelists did not address potential consumer consent requirements. However, recommendations for transparency regarding the design and data practices for digital phenotyping projects were viewed as a way to address privacy concerns. The panelists agreed that information regarding the collection, storage, and dissemination of raw data should be available to users. Reports regarding the findings of digital phenotyping tools should also be available to users. Such information would need to be available at an appropriate reading level, such as a sixth-grade reading level for users.

As many institutional review boards may not have members with expertise in data privacy or predictive algorithms, the need for institutional review boards to have access to adequate educational materials was noted. Panelists also agreed that independent ethics review of digital phenotyping was useful but what that means in practice could take different forms, with emphasis being on the need for such reviews to have transparent processes and independence in their judgment.

Although clinical applications of digital phenotyping are subject to FDA oversight of validation and safety, the panel identified some specific concerns regarding validating tools for specific contexts and applications. Standards for evaluating validity, accuracy, and effectiveness for specific uses, as well as the mechanisms for performing these evaluations, are still evolving and vary across different contexts. There was consensus regarding the need to have a mechanism for review or auditing of the validity of digital phenotyping tools beyond their initial deployment, such as evaluating software updates or device uses deployed in new contexts. There was also general agreement regarding the need for a mechanism through which the data processing and architecture of the digital phenotyping algorithms could be available for independent third-party reviews. However, a statement that set out a proposal for a *continual review* of digital phenotyping devices received feedback from several panelists, indicating that it would be too burdensome to have such a requirement. Decisions concerning how often and in what situations to re-evaluate a device would depend on usage contexts and specific projects. Survey feedback also noted that *evaluation* of an algorithm used for digital phenotyping could entail different degrees of thoroughness. Thus, these types of specific details regarding evaluation were seen as the domain of professional organizations to establish appropriate technical standards for the evaluation of specific types of devices. Another panelist stated that *explainability* for the algorithms could be a desirable goal, but it is not something that would be feasible to require currently.

The issue of interoperability posed another area in which panelists agreed upon necessity but not upon feasibility. *Interoperability* refers to the ability of data systems and services to have clear, shared standards for the content, context, and meaning of data [102]. Most panelists viewed the ability of data to be used by different systems as necessary to facilitate scientific research using digital phenotyping data. At the same time, comments made in the survey by panelists with computer science expertise noted that although interoperability is a desirable goal, it has encountered practical challenges for implementation in health data that would make it impracticable to put forth as a requirement [103].

The lack of diversity in research participants and the data used for research, such as lack of panelists according to race, gender, or disability, presents concerns for ensuring equity and fairness in digital phenotyping for mental health. The potential for bias needs to be addressed at the different stages of the development process for digital phenotyping tools, from how the initial research questions are formulated, how data are selected and used within these stages, and the potential for disparities resulting from implementation of these tools in different contexts. In particular, practices during the design and development processes are needed to ensure that digital phenotyping tools can be used in different communities and contexts while mitigating potential harm to populations, such as marginalized racial, linguistic, or socioeconomic groups. There was strong agreement regarding the need to address bias and fairness; however, as one panelist stated, "To assess feasibility, context is important and - depending on context, there will be unique barriers and facilitators to implementation." Another panelist noted that it is "[h]ard to predict where bias might arise, thus this is challenging work, requiring constant vigilance." Organizations such as the American Medical Informatics Association have been working on specific standards and principles for addressing bias and fairness in algorithms [104-106]. The Delphi panel identified some practices in the development of digital phenotyping that can be useful in identifying areas of potential bias, such as having diverse research teams and engagement of key stakeholders at different stages of the development process.

Limitations

Although this study met the stated recommendations regarding the size of the Delphi panel and selection of experts, given the size of the panel, some relevant viewpoints might not have been included in the panel. As noted in the *Methods* section, despite efforts to recruit additional panelists who have lived mental health experience, we had a notably smaller number of panelists in that category and thus did not have the benefit of additional insights from that perspective.

The digital phenotyping literature review raised several areas of ethical concern that were not directly engaged by the Delphi panel in the consensus statements. For example, concerns regarding the potential impact of digital phenotyping on the therapeutic alliance or the impact of continuous monitoring on the experience of patients and participants were not addressed in the consensus statements. This Delphi approach was not intended to comprehensively address all of the potential ethical concerns regarding mental health applications of digital phenotyping. The Delphi process served to identify priority areas of ethical concern for an emerging technology. The consensus aspects of the approach meant that there were relevant ethical issues that did not ultimately be prioritized for inclusion in the recommendations. Nonetheless, excluded ethical concerns, such as impact on the therapeutic relationship, remain relevant and merit scrutiny and empirical research as mental health applications of digital phenotyping become more common.

Conclusions

This Delphi study found agreement on a number of ethical issues to prioritize in the development of digital phenotyping for mental health applications. Standards and guidelines for key areas of digital phenotyping, such as privacy and data protection outside of health care institutions and the regulation of digital medical devices, are still evolving. The Delphi consensus statements identified general recommendations and principles regarding the ethical application of digital phenotyping to mental health. As digital phenotyping for mental health is implemented in clinical care, there remains a need for empirical research and consultation with relevant stakeholders to further understand and address relevant ethical issues.

Acknowledgments

This study was funded by a Greenwall Foundation Making a Difference in Real-World Bioethics grant. The work of NMM was also supported by NIH/NIMH grant K01 MH118375-01A1.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the conception and design of the study. Data collection and paper preparation was performed by NMM. Data analysis and interpretation were done by NMM and MKC. All authors edited the paper and provided their final approval.

Conflicts of Interest

None declared.

References

1. 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]
2. Spinazze P, Rykov Y, Bottle A, Car J. Digital phenotyping for assessment and prediction of mental health outcomes: a scoping review protocol. BMJ Open 2019 Dec 30;9(12):e032255 [FREE Full text] [doi: 10.1136/bmjopen-2019-032255] [Medline: 31892655]
3. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat 2012 May;33(5):777-780. [doi: 10.1002/humu.22080] [Medline: 22504886]
4. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med 2018;1:- [FREE Full text] [doi: 10.1038/s41746-018-0075-8] [Medline: 31211249]
5. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
6. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 7;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
7. Dagum P. Digital biomarkers of cognitive function. NPJ Digit Med 2018;1:10 [FREE Full text] [doi: 10.1038/s41746-018-0018-4] [Medline: 31304295]
8. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
9. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019 Jun 1;26(6):561-576 [FREE Full text] [doi: 10.1093/jamia/ocz009] [Medline: 30908576]

10. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166. [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]
11. Hsu M, Ahern DK, Suzuki J. Digital phenotyping to enhance substance use treatment during the covid-19 pandemic. JMIR Ment Health 2020 Oct 26;7(10):e21814 [FREE Full text] [doi: 10.2196/21814] [Medline: 33031044]
12. Barnett S, Huckvale K, Christensen H, Venkatesh S, Mouzakis K, Vasa R. Intelligent sensing to inform and learn (INSTIL): a scalable and governance-aware platform for universal, smartphone-based digital phenotyping for research and clinical applications. J Med Internet Res 2019 Nov 6;21(11):e16399 [FREE Full text] [doi: 10.2196/16399] [Medline: 31692450]
13. Sariyska R, Rathner E, Baumeister H, Montag C. Feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones. Front Neurosci 2018 Dec 18;12:945-946 [FREE Full text] [doi: 10.3389/fnins.2018.00945] [Medline: 30618574]
14. Corcoran CM, Cecchi GA. Computational approaches to behavior analysis in psychiatry. Neuropsychopharmacology 2018 Jan;43(1):225-226 [FREE Full text] [doi: 10.1038/npp.2017.188] [Medline: 29192659]
15. Saxon L, DiPaula B, Fox GR, Ebert R, Duhaime J, Nocera L, et al. Continuous measurement of reconnaissance marines in training with custom smartphone app and watch: observational cohort study. JMIR Mhealth Uhealth 2020 Jun 15;8(6):e14116 [FREE Full text] [doi: 10.2196/14116] [Medline: 32348252]
16. 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 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
17. Mohr DC, Shilton K, Hotopf M. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. NPJ Digit Med 2020 Mar 25;3(1):45-42 [FREE Full text] [doi: 10.1038/s41746-020-0251-5] [Medline: 32219186]
18. Liang Y, Zheng X, Zeng DD. A survey on big data-driven digital phenotyping of mental health. Information Fusion 2019 Dec;52:290-307. [doi: 10.1016/j.inffus.2019.04.001]
19. Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
20. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018 Jul;35(7):601-608. [doi: 10.1002/da.22730] [Medline: 29637663]
21. Meyer N, Joyce D, Karr C. P029 Sleep and circadian rhythm disturbances and relapse in schizophrenia: a digital phenotyping study. BMJ Open Resp Res 2019:6. [doi: 10.1136/bmjresp-2019-bssconf.29]
22. Hahn L, Eickhoff SB, Habel U, Stickeler E, Schnakenberg P, Goecke TW, et al. Early identification of postpartum depression using demographic, clinical, and digital phenotyping. Transl Psychiatry 2021 Feb 11;11(1):121 [FREE Full text] [doi: 10.1038/s41398-021-01245-6] [Medline: 33574229]
23. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
24. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
25. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. J Nerv Ment Dis 2019;207(10):893-896. [doi: 10.1097/nmd.0000000000001042]
26. Tamersoy A, de Choudhury M, Chau DH. Characterizing smoking and drinking abstinence from social media. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:139-148 [FREE Full text] [doi: 10.1145/2700171.2791247] [Medline: 26640831]
27. Lydon-Staley DM, Barnett I, Satterthwaite TD, Bassett DS. Digital phenotyping for psychiatry: accommodating data and theory with network science methodologies. Curr Opin Biomed Eng 2019 Mar;9:8-13 [FREE Full text] [doi: 10.1016/j.cobme.2018.12.003] [Medline: 31650093]
28. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]
29. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
30. García-Gutiérrez MS, Navarrete F, Sala F, Gasparyan A, Austrich-Olivares A, Manzanares J. Biomarkers in psychiatry: concept, definition, types and relevance to the clinical reality. Front Psychiatry 2020;11:432 [FREE Full text] [doi: 10.3389/fpsyt.2020.00432] [Medline: 32499729]
31. Boksa P. A way forward for research on biomarkers for psychiatric disorders. J Psychiatry Neurosci 2013 Mar;38(2):75-77 [FREE Full text] [doi: 10.1503/jpn.130018] [Medline: 23422052]
32. Chauvin JJ, Insel TR. Building the thermometer for mental health. Cerebrum 2018;2018:- [FREE Full text] [Medline: 30746032]
33. Zhong H, Xiao J. Big data analytics on customer behaviors with kinect sensor network. Int J Recent Trends Hum Comput Interac 2015 Mar;6(2):36-47 [FREE Full text]

34. Prudentov RV. Private life and surveillance in a digital era: human rights in European perspective. DLJ 2020 Aug 26;1(2):41-52 [FREE Full text] [doi: 10.38044/2686-9136-2020-1-2-41-52]
35. Parshley L. Our Phones Can Now Detect Health Problems From Parkinson's to Depression: Is That a Good Thing? Vox. 2020 Feb 12. URL: https://www.vox.com/the-highlight/2020/2/5/21056921/phones-health-apps-data-digital-phenotyping [accessed 2021-03-21]
36. Cohen AS, Schwartz E, Le T, Cowan T, Cox C, Tucker R, et al. Validating digital phenotyping technologies for clinical use: the critical importance of 'resolution'. World Psychiatry 2020 Feb 10;19(1):114-115 [FREE Full text] [doi: 10.1002/wps.20703] [Medline: 31922662]
37. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
38. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. Physiol Meas 2018 May 15;39(5):05TR01 [FREE Full text] [doi: 10.1088/1361-6579/aabf64] [Medline: 29671754]
39. Vaidyam A, Halamka J, Torous J. Actionable digital phenotyping: a framework for the delivery of just-in-time and longitudinal interventions in clinical healthcare. Mhealth 2019 Aug;5:25 [FREE Full text] [doi: 10.21037/mhealth.2019.07.04] [Medline: 31559270]
40. Rothstein MA, Wilbanks JT, Beskow LM, Brelsford KM, Brothers KB, Doerr M, et al. Unregulated health research using mobile devices: ethical considerations and policy recommendations. J Law Med Ethics 2020 Mar;48(1_suppl):196-226. [doi: 10.1177/1073110520917047] [Medline: 32342752]
41. Perez-Pozuelo I, Spathis D, Gifford-Moore J, Morley J, Cowls J. Digital phenotyping and sensitive health data: Implications for data governance. J Am Med Inform Assoc 2021 Feb 27:-. [doi: 10.1093/jamia/ocab012] [Medline: 33647989]
42. Stanghellini G, Leoni F. Digital phenotyping: ethical issues, opportunities, and threats. Front Psychiatry 2020 May 27;11:473 [FREE Full text] [doi: 10.3389/fpsyt.2020.00473] [Medline: 32536882]
43. Cortez N. The evolving law and ethics of digital health. In: Rivas H, Wac K, editors. Digital Health: Scaling Healthcare to the World. New York City, USA: Springer International Publishing; 2018:249-269.
44. 45 CFR Part 160 and Subparts A and E of Part 164. Health Insurance Portability and Accountability Act, Privacy Rule. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [accessed 2021-02-01]
45. Dasgupta I. Developments in neuroethics and bioethics. Vol 3. Ethical dimensions of commercial and DIY neurotechnologies. In: Bard I, Hildt E, editors. Assessing Current Mechanisms for the Regulation of Direct-to-consumer Neurotechnology. Amsterdam, Netherlands: Elsevier Academic Press; 2020:233-265.
46. Keller AS, Ball TM, Williams LM. Deep phenotyping of attention impairments and the 'inattention biotype' in major depressive disorder. Psychol Med 2020 Oct;50(13):2203-2212 [FREE Full text] [doi: 10.1017/S0033291719002290] [Medline: 31477195]
47. Perakslis ED. Cybersecurity in health care. N Engl J Med 2014 Jul 31;371(5):395-397. [doi: 10.1056/NEJMp1404358] [Medline: 25075831]
48. Puranik M. Are Current HIPAA Regulations Enough To Protect Sensitive Data? Forbes. 2019 Mar 26. URL: https://www.forbes.com/sites/forbestechcouncil/2019/03/26/are-current-hipaa-regulations-enough-to-protect-sensitive-data/ [accessed 2021-03-21]
49. Rothstein M. Debate over patient privacy control in electronic health records. Hastings Center. 2011. URL: https://papers.ssrn.com/abstract=1764002 [accessed 2021-07-13]
50. Becker J. Insufficient Protections for Health Data Privacy: Lessons from Dinerstein v Google. Bill of Health. 2020 Sep 28. URL: http://blog.petrieflom.law.harvard.edu/2020/09/28/dinerstein-google-health-data-privacy/ [accessed 2021-05-05]
51. De-identification of Protected Health Information: How to Anonymize PHI. HIPAA Journal. 2017 Oct 18. URL: https://www.hipaajournal.com/de-identification-protected-health-information/ [accessed 2021-04-05]
52. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]
53. Yoo J, Thaler A, Sweeney L, Zang J. Risks to Patient Privacy: A Re-identification of Patients in Maine and Vermont Statewide Hospital Data. Technology Science. 2018 Oct. URL: https://techscience.org/a/2018100901/#Citation [accessed 2021-07-13]
54. Culnane C, Rubinstein B, Teague V. Health Data in an Open World. arXivcs. 2017 Dec 17. URL: http://arxiv.org/abs/1712.05627 [accessed 2021-07-13]
55. Crain M. The limits of transparency: data brokers and commodification. New Media Soc 2016 Jul 7;20(1):88-104. [doi: 10.1177/1461444816657096]
56. 100 - 1798. California Consumer Privacy Act of 2018. URL: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5 [accessed 2020-12-01]
57. Skiljic A. The Status Quo of Health Data Inferences. International Association of Privacy Professionals: Privacy Perspectives. URL: https://iapp.org/news/a/the-status-quo-of-health-data-inferences/ [accessed 2021-04-02]
58. Kröger J, Raschke P, Bhuiyan T. Privacy Implications of Accelerometer Data: a Review of Possible Inferences. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy. 2019 Presented at: CSP'19; January 19-21, 2019; Kuala Lumpur, Malaysia. [doi: 10.1145/3309074.3309076]

59. Allen M. Health Insurers Are Vacuuming Up Details About You And It Could Raise Your Rates. ProPublica. 2018 Jul 17. URL: https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates [accessed 2021-05-05]
60. Wachter S. Data protection in the age of big data. Nat Electron 2019 Jan 16;2(1):6-7. [doi: 10.1038/s41928-018-0193-y]
61. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 5;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
62. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating ai and practical implications. Yearb Med Inform 2019 Aug;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
63. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar;28(3):231-237 [FREE Full text] [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
64. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
65. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
66. Digital Health Software Precertification (Pre-Cert) Program. US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program [accessed 2021-04-21]
67. Warren E, Murray P, Smith T. Letter to FDA on Regulation of Software as Medical Device. US Senator Elizabeth Warren of Massachusetts. 2018. URL: https://www.warren.senate.gov/oversight/letters/warren-murray-smith-press-fda-on-oversight-of-digital-health-devices [accessed 2021-04-05]
68. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. NPJ Digit Med 2020 Apr 7;3(1):53-54 [FREE Full text] [doi: 10.1038/s41746-020-0262-2] [Medline: 32285013]
69. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
70. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artif Intell Healthcare 2020:295-336. [doi: 10.1016/b978-0-12-818438-7.00012-5]
71. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med 2018 Mar 15;378(11):981-983 [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]
72. Binns R. Fairness in Machine Learning: Lessons from Political Philosophy. ArXiv. 2017 Dec 10. URL: http://arxiv.org/abs/1712.03586 [accessed 2019-04-15]
73. Kaushal A, Altman R, Langlotz C. Geographic distribution of us cohorts used to train deep learning algorithms. J Am Med Assoc 2020 Sep 22;324(12):1212-1213 [FREE Full text] [doi: 10.1001/jama.2020.12067] [Medline: 32960230]
74. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc 2020 Dec 9;27(12):2020-2023 [FREE Full text] [doi: 10.1093/jamia/ocaa094] [Medline: 32574353]
75. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 2016 Apr;22(2):303-341. [doi: 10.1007/s11948-015-9652-2] [Medline: 26002496]
76. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell 2019 Nov 4;1(11):501-507. [doi: 10.1038/s42256-019-0114-4]
77. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]
78. Roberts LW, Warner TD, Brody JL. Perspectives of patients with schizophrenia and psychiatrists regarding ethically important aspects of research participation. Am J Psychiatry 2000 Jan;157(1):67-74. [doi: 10.1176/ajp.157.1.67] [Medline: 10618015]
79. Sequeira L, Battaglia M, Perrotta S, Merikangas K, Strauss J. Digital phenotyping with mobile and wearable devices: advanced symptom measurement in child and adolescent depression. J Am Acad Child Adolesc Psychiatry 2019 Sep;58(9):841-845. [doi: 10.1016/j.jaac.2019.04.011] [Medline: 31445619]
80. Thomson A, Griffiths H, Fisher R, McCabe R, Abbott-Smith S, Schwannauer M. Treatment outcomes and associations in an adolescent-specific early intervention for psychosis service. Early Interv Psychiatry 2019 Jun;13(3):707-714. [doi: 10.1111/eip.12778] [Medline: 30690896]
81. Sundby A, Boolsen MW, Burgdorf KS, Ullum H, Hansen TF, Mors O. Attitudes of stakeholders in psychiatry towards the inclusion of children in genomic research. Hum Genomics 2018 Mar 5;12(1):12 [FREE Full text] [doi: 10.1186/s40246-018-0144-8] [Medline: 29506557]
82. Martinez-Martin N, Dunn LB, Roberts LW. Is it ethical to use prognostic estimates from machine learning to treat psychosis? AMA J Ethics 2018 Sep 1;20(9):E804-E811. [doi: 10.1001/amajethics.2018.804] [Medline: 30242810]

83. Potier R. The digital phenotyping project: a psychoanalytical and network theory perspective. Front Psychol 2020;11:1218 [FREE Full text] [doi: 10.3389/fpsyg.2020.01218] [Medline: 32760307]
84. Dagum P, Montag C. Ethical considerations of digital phenotyping from the perspective of a healthcare practitioner. In: Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Studies in Neuroscience, Psychology and Behavioral Economics. New York City, USA: Springer International Publishing; 2019:13-28.
85. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014 Jul;33(7):1139-1147. [doi: 10.1377/hlthaff.2014.0048] [Medline: 25006139]
86. Dalkey N, Helmer O. An experimental application of the delphi method to the use of experts. Manag Sci 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]
87. Delbecq A, Ven A, Gustafson D. Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Middleton, Wisconsin, United States: Green Briar Press; 1986.
88. Hohmann E, Brand JC, Rossi MJ, Lubowitz JH. Expert opinion is necessary: delphi panel methodology facilitates a scientific approach to consensus. Arthroscopy 2018 Feb;34(2):349-351. [doi: 10.1016/j.arthro.2017.11.022] [Medline: 29413182]
89. Eubank BH, Mohtadi NG, Lafave MR, Wiley JP, Bois AJ, Boorman RS, et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. BMC Med Res Methodol 2016 May 20;16:56 [FREE Full text] [doi: 10.1186/s12874-016-0165-8] [Medline: 27206853]
90. Landeta J. Current validity of the Delphi method in social sciences. Technol Forecast Soc Change 2006 Jun;73(5):467-482. [doi: 10.1016/j.techfore.2005.09.002]
91. Vojtila L, Ashfaq I, Ampofo A, Dawson D, Selby P. Engaging a person with lived experience of mental illness in a collaborative care model feasibility study. Res Involv Engagem 2021 Jan 8;7(1):5 [FREE Full text] [doi: 10.1186/s40900-020-00247-w] [Medline: 33419484]
92. Witkin BR, Altschuld JW. Planning and Conducting Needs Assessments: A Practical Guide. Thousand Oaks, California: Sage Publications; 1995.
93. Hsu CC, Sandford B. The Delphi Technique: Making Sense of Consensus. Pract Assess Res Eval 2007;12:1-8 [FREE Full text]
94. Ludwig B. Predicting the future: have you considered using the delphi methodology? J Exten 1997;35(5):1-5 [FREE Full text]
95. Castleberry A, Nolen A. Thematic analysis of qualitative research data: Is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
96. Vaughn P, Turner C. Decoding via coding: analyzing qualitative text data through thematic coding and survey methodologies. Journal of Library Administration 2015 Nov 23;56(1):41-51. [doi: 10.1080/01930826.2015.1105035]
97. Nadir N, Hart D, Cassara M, Noelker J, Moadel T, Kulkarni M, et al. Simulation-based remediation in emergency medicine residency training: a consensus study. West J Emerg Med 2019 Jan;20(1):145-156 [FREE Full text] [doi: 10.5811/westjem.2018.10.39781] [Medline: 30643618]
98. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
99. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul 23;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]
100. Wachter S, Mittelstadt B. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Business Law Review. 2019. URL: https://papers.ssrn.com/abstract=3248829 [accessed 2021-07-13]
101. Snell E. De-Identification of Data: Breaking Down HIPAA Rules. HealthITSecurity. 2018. URL: https://healthitsecurity.com/news/de-identifcation-of-data-breaking-down-hipaa-rules [accessed 2018-12-11]
102. Kadadi A, Agrawal R, Nyamful C, Atiq R. Challenges of Data Integration and Interoperability in Big Data. In: IEEE International Conference on Big Data. 2014 Presented at: CBD'14; July 17-19, 2014; Washington, DC. [doi: 10.1109/BigData.2014.7004486]
103. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Sep;42(9):572-575 [FREE Full text] [Medline: 28890644]
104. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. Front Artif Intell 2020;3:561802 [FREE Full text] [doi: 10.3389/frai.2020.561802] [Medline: 33981989]
105. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. arXivcs. 2019 Aug 23. URL: http://arxiv.org/abs/1908.09635 [accessed 2021-07-13]
106. McCradden MD, Joshi S, Anderson JA, Mazwi M, Goldenberg A, Zlotnik Shaul R. Patient safety and quality improvement: ethical principles for a regulatory approach to bias in healthcare machine learning. J Am Med Inform Assoc 2020 Dec 9;27(12):2024-2027 [FREE Full text] [doi: 10.1093/jamia/ocaa085] [Medline: 32585698]

Abbreviations

EHR: electronic health record
FDA: Food and Drug Administration
HIPAA: Health Information Portability and Accountability Act

Edited by L Buis; submitted 22.01.21; peer-reviewed by M Himelein-Wachowiak, P Gooding, M Doerr; comments to author 12.03.21; revised version received 06.05.21; accepted 21.05.21; published 28.07.21.

Please cite as:
Martinez-Martin N, Greely HT, Cho MK
Ethical Development of Digital Phenotyping Tools for Mental Health Applications: Delphi Study
JMIR Mhealth Uhealth 2021;9(7):e27343
URL: https://mhealth.jmir.org/2021/7/e27343
doi:10.2196/27343
PMID:34319252

Original Paper

Predicting Depressive Symptom Severity Through Individuals' Nearby Bluetooth Device Count Data Collected by Mobile Phones: Preliminary Longitudinal Study

Yuezhou Zhang[1], MSc; Amos A Folarin[1,2,3,4,5], PhD; Shaoxiong Sun[1], PhD; Nicholas Cummins[1], PhD; Yatharth Ranjan[1], MSc; Zulqarnain Rashid[1], PhD; Pauline Conde[1], BSc; Callum Stewart[1], MSc; Petroula Laiou[1], PhD; Faith Matcham[6], PhD; Carolin Oetzmann[6], MSc; Femke Lamers[7], PhD; Sara Siddi[8,9,10], PhD; Sara Simblett[11], PhD; Aki Rintala[12,13], MSc; David C Mohr[14], PhD; Inez Myin-Germeys[12], PhD; Til Wykes[3,11], PhD; Josep Maria Haro[8,9,10], MD; Brenda W J H Penninx[7], PhD; Vaibhav A Narayan[15], PhD; Peter Annas[16], PhD; Matthew Hotopf[3,6], PhD; Richard J B Dobson[1,2,3,4,5], PhD; RADAR-CNS Consortium

[1]Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]Institute of Health Informatics, University College London, London, United Kingdom

[3]NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom

[4]Health Data Research UK London, University College London, London, United Kingdom

[5]NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, United Kingdom

[6]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[7]Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam University Medical Centre, Vrije Universiteit and GGZ InGeest, Amsterdam, Netherlands

[8]Teaching Research and Innovation Unit, Parc Sanitari Sant Joan de Déu, Fundació Sant Joan de Déu, Barcelona, Spain

[9]Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain

[10]Faculty of Medicine and Health Sciences, Universitat de Barcelona, Barcelona, Spain

[11]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[12]Center for Contextual Psychiatry, Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium

[13]Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lahti, Finland

[14]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University, Evanston, IL, United States

[15]Janssen Research and Development LLC, Titusville, NJ, United States

[16]H Lundbeck A/S, Copenhagen, Denmark

Corresponding Author:
Richard J B Dobson, PhD
Department of Biostatistics & Health Informatics
Institute of Psychiatry, Psychology and Neuroscience
King's College London
SGDP Centre, IoPPN, Box PO 80
De Crespigny Park, Denmark Hill
London, SE5 8AF
United Kingdom
Phone: 44 20 7848 0473
Email: richard.j.dobson@kcl.ac.uk

Abstract

Background: Research in mental health has found associations between depression and individuals' behaviors and statuses, such as social connections and interactions, working status, mobility, and social isolation and loneliness. These behaviors and statuses can be approximated by the nearby Bluetooth device count (NBDC) detected by Bluetooth sensors in mobile phones.

Objective: This study aimed to explore the value of the NBDC data in predicting depressive symptom severity as measured via the 8-item Patient Health Questionnaire (PHQ-8).

Methods: The data used in this paper included 2886 biweekly PHQ-8 records collected from 316 participants recruited from three study sites in the Netherlands, Spain, and the United Kingdom as part of the EU Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) study. From the NBDC data 2 weeks prior to each PHQ-8 score, we extracted 49 Bluetooth features, including statistical features and nonlinear features for measuring the periodicity and regularity of individuals' life rhythms. Linear mixed-effect models were used to explore associations between Bluetooth features and the PHQ-8 score. We then applied hierarchical Bayesian linear regression models to predict the PHQ-8 score from the extracted Bluetooth features.

Results: A number of significant associations were found between Bluetooth features and depressive symptom severity. Generally speaking, along with depressive symptom worsening, one or more of the following changes were found in the preceding 2 weeks of the NBDC data: (1) the amount decreased, (2) the variance decreased, (3) the periodicity (especially the circadian rhythm) decreased, and (4) the NBDC sequence became more irregular. Compared with commonly used machine learning models, the proposed hierarchical Bayesian linear regression model achieved the best prediction metrics (R^2=0.526) and a root mean squared error (RMSE) of 3.891. Bluetooth features can explain an extra 18.8% of the variance in the PHQ-8 score relative to the baseline model without Bluetooth features (R^2=0.338, RMSE=4.547).

Conclusions: Our statistical results indicate that the NBDC data have the potential to reflect changes in individuals' behaviors and statuses concurrent with the changes in the depressive state. The prediction results demonstrate that the NBDC data have a significant value in predicting depressive symptom severity. These findings may have utility for the mental health monitoring practice in real-world settings.

(JMIR Mhealth Uhealth 2021;9(7):e29840) doi:10.2196/29840

KEYWORDS

mental health; depression; digital biomarkers; digital phenotyping; digital health; Bluetooth; hierarchical Bayesian model; mobile health; mHealth; monitoring

Introduction

Existing studies have demonstrated that depression is significantly associated with individuals' behaviors and statuses, such as social connections and interactions, working status, mobility, and social isolation and loneliness [1-4]. For example, individuals reporting fewer social network connections or less social support tend to have higher depressive symptomatology [1]. As the depressive mood and medical comorbidity can make people unable to work, the unemployment rate in depression is high [2]. Reduced mobility and physical activity are associated with depressive symptoms [3]. Loneliness is a specific risk factor for depression, and a significant proportion of suicides have a history of social isolation [1,4]. Although these findings have been replicated in different populations, these studies relied on participant self-report, which is susceptible to recall bias and typically does not capture dynamic information [5].

Mobile phone technology provides an unobtrusive, continuous, and cost-efficient means to capture individuals' daily behaviors and statuses using a number of embedded sensors, such as accelerometers, GPS sensors, and Bluetooth sensors [6]. The embedded Bluetooth sensor can be used to record individuals' local proximity information, such as the nearby Bluetooth device count (NBDC) that includes the Bluetooth signal of other phone users [7]. The continuously recorded NBDC data represents a mixed signal that has been used to estimate individuals' behaviors and statuses, including face-to-face social interactions [8-10], working status [11], mobility [12], and isolation and loneliness [13,14]. Therefore, the NBDC data have the potential to reflect changes in people's behaviors and statuses during the depressive state.

There have been a few studies exploring the relationship between the NBDC data and depression directly. Wang et al found a negative association (r=−0.362, P=.03) between the NBDC and self-reported depressive symptoms on the StudentLife data set, which contained mobile phone data from 48 students across a 10-week term at Dartmouth College [15]. Boonstra et al illustrated the feasibility of collecting nearby Bluetooth device information for the depression recognition task, but they did not provide further findings [5].

Several recent studies have investigated the relationships between Bluetooth proximity data and mental health [16-18]. Moturu et al found that individuals with lower sociability (estimated by the NBDC) tend to report lower mood more often [16]. Bogomolov et al established machine learning models to recognize happiness and stress with features of Bluetooth records, calls, and text messages, which obtained accuracy rates of 80.81% and 72.28%, respectively [17,18]. The above three studies were all performed on the "Friends and Family" data set, including 8 weeks of mobile phone data from 117 participants living in a major US university's married graduate student residency.

Previous studies [15-18] have been performed on relatively small (approximately 100 participants) homogeneous (eg, university students) cohorts of participants over relatively short periods (8-10 weeks), which may limit their generalizability. Besides, Bluetooth features used in these studies [15-18] have been limited to basic statistical features (eg, sum, mean, and standard deviation), which are unable to characterize some nonlinear aspects (such as complexity, regularity, and periodicity) of the Bluetooth data. These nonlinear characteristics can reflect individuals' life rhythms, such as circadian and social rhythms, which are affected by depressive symptoms [19]. Therefore, the associations between the NBDC data and depression are yet to be fully explored.

In this paper, we aimed to explore the value of the NBDC data in predicting self-reported depressive symptom severity in a relatively large cohort of individuals with a history of recurrent major depressive disorder. Our first objective was to explore the associations between statistical Bluetooth features and depressive symptom severity. Our second objective was to extract nonlinear features for quantifying complexity, regularity, and periodicity from the NBDC data and test their associations with depression. The third objective was to leverage appropriate machine learning models to predict the severity of depressive symptoms using extracted Bluetooth features.

Methods

Data Set

Study Participants and Settings

The data used in this study were collected from a major EU Innovative Medicines Initiative (IMI) research program Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) [20]. The project aimed to investigate the use of remote measurement technologies (RMTs) to monitor people with depression, epilepsy, and multiple sclerosis in real-world settings. The study protocol for the depression component (Remote Assessment of Disease and Relapse-Major Depressive Disorder; RADAR-MDD) has been described in detail by Matcham et al [21]. The RADAR-MDD project aimed to recruit 600 participants with a recent history of depression from three study sites in Spain (Centro de Investigación Biomédican en Red [CIBER], Barcelona), the Netherlands (Vrije Universiteit Medisch Centrum [VUmc], Amsterdam]), and the United Kingdom (King's College London [KCL]). Recruitment procedures varied slightly across sites with eligible participants identified through existing research infrastructures (in KCL and VUmc) where consent to be contacted for research purposes exists; advertisements in general practices, psychologist practices, and newspapers; Hersenonderzoek.nl [22], a Dutch online registry (VUmc); and mental health services (in KCL and CIBER) [21].

Participants were asked to install passive and active remote monitoring technology (pRMT and aRMT, respectively) apps and use an activity tracker for up to 2 years of follow-up. Many categories of passive and active data were collected and uploaded to an open-source platform, RADAR-base [23].

As the purpose of this paper was to explore the value of the NBDC data in predicting self-reported depressive symptom severity, we focused on the NBDC data, 8-item Patient Health Questionnaire (PHQ-8) data [24], and baseline demographics. However, according to our previous research, the COVID-19 pandemic and related lockdown policies greatly impacted the behaviors (particularly mobility, social interactions, and working environment [working from home]) of European people [25]. To exclude the impact of the COVID-19 pandemic, we performed a preliminary analysis with the data before February 2020.

PHQ-8 Data

The variability of each participant's depressive symptom severity was measured via the PHQ-8, conducted by mobile phones every 2 weeks. The PHQ-8 score ranges from 0 to 24 (increasing severity) [24]. According to the PHQ-8 score, the severity of depression can usually be divided into the following five levels: asymptomatic (PHQ-8 <5), mild (5 ≤ PHQ-8 < 10), moderate (10 ≤ PHQ-8 < 15), moderately severe (15 ≤ PHQ-8 < 20), and severe (PHQ-8 ≥20) [24].

NBDC Data

The RADAR-base pRMT app scanned other Bluetooth devices in the participant's physical proximity once every hour. To avoid privacy leaks from participants and passers, the Media Access Control (MAC) address and types of Bluetooth devices were not recorded in this study. The NBDC was uploaded to the RADAR-base platform for further analyses.

Figure 1 is a schematic diagram showing an individual's NBDC in different scenarios in daily activities and life. At home, the NBDC is related to the number of family members and Bluetooth devices in the house, reflecting the participant's connections with family (whether living alone) and the number of other Bluetooth devices. In public transportation (such as the train, subway, and bus), the NBDC is affected by the number of surrounding passengers' Bluetooth devices, reflecting the participant's social connections with strangers. Studies have shown that whether feeling comfortable in the presence of strangers is related to the intensity of social connections [26]. In the company, the NBDC can reflect the participant's social connections and interactions with co-workers. After work, the NBDC can reflect whether the participant joins other social activities, such as going to the park or bar. Therefore, the NBDC data contain information about participants' social connections and interactions with family, friends, co-workers, and strangers, and the data can also reflect participants' time at home, mobility, social isolation, and working status, as well as the number of other Bluetooth devices in the house and working environment.

Figure 2 shows an example of two NBDC sequences collected over 14 days (336 hours) before two PHQ-8 records from one participant at two different depression severity levels (mild vs moderately severe).

Figure 1. A schematic diagram showing an individual's nearby Bluetooth devices count (NBDC) in different scenarios in daily activities and life.

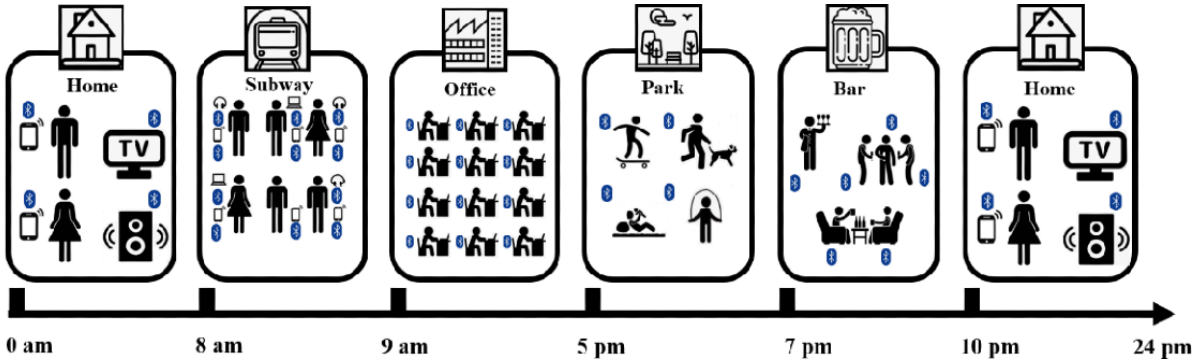
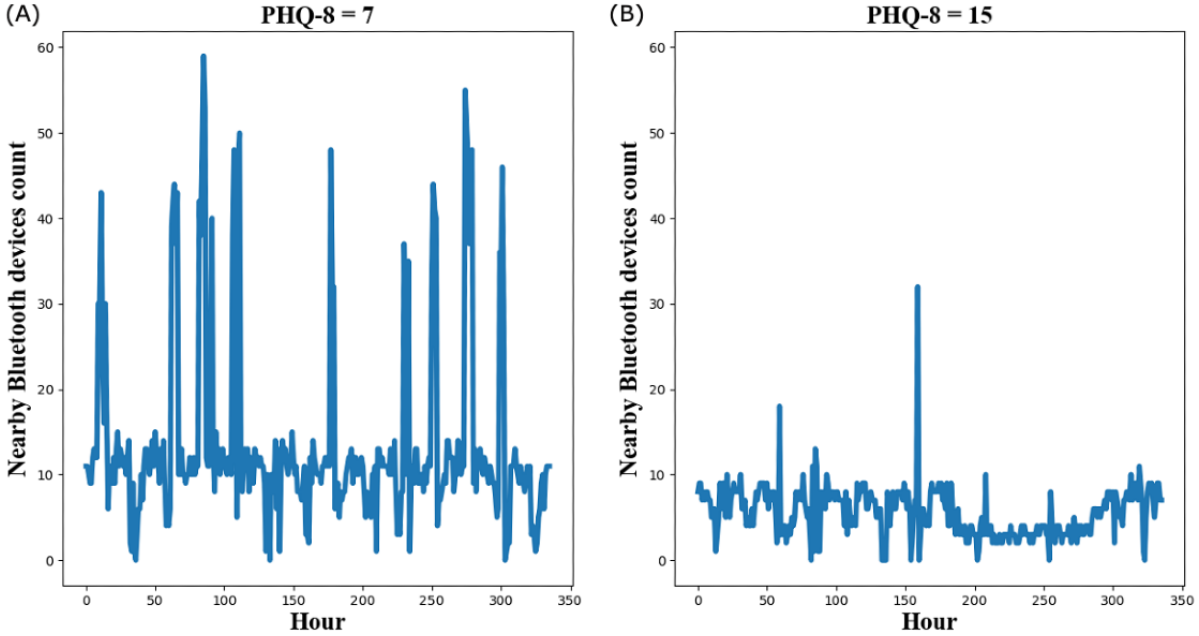


Figure 2. An example of two 14-day nearby Bluetooth devices count (NBDC) sequences from the same participant at the mild depression level (A) and moderately severe level (B). PHQ-8: 8-item Patient Health Questionnaire.



Demographics

Participants' demographics were recorded during the enrollment session. According to previous studies [27,28], baseline age, gender, and education level were considered as covariates in our analyses. Due to the different educational systems in the three countries in our data set, we used the number of years in education to represent education level.

Data Inclusion Criteria and Data Preprocessing

For each PHQ-8 record, we considered a "PHQ-8 interval" of 14 days before the day when the participant fills in the PHQ-8 questionnaire, as the PHQ-8 score is used to represent the depressive symptom severity of the participant for the past 2 weeks. To reduce the impact of the COVID-19 pandemic and missing data on our analysis, we specified the following two data inclusion criteria:

1. As mentioned in the data set section, to exclude the impact of the COVID-19 pandemic, we restricted our analysis to PHQ-8 records prior to February 2020.
2. Saeb et al [29] and Farhan et al [30] used 50% as each day's completeness threshold for passive data. In our data set, 89.62% of days have 50% (12 hours) or more of the NBDC data. We considered one day as a "valid day" if it contained at least 12 hours of the NBDC data. Then, we empirically selected PHQ-8 intervals with at least 10 valid days as valid PHQ-8 intervals to retain the majority (81.78%) of PHQ-8 intervals.

For the NBDC sequence in each selected PHQ-8 interval, we used linear interpolation to impute the missing hours in all valid days and discarded the NBDC data that did not belong to a valid day. The "NBDC sequence" in the rest of this paper refers to the preprocessed NBDC data in the 14-day PHQ-8 interval.

Feature Extraction

According to past Bluetooth-related research [15-18] and research on nonlinear features of signal processing [31,32], we extracted 49 Bluetooth features from the NBDC sequence in the PHQ-8 interval in the following three categories: second-order statistics, multiscale entropy (MSE), and frequency domain (FD). Table 1 summarizes all Bluetooth features extracted in this paper.

Table 1. Summary of 49 Bluetooth features used in this paper and their short descriptions.

Category	Abbreviation	Description	Number of features (N=49)
Statistical features	[Second-order feature]_[Daily feature], eg, Max_Mean	Second-order features (max, min, mean, and standard deviation) calculated in the PHQ-8[a] interval based on daily statistical Bluetooth features (max, min, mean, and standard deviation).	16
Multiscale entropy (MSE)	MSE_1, MSE_2, …, MSE_24	Multiscale entropy of the NBDC[b] sequences from scale 1 to scale 24.	24
Frequency domain[c]	LF_sum, MF_sum, HF_sum	The sums of spectrum power in LF, MF, and HF.	3
Frequency domain	LF_pct, MF_pct, HF_pct	The percentages of spectrum power in LF, MF, and HF to the total spectrum power.	3
Frequency domain	LF_se, MF_se, HF_se	Spectral entropy in LF, MF, and HF.	3

[a]PHQ-8: 8-item Patient Health Questionnaire.

[b]NBDC: nearby Bluetooth device count.

[c]LF: low frequency (0-0.75 cycles/day); MF: middle frequency (0.75-1.25 cycles/day); HF: high frequency (>1.25 cycles/day).

Second-Order Statistical Features

We first calculated four daily features (max, min, mean, and standard deviation) of daily NBDC data from all valid days in the PHQ-8 interval. For each daily feature, we calculated four second-order features (max, min, mean, and standard deviation) to reflect the amount and variance of the NBDC in the PHQ-8 interval. These features were denoted in the following format: [Second-order feature]_[Daily feature]. For example, the average value of the daily maximum number of the NBDC in the PHQ-8 interval was denoted as *Mean_Max*. A total of 16 second-order statistical features were extracted.

Nonlinear Bluetooth Features

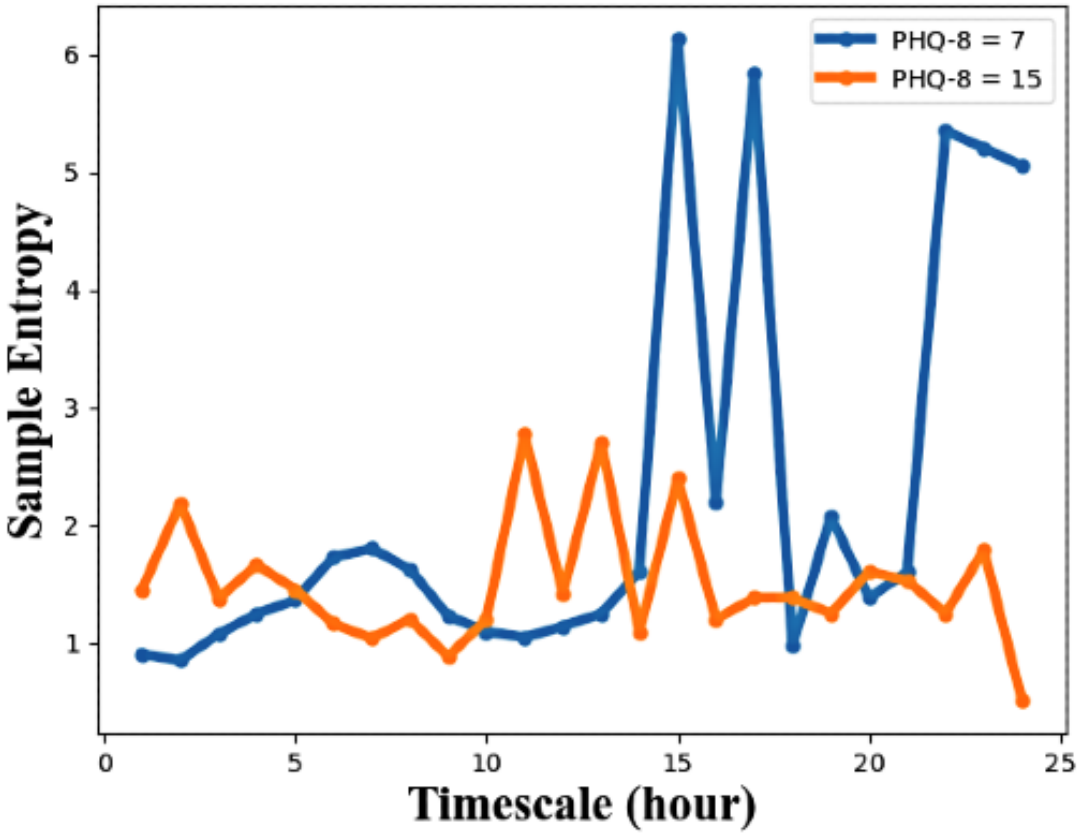
The second-order statistical features can only reflect the amount (max, min, and mean) and variance (standard deviation) of the NBDC data. To exploit more information embedded in the NBDC data, we proposed MSE and FD features to measure the nonlinear characteristics, such as regularity, complexity, and periodicity, of the NBDC sequence.

Multiscale Entropy Features

MSE analysis has been used to provide insights into the complexity and periodicity of signals over a range of timescales since the method was proposed by Costa et at [31]. It has been widely used in the field of signal analysis, such as heart rate variability analysis [33], electroencephalogram analysis [34], and gait dynamics analysis [35]. Compared with other entropy techniques (eg, sample entropy and approximate entropy), the advantage of MSE analysis is that the assessments of complexity at shorter and longer timescales can be analyzed separately [36]. The MSE at short timescales reflects the complexity of the sequence. The larger the MSE at short timescales, the more chaotic and irregular the signal. The MSE at relatively long timescales assesses fluctuations occurring at a certain period, reflecting the periodicity of the signal.

To explore the complexity and periodicity of the NBDC sequence on different timescales (from 1 hour to 24 hours), we calculated MSE features of the NBDC sequences from scale 1 to scale 24, denoted as *MSE_1*, *MSE_2*, …, *MSE_24*. Figure 3 shows an example of MSE features calculated on two NBDC sequences at different depression severity levels from the same participant shown in Figure 2. In this example, the NBDC sequence at the mild depression level (PHQ-8=7) has lower MSE at relatively short timescales (scale 1-3) and higher MSE at relatively long timescales than the sequence at the moderately severe depression level (PHQ-8=15). This indicated that this participant's NBDC sequence at the mild depression level was more regular and periodic than the NBDC sequence at the moderately severe depression level.

Figure 3. An example of multiscale entropy (scale 1-24) of two 14-day nearby Bluetooth device count (NBDC) sequences at the mild depression level (blue) and the moderately severe level (orange) from the same participant as in Figure 2. PHQ-8: 8-item Patient Health Questionnaire.



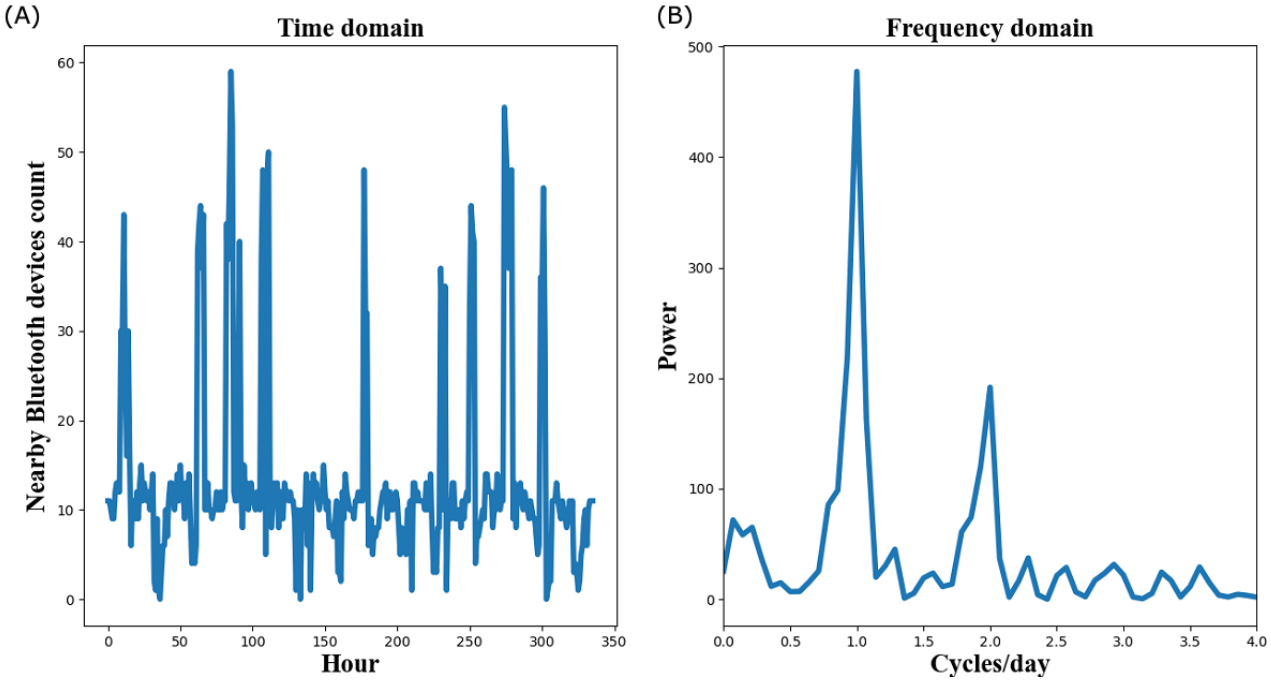
FD Features

FD analysis has been widely used in the signal processing field, especially for signals with periodic characteristics [32]. People's behaviors follow a quasiperiodic routine, such as sleeping at night, working on weekdays, and gathering with friends on weekends [19,37]. We therefore leveraged FD analysis to explore the periodic patterns in the NBDC data. Fast Fourier transformation (FFT) was performed to transform the NBDC sequence from the time domain to the FD. We set the sample rate to 24 hours, and then, the spectrum generated by FFT had the frequency axis scaled to reflect cycles per day.

Figure 4 is an example of a NBDC sequence in the time domain and its spectrum in the FD. According to the spectrum's definition, spectrum power around 1 cycle per day reflects the participant's circadian rhythm (approximately 24-hour rhythm) [19]. To explore the periodic rhythms of different period lengths, we empirically defined the following three frequency intervals: low frequency (LF) (0-0.75 cycles/day), middle frequency (MF) (0.75-1.25 cycles/day), and high frequency (HF) (>1.25 cycles/day). The power in MF represents the circadian rhythm. Similarly, the power in LF represents the long-term (>1 day) rhythm, while the power in HF represents the short-term (<1 day) rhythm.

The sums of spectrum power in these three frequency intervals were calculated and denoted as *LF_sum*, *MF_sum*, and *HF_sum*, respectively. The percentages of spectrum powers in these three frequency intervals to the total spectrum power were extracted and denoted as *LF_pct*, *MF_pct*, and *HF_pct*, respectively. To estimate the complexity and regularity of the spectrum, we calculated spectral entropy (SE) [38] in these three intervals, denoted as *LF_se*, *MF_se*, and *HF_se*, respectively.

Figure 4. An example of a 14-day nearby Bluetooth devices count (NBDC) sequence in the time domain (A) and its spectrum in the frequency domain (B).



Statistical Methods

The linear mixed-effect model contains both fixed and random effects, allowing for both within-participant and between-participants variations over repeated measurements [39]. Therefore, we used linear mixed-effect models in our statistical analyses.

Pairwise Association Analyses

To explore the association between each Bluetooth feature and depression severity, a series of pairwise linear mixed-effect models with random participant intercepts were performed to regress the PHQ-8 score with each of the Bluetooth features. All mixed-effect models, baseline age, gender, and years in education were considered as covariates. The z-test was used to evaluate the statistical significance of the coefficient of each model. The Benjamini-Hochberg method [40] was used for correction of multiple comparisons, and the significant level for the adjusted *P* value was set to .05. All linear mixed-effect models were implemented by using the R package "lmerTest," and the Benjamini-Hochberg method was performed by using the command "p.adjust" in R software (R Foundation for Statistical Computing).

Likelihood Ratio Test

One objective of this paper was to assess what value these Bluetooth features provide beyond other information that might be readily available, such as baseline demographics. The likelihood ratio test is a statistical test of goodness of fit between two nested models [41]. If the model with more parameters fits the data significantly better, it indicates that additional parameters provide more information and improve the model's fitness [41]. Therefore, we built three nested linear mixed-effect models with random participant intercepts (model A, model B, and model C). The predictors of model A were only demographics. The predictors of model B were demographics and 16 second-order statistical features. The predictors of model C were demographics and all 49 Bluetooth features. The likelihood ratio tests were performed to test whether these Bluetooth features have a significant value in fitting the PHQ-8 score regression model.

Prediction Models

Another objective of this paper was to examine whether it is possible to predict participants' depressive symptom severity using Bluetooth features combined with some known information (demographics and previous PHQ-8 scores). A subset of PHQ-8 intervals was selected for the prediction task based on the following two additional criteria:

1. To ensure that each participant had sufficient PHQ-8 intervals for the time-series cross-validation (described in the following model evaluation section), the number of valid PHQ-8 intervals for each participant should be at least 3.
2. To test whether the model can predict variability of depression severity, the difference of one participant's PHQ-8 scores should be more than or equal to 5 (clinically meaningful change) [42].

Hierarchical Bayesian Linear Regression Model

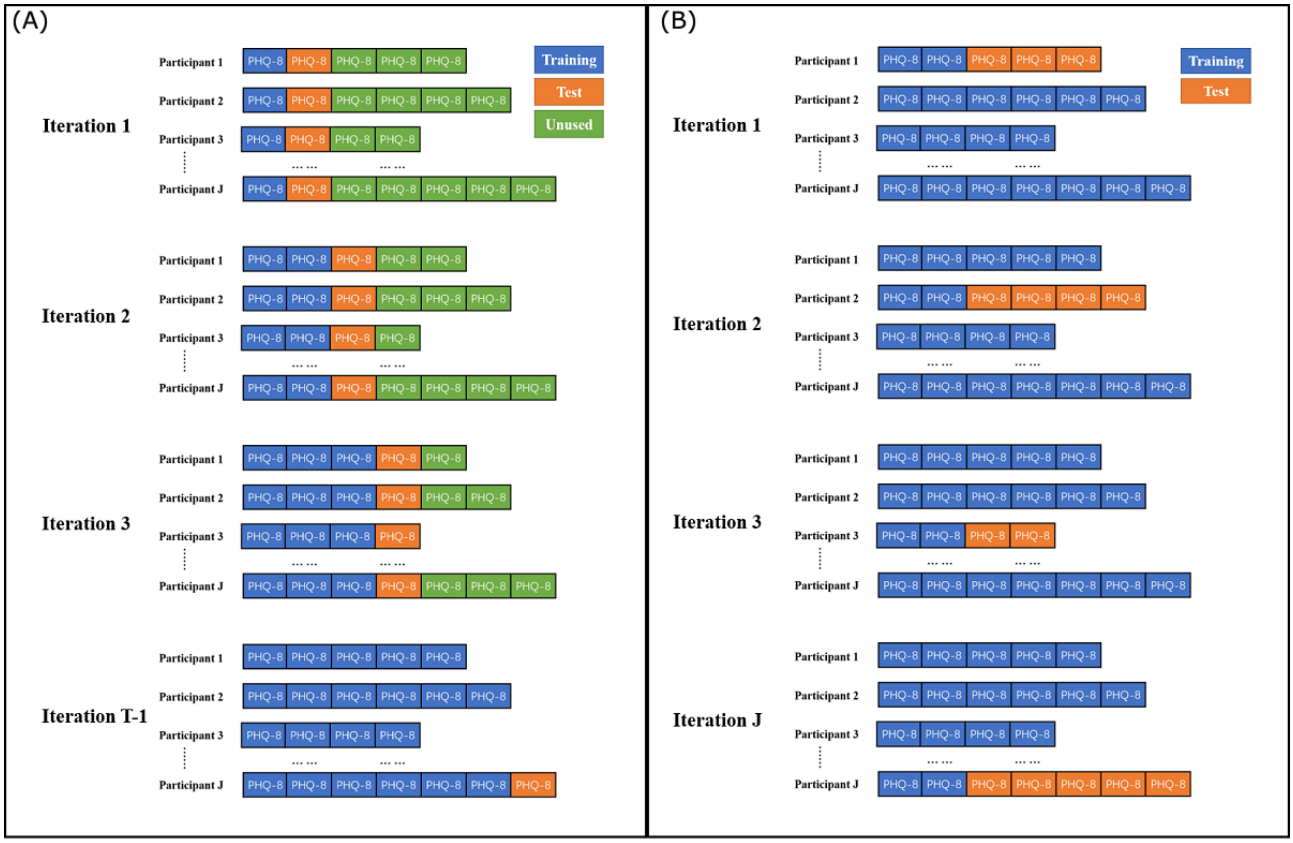
The hierarchical Bayesian approach is an intermediate method compared to the completely pooled model and individualized model, capturing the whole population's characteristics while allowing individual differences [43]. We leveraged the hierarchical Bayesian linear regression model to predict participants' PHQ-8 scores using Bluetooth features, demographics (age, gender, and years in education), and the last observed PHQ-8 score. In this study, we implemented the hierarchical Bayesian linear regression using the "PyMC3" package [44] in Python. To compare the results with other commonly used machine learning models, we also implemented

the LASSO regression model [45] and XGBoost regression model [46] using the Scikit-learn machine learning library [47] in Python. As depressive mood has a strong autocorrelation [48], we considered a baseline hierarchical Bayesian linear regression model with the last observed PHQ-8 score and demographics as predictors.

Model Evaluation

We selected root mean squared error (RMSE) and the predicted coefficient of determination (R^2) as two metrics for model discrimination evaluation. As we used the temporal data, "future data" should not predict "past data." Therefore, only the data observed before test data can be included in the training set. We applied leave-all-out (LAO) and leave-one-out (LOO) time-series cross-validation [48]. As the number of PHQ-8 intervals of each participant in our data was different, we made some minor modifications to these two schemes (Figure 5).

Figure 5. Two schematic diagrams of leave-all-out time-series cross-validation (A) and leave-one-out time-series cross-validation (B), where T is the maximum number of PHQ-8 intervals of one participant, J is the number of participants, the training set is indicated by blue, the test set is indicated by orange, and unused data are indicated by green. PHQ-8: 8-item Patient Health Questionnaire.



LAO Time-Series Cross-Validation

Each participant's data were divided into a sequence of t consecutive same-sized test sets, where the size of each test set is the length of one PHQ-8 interval (14 days) and t is the number of PHQ-8 intervals of this participant. The corresponding training set included all PHQ-8 intervals before each test set. Then, test sets and training sets were pooled across all participants. This process generated T-1 test and training set pairs (no prior data to predict the first PHQ-8 score), where T is the maximum number of PHQ-8 intervals of one participant in our data set ($t \leq T$).

LOO Time-Series Cross-Validation

Each participant's data were divided into a training set and a test set. The training set was constructed using the first two PHQ-8 intervals of a participant, with the test set containing the rest of the participant's PHQ-8 intervals. Then, the training set was pooled with all data from all other participants. This scheme generated J training and test set pairs, where J is the number of participants in our data set.

Results

Data Summary

According to our date inclusion criteria, from June 2018 to February 2020, 2886 PHQ-8 intervals from 316 participants collected from three study sites were selected for our analysis. Table 2 shows the descriptive statistics for all 49 Bluetooth features, and Figure 6 presents pairwise Spearman correlation coefficients between all features. Table 3 presents a summary of the demographics and distribution of PHQ-8 records of all selected participants. Figure 7 presents boxplots of the NBDC for every hour in the whole population.

Table 2. Descriptive statistics for all 49 Bluetooth features.

Feature[a]	Mean	SD	Min	Q1	Median	Q3	Max
Second-order statistics							
Max_Max	49.79	48.48	1.00	25.00	40.00	60.00	621.00
Min_Max	5.09	6.22	0.00	2.00	4.00	6.00	90.00
Mean_Max	18.56	18.94	0.75	9.23	14.07	21.62	268.29
Std_Max	13.14	14.05	0.00	6.14	10.45	16.22	195.19
Max_Min	1.59	2.08	0.00	0.00	1.00	2.00	43.00
Min_Min	0.06	0.27	0.00	0.00	0.00	0.00	3.00
Mean_Min	0.58	0.88	0.00	0.00	0.21	0.79	13.71
Std_Min	0.50	0.62	0.00	0.00	0.42	0.70	11.94
Max_Std	12.31	12.76	0.34	5.60	9.51	15.39	185.98
Min_Std	1.20	1.45	0.00	0.56	0.87	1.32	21.61
Mean_Std	4.55	4.87	0.16	2.17	3.25	5.24	70.65
Std_Std	3.24	3.71	0.09	1.34	2.43	4.04	62.52
Max_Mean	9.32	9.34	0.17	4.38	6.88	11.04	136.10
Min_Mean	1.88	2.14	0.00	0.50	1.42	2.50	32.00
Mean_Mean	4.42	4.19	0.07	2.19	3.40	5.28	49.55
Std_Mean	2.13	2.59	0.05	0.84	1.45	2.54	49.37
Multiscale entropy (MSE)							
MSE_1	0.80	0.46	0.05	0.42	0.71	1.13	2.44
MSE_2	0.97	0.54	0.04	0.56	0.85	1.31	3.58
MSE_3	1.12	0.66	0.09	0.70	1.01	1.42	9.41
MSE_4	1.23	0.69	0.05	0.82	1.15	1.51	8.83
MSE_5	1.35	0.82	0.10	0.93	1.27	1.62	8.51
MSE_6	1.38	0.84	0.08	0.97	1.28	1.63	8.00
MSE_7	1.47	0.97	0.10	1.01	1.33	1.70	7.72
MSE_8	1.50	1.07	0.10	1.00	1.30	1.67	7.40
MSE_9	1.58	1.22	0.10	0.99	1.32	1.72	7.30
MSE_10	1.58	1.23	0.08	0.97	1.30	1.72	7.08
MSE_11	1.58	1.29	0.09	0.95	1.25	1.67	7.02
MSE_12	1.59	1.33	0.10	0.92	1.23	1.66	6.70
MSE_13	1.74	1.46	0.11	0.98	1.30	1.79	6.55
MSE_14	1.85	1.53	0.11	1.01	1.36	1.87	6.70
MSE_15	1.96	1.62	0.13	1.03	1.39	1.95	6.55
MSE_16	1.98	1.62	0.13	1.03	1.39	1.95	6.40
MSE_17	2.04	1.67	0.14	1.02	1.39	2.08	6.14
MSE_18	2.03	1.65	0.15	1.01	1.39	2.08	6.04
MSE_19	2.09	1.69	0.17	1.01	1.39	2.08	6.04
MSE_20	2.09	1.67	0.17	0.98	1.39	2.08	5.94
MSE_21	2.10	1.66	0.18	0.98	1.39	2.20	5.83
MSE_22	2.13	1.68	0.18	0.98	1.39	2.30	5.83
MSE_23	2.17	1.69	0.18	0.98	1.39	4.28	5.61
MSE_24	2.27	1.70	0.20	0.98	1.39	4.28	5.35

Feature[a]	Mean	SD	Min	Q1	Median	Q3	Max
Frequency domain (FD)							
LF[b]_sum	330.66	2469.74	0.05	17.41	53.87	184.80	85956.16
MF[c]_sum	157.24	1166.32	0.02	8.16	25.77	83.05	34970.35
HF[d]_sum	602.22	3272.44	0.47	55.72	151.74	403.38	64127.16
LF_pct[e]	0.25	0.10	0.03	0.17	0.23	0.31	0.63
MF_pct	0.13	0.10	0.01	0.07	0.11	0.17	0.74
HF_pct	0.62	0.15	0.12	0.53	0.64	0.72	0.92
LF_se[f]	0.83	0.10	0.38	0.78	0.85	0.90	1.00
MF_se	0.82	0.09	0.40	0.77	0.83	0.88	0.99
HF_se	0.90	0.04	0.72	0.88	0.90	0.92	0.99

[a]Definitions of Bluetooth features in this table are shown in Table 1.

[b]LF: low frequency (0-0.75 cycles/day).

[c]MF: middle frequency (0.75-1.25 cycles/day).

[d]HF: high frequency (>1.25 cycles/day).

[e]pct: percentage of spectrum power.

[f]se: spectral entropy.

Figure 6. A correlation plot of pairwise Spearman correlations between all 49 Bluetooth features. Definitions of Bluetooth features in this figure are shown in Table 1.

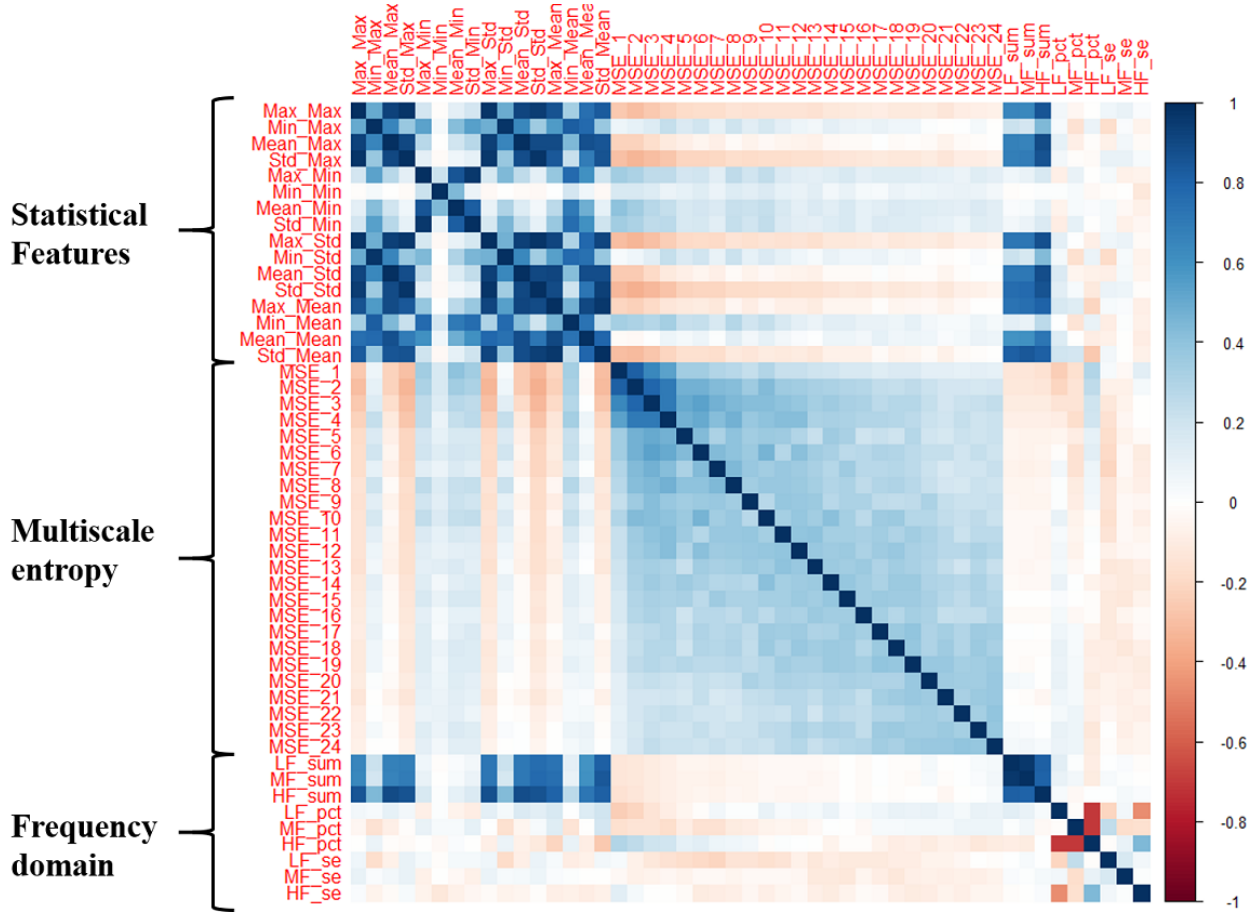
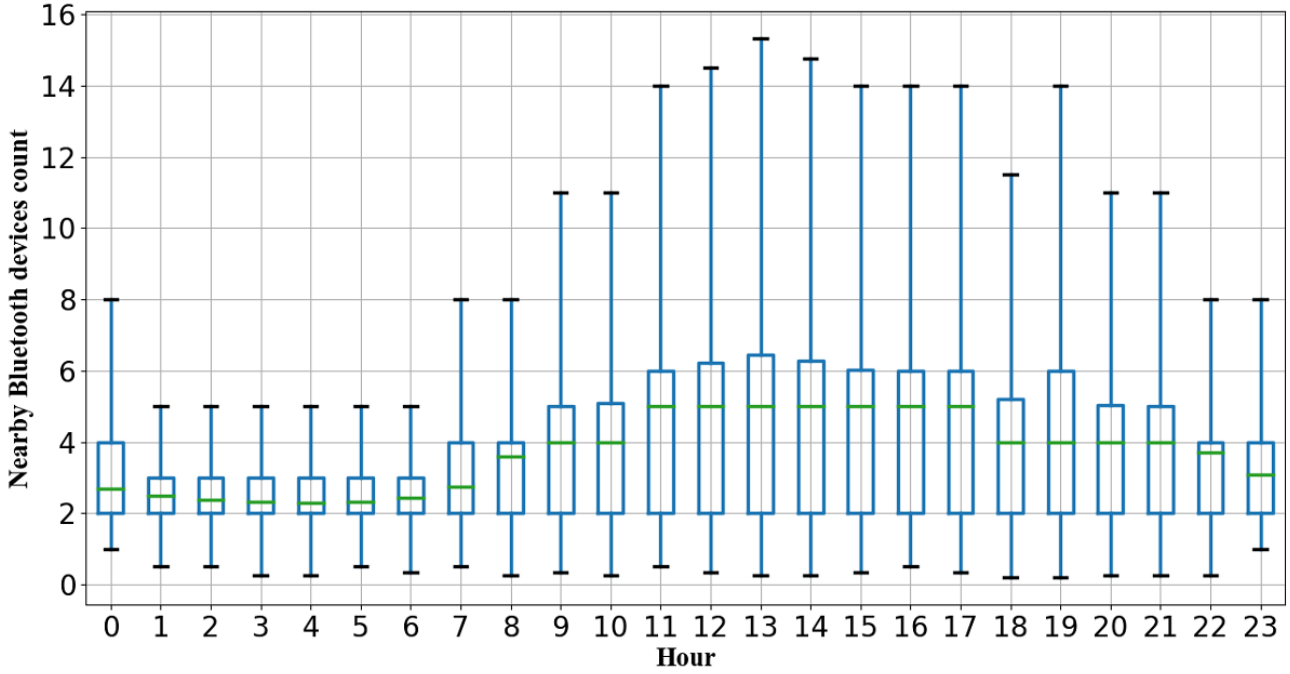


Table 3. Summary of the demographics and 8-item Patient Health Questionnaire (PHQ-8) record distribution of all selected participants.

Characteristic	Value
Number of participants	316
Demographics	
Age at baseline, median (Q1, Q3)	51.0 (35.0, 59.0)
Female sex, n (%)	234 (74.1%)
Number of years in education, median (Q1, Q3)	16.0 (14.0, 19.0)
PHQ-8 record distribution	
Number of PHQ-8 intervals	2886
Number of PHQ-8 intervals for each participant, median (Q1, Q3)	8.0 (3.0, 14.0)
PHQ-8 score, median (Q1, Q3)	9.0 (5.0, 15.0)

Figure 7. Boxplots of the nearby Bluetooth devices count (NBDC) for every hour in the whole population. Boxes extend between the 25th and 75th percentiles, and green solid lines inside the boxes are medians. Note the relative stationary NBDC during the night-time hours.



Association Analysis Results

The significant associations between depression severity (the PHQ-8 score) and Bluetooth features are presented in Table 4.

Associations Between the PHQ-8 Score and Second-Order Statistical Features

There were 10 second-order statistical features significantly associated with the PHQ-8 score. All these significant associations were negative, that is, the larger the value of these features, the lower the PHQ-8 score. Notably, *Min_Max* (the minimum value of daily maximum NBDC in the past 14 days) had the strongest association (z=−4.431, *P*<.001), which indicated that participants with a lower PHQ-8 score tended to have more daily social activities (such as social interactions and traveling) in the past 2 weeks. In addition, four features related to daily variance (*Max_Std*, *Min_Std*, *Mean_Std*, and *Std_Std*) of the NBDC were all significantly and negatively associated with depression.

Associations Between the PHQ-8 Score and Multiscale Entropy Features

MSE at scale 1, scale 2, and scale 3 (*MSE_1*, *MSE_2*, and *MSE_3*) were significantly and positively associated with the PHQ-8 score, while MSE at scale 16 and scale 22 (*MSE_16* and *MSE_22*) were significantly and negatively associated with depressive symptom severity. According to the explanations of MSE we mentioned in the Methods section, these associations indicated that participants with more irregular and chaotic NBDC sequences were likely to have more severe depressive symptoms, while those with periodic and regular NBDC sequences may have lower PHQ-8 scores.

Associations Between the PHQ-8 Score and FD Features

There were five FD features significantly associated with the PHQ-8 score. The spectrum power was related to both the amount and frequency components of the NBDC sequence, so it had relatively strong correlations with second-order statistical features (Figure 6). Therefore, the spectrum power of three

frequency intervals (*LF_sum*, *MF_sum*, and *HF_sum*) were all significantly and negatively associated with the PHQ-8 score. Among them, the *MF_sum* had the strongest association (z=−4.766, *P*<.001) with depression, which indicated that the circadian rhythm of the NBDC sequence is important to reflect the severity of depression. Likewise, the percentage of middle-frequency power (*MF_pct*) was significantly and negatively associated with depressive symptom severity. The spectral entropy of HF (*HF_se*) was significantly and positively associated with depression. This indicated that participants with irregular short-term (<1 day) rhythms were likely to have more severe depressive symptoms.

Table 4. Coefficient estimates, standard error, z-test statistics, and *P* values from pairwise linear mixed-effect models for exploring associations between Bluetooth features and the depressive symptom severity (8-item Patient Health Questionnaire).

Feature[a]	Estimate	SE	z score	Adjusted *P* value[b,c]
Second-order statistics				
Min_Max	−0.052	0.012	−4.431	<.001
Mean_max	−0.016	0.006	−2.809	.005
Max_Std	−0.015	0.006	−2.657	.008
Min_Std	−0.215	0.056	−3.838	<.001
Mean_Std	−0.065	0.023	−2.802	.005
Std_Std	−0.048	0.020	−2.385	.02
Max_Mean	−0.030	0.008	−3.498	<.001
Min_Mean	−0.093	0.046	−2.036	.04
Mean_Mean	−0.083	0.026	−3.225	.001
Std_Mean	−0.095	0.027	−3.464	.001
Multiscale entropy (MSE)				
MSE_1	0.642	0.225	2.853	.005
MSE_2	0.433	0.192	2.255	.02
MSE_3	0.401	0.202	1.985	.04
MSE_16	−0.102	0.042	−2.429	.01
MSE_22	−0.123	0.043	−2.860	.005
Frequency domain (FD)				
LF[d]_sum	−0.021	0.005	−3.865	<.001
MF[e]_sum	−0.067	0.014	−4.766	<.001
HF[f]_sum	−0.027	0.010	−2.606	.009
MF_pct[g]	−1.834	0.812	−2.259	.02
HF_se[h]	3.821	1.820	2.099	.04

[a]Definitions of Bluetooth features in this table are shown in Table 1.

[b]Only significant associations (adjusted *P* value <.05) are reported.

[c]*P* values were adjusted by the Benjamini-Hochberg method for correction of multiple comparisons.

[d]LF: low frequency (0-0.75 cycles/day).

[e]MF: middle frequency (0.75-1.25 cycles/day).

[f]HF: high frequency (>1.25 cycles/day).

[g]pct: percentage of spectrum power.

[h]se: spectral entropy.

Results of Likelihood Ratio Tests

The results of the likelihood ratio tests are presented in Table 5. Model B (with second-order statistical Bluetooth features) and model C (with all Bluetooth features) fitted data significantly better than model A (without Bluetooth features), indicating that Bluetooth features could improve the statistical model significantly. The goodness of fit of model C was significantly better than that of model B, indicating that nonlinear Bluetooth features (MSE and FD features) provided additional information to the statistical model.

Table 5. Results of the likelihood ratio tests of the three nested linear mixed-effect models.

Model	Difference of parameters	Chi-square[a]	*P* value
Model B[b] vs model A[c]	16	31.04	.01
Model C[d] vs model A	49	135.19	<.001
Model C vs model B	33	104.15	<.001

[a]The critical values of the likelihood ratio statistic are as follows: $\chi^2_{0.05}(16)=26.296$, $\chi^2_{0.05}(33)=47.400$, and $\chi^2_{0.05}(49)=66.339$.

[b]Predictors of model B: demographics + 16 second-order statistical features.

[c]Predictors of model A: demographics.

[d]Predictors of model C: demographics + 16 second-order statistical features + 24 multiscale entropy features + nine frequency domain features.

Performance of Prediction Models

A subset of 183 participants was selected for the prediction models. The results of the LAO and LOO time-series cross-validation are presented in Table 6. The R^2 score of the baseline model was 0.338 in LAO time-series cross-validation, which showed that more than 30% variance could be explained by the last observed PHQ-8 score and baseline demographics. In LOO time-series cross-validation, the R^2 score of the baseline model was negative, which indicated that the baseline model did not explain any variance in the LOO time-series cross-validation. To assess the improvement from nonlinear Bluetooth features, we tested the hierarchical Bayesian model with and without nonlinear Bluetooth features separately.

Table 6. Results of the leave-all-out time-series cross-validation and leave-one-out time-series cross-validation of the hierarchical Bayesian linear regression model, commonly used machine learning models, and the baseline model.

Model	Leave-all-out		Leave-one-out	
	R^2	RMSE[a]	R^2	RMSE
Baseline model[b]	0.338	4.547	−0.074	5.802
LASSO regression	0.458	4.114	0.144	5.178
XGBoost regression	0.464	4.092	0.346	4.523
Hierarchical Bayesian linear (second-order statistical features)	0.481	4.026	0.353	4.501
Hierarchical Bayesian linear (all Bluetooth features)	0.526	3.891	0.387	4.426

[a]RMSE: root mean squared error.

[b]The baseline model is the hierarchical Bayesian linear regression model with only the last observed 8-item Patient Health Questionnaire score and demographics as predictors.

In the subset, the maximum number of PHQ-8 intervals of one participant was 27, so the LAO time-series cross-validation went through T-1=26 iterations. The hierarchical Bayesian linear regression model with all Bluetooth features achieved the best result (R^2=0.526, RMSE=3.891), beating the LASSO and XGBoost regression models. Compared with the result of the baseline model (R^2=0.338), the improvement in the R^2 score was 0.188, which means the Bluetooth features explained an additional 18.8% of data variance. The nonlinear Bluetooth features explained an additional 4.5% of data variance in the hierarchical Bayesian model.

The number of subset participants was 183, so J=183 iterations of the LOO time-series cross-validation were performed. The hierarchical Bayesian linear model with all Bluetooth features had the best performance (R^2=0.387, RMSE=4.426), but the result was close to that of the XGBoost regression model (R^2=0.346, RMSE=4.523).

The performance of the hierarchical Bayesian linear regression model evaluated by the LAO cross-validation was better than the LOO cross-validation performance. One potential reason is that only the first two PHQ-8 intervals of one participant were used for training in the LOO cross-validation, which may have caused the model to underfit the patterns at the participant level.

Discussion

Principal Findings

This paper explored the value of the NBDC data in predicting depression severity. Compared with previous Bluetooth-related studies [15-18], our study was performed on a larger (N=316) multicenter data set with a longer follow-up (median 4 months). We extracted 49 features from the NBDC sequences in the following three categories: second-order statistical features, MSE features, and FD features. To the best of our knowledge, this is the first time that MSE and FD features have been used in NBDC and depression data analyses. According to the results of association analyses (Table 4), when depression symptoms worsened (increase in the PHQ-8 score), one or more of the following changes were seen in the preceding 14 days of the NBDC sequence: (1) the amount decreased, which is consistent with the finding by Wang et al [15], (2) the variance decreased,

(3) the periodicity (especially the circadian rhythm) decreased, and (4) the NBDC sequence became more irregular and chaotic.

These changes in the NBDC data can be explained by depression symptoms. The main manifestations of depression include negative feelings (such as sadness, guilt, stress, and tiredness) and loss of interest or pleasure [49]. This may lead to changes in behaviors, such as increased time at home [29,50], decreased mobility [3,29], loss of the ability to work or study [2,49], reduced intensity of social interactions [1], unstable and irregular sleep [51], and decreased engagement in activities [52]. The increased time at home, inability to work or study, and diminished social interactions are reflected in the reduced amount of the NBDC sequence. The decreased mobility and engagement in activities may be possible reasons why participants with higher PHQ-8 scores have lower variance-related features (*Max_Std*, *Min_Std*, *Mean_Std*, and *Std_Std*). Depression also may lead to misalignment of the circadian rhythm and make people's life rhythms (such as sleep rhythms and social rhythms) more irregular [19]. This can be reflected in reduced periodicity and increased irregularity of the NBDC sequence. Saeb et al [29] and Farhan et al [30] found similar findings in GPS data, and showed that the circadian rhythm of the GPS signal was significantly and negatively correlated with depression.

From the perspective of the statistical model, Bluetooth features extracted in this paper significantly improved the goodness of fit for the PHQ-8 score, and nonlinear Bluetooth features (MSE and FD features) can provide additional information to second-order statistical features (Table 5). From the perspective of the prediction model, these 49 Bluetooth features explained an extra 18.8% of the variance in the PHQ-8 score relative to the baseline model, containing only the last PHQ-8 score and demographics, and MSE and FD features explained an extra 4.5% of data variance in the hierarchical Bayesian model (Table 6). From the perspective of the correlations between Bluetooth features (Figure 6), we can observe that, except for three FD features related to the spectrum power that had relatively strong correlations with second-order statistical features, the correlations between other nonlinear Bluetooth features and second-order statistical features were not obvious. This indicated that the MSE and FD features captured dimensions of information to second-order statistical features.

In our prediction model, the hierarchical Bayesian linear regression model achieved the best results in both the LAO and LOO time-series cross-validation. Compared with other models, one of the advantages of the hierarchical Bayesian model is that it performs individual predictions while considering the population's common characteristics [43]. Therefore, the hierarchical Bayesian model can be considered a suitable prediction modelling method for longitudinal data. The LOO time-series cross-validation results illustrated that the hierarchical Bayesian model could predict depression for participants with few observations (only two PHQ intervals in the training set) that overcomes the cold start problem. The hierarchical Bayesian linear model achieved a better result in the LAO time-series cross-validation, which indicated that the prediction results gradually became more accurate and individualized when each participant had more data available in the training set.

Limitations

The RADAR-MDD project was designed for long-term monitoring (up to 2 years) and collecting many other passive data, such as GPS data, acceleration data, app usage, and screen lightness, which need to be collected simultaneously through the mobile phone. Therefore, to avoid excessive battery consumption, nearby Bluetooth devices were scanned hourly in this study. However, some past studies suggested scanning nearby Bluetooth devices every 5 minutes to achieve high enough temporal resolution [9,18]. Although hourly NBDC data can also reflect individuals' behaviors and statuses, our lower data resolution may cause the loss of some dynamic information. On the other hand, using the relatively low resolution enabled us to collect multimodal data without excessive battery consumption. As the NBDC data are related to individuals' movement and location information, we will combine the NBDC data with GPS and acceleration data for future analysis to understand the context of the Bluetooth data.

As we mentioned in the Methods section, the MAC addresses and types of Bluetooth devices were not recorded for private issues. This made it impossible to distinguish between mobile phones and other Bluetooth devices (such as headphones, printers, and laptops), and between strangers' and acquaintances' devices. The advantage of the NBDC data is that the data contain mixed and rich information. The disadvantage is that it is difficult to explain the specific reasons for changes in the NBDC, that is, we cannot know whether the changes in the NBDC are caused by social interactions, working status, traveling, or isolation. Therefore, this paper did not explain in depth the actual meaning behind the Bluetooth features. For this limitation, we plan to use hashed MAC addresses in future research.

For the FD features, the division of the frequency intervals of the spectrum of the NBDC sequence in this paper was manually specified by our experience. The purpose of extracting these FD features was to prove that the NBDC sequence's FD has the potential to provide more information about individuals' behaviors and life rhythms. It is necessary to discuss the optimal boundaries of frequency intervals of the NBDC data in future research.

This paper applied the hierarchical Bayesian linear regression model to explore the linear relationships between Bluetooth features and depression. However, there may be nonlinear relationships between social connections and depressive symptom severity. The Gaussian process [53], using the kernel method to find nonlinear relationships, will be considered in future research.

Conclusion

Our statistical results indicated that the NBDC data have the potential to reflect changes in individuals' behaviors and statuses during a depressive state. The prediction results demonstrated that the NBDC data have significant value in predicting depressive symptom severity. The nonlinear Bluetooth features proposed in this paper provide additional information to

statistical and prediction models. The hierarchical Bayesian model is an appropriate prediction model for predicting depression with longitudinal data, as both participant-level and population-level characteristics are considered in the model. These findings may support the mental health monitoring practice in real-world settings.

Acknowledgments

The Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) project has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant agreement number 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA) (www.imi.europa.eu). This communication reflects the views of the RADAR-CNS consortium, and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. The funding body was not involved in the design of the study, the collection or analysis of the data, or the interpretation of the data. Participants in the CIBER site came from the following four clinical communities in Spain: Parc Sanitari Sant Joan de Déu Network services, Institut Català de la Salut, Institut Pere Mata, and Hospital Clínico San Carlos. Participant recruitment in Amsterdam was partially accomplished through Hersenonderzoek.nl, a Dutch online registry that facilitates participant recruitment for neuroscience studies. Hersenonderzoek.nl is funded by ZonMw-Memorabel (project number 73305095003), a project in the context of the Dutch Deltaplan Dementie, Gieskes-Strijbis Foundation, the Alzheimer's Society in the Netherlands, and Brain Foundation Netherlands. This paper represents independent research partly funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health and Social Care. We thank all the members of the RADAR-CNS patient advisory board for their contribution to the device selection procedures and for their invaluable advice throughout the study protocol design. This research was reviewed by a team with experience in mental health problems and their careers who have been specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers (FAST-R), a free confidential service in England provided by the National Institute for Health Research Maudsley Biomedical Research Centre via King's College London and South London and Maudsley NHS Foundation Trust. Remote Assessment of Disease and Relapse-Major Depressive Disorder (RADAR-MDD) will be conducted according to the Declaration of Helsinki and Good Clinical Practice, adhering to principles outlined in the NHS Research Governance Framework for Health and Social Care (2nd edition). Ethical approval has been obtained in London from the Camberwell St Giles Research Ethics Committee (REC reference: 17/LO/1154), in London from the CEIC Fundacio Sant Joan de Deu (CI: PIC-128-17), and in the Netherlands from the Medische Ethische Toetsingscommissie VUms (METc VUmc registratienummer: 2018.012 – NL63557.029.17). RJBD is supported by the following: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, UK; (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation, and Wellcome Trust; (3) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement number 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and the European Society of Cardiology; (4) The National Institute for Health Research University College London Hospitals Biomedical Research Centre; (5) The NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; (6) The UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare; (7) The NIHR Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. This paper represents independent research partly funded by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. MH and TW have been supported by NIHR Senior Investigator Awards.

Conflicts of Interest

VAN is an employee of Janssen Research and Development LLC. PA is employed by the pharmaceutical company H. Lundbeck A/S. DCM has accepted honoraria and consulting fees from Apple, Inc, Otsuka Pharmaceuticals, Pear Therapeutics, and the One Mind Foundation; has received royalties from Oxford Press; and has an ownership interest in Adaptive Health, Inc.

References

1. Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. Psychol Aging 2006 Mar;21(1):140-151. [doi: 10.1037/0882-7974.21.1.140] [Medline: 16594799]

2. Rizvi SJ, Cyriac A, Grima E, Tan M, Lin P, Gallaugher LA, et al. Depression and employment status in primary and tertiary care settings. Can J Psychiatry 2015 Jan;60(1):14-22 [FREE Full text] [doi: 10.1177/070674371506000105] [Medline: 25886545]
3. Lampinen P, Heikkinen E. Reduced mobility and physical activity as predictors of depressive symptoms among community-dwelling older adults: an eight-year follow-up study. Aging Clin Exp Res 2003 Jun;15(3):205-211. [doi: 10.1007/BF03324501] [Medline: 14582683]
4. Burgess P, Pirkis J, Morton J, Croke E. Lessons from a comprehensive clinical audit of users of psychiatric services who committed suicide. Psychiatr Serv 2000 Dec;51(12):1555-1560. [doi: 10.1176/appi.ps.51.12.1555] [Medline: 11097653]
5. Boonstra T, Werner-Seidler A, O'Dea B, Larsen M, Christensen H. Smartphone app to investigate the relationship between social connectivity and mental health. 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 11-15, 2017; Jeju, Korea (South) p. 287-290. [doi: 10.1109/EMBC.2017.8036818]
6. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
7. Yan Z, Yang J, Tapia EM. Smartphone bluetooth based social sensing. In: UbiComp '13 Adjunct: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. 2013 Presented at: ACM conference on Pervasive and ubiquitous computing adjunct publication; September 8-12, 2013; Zurich, Switzerland p. 95-98. [doi: 10.1145/2494091.2494118]
8. Eagle N, Pentland A, Lazer D. Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 2009 Sep 08;106(36):15274-15278 [FREE Full text] [doi: 10.1073/pnas.0900282106] [Medline: 19706491]
9. Eagle N, (Sandy) Pentland A. Reality mining: sensing complex social systems. Pers Ubiquit Comput 2005 Nov 3;10(4):255-268. [doi: 10.1007/s00779-005-0046-3]
10. Dissing AS, Jørgensen TB, Gerds TA, Rod NH, Lund R. High perceived stress and social interaction behaviour among young adults. A study based on objective measures of face-to-face and smartphone interactions. PLoS One 2019 Jul 26;14(7):e0218429 [FREE Full text] [doi: 10.1371/journal.pone.0218429] [Medline: 31348787]
11. Clark BK, Winkler EA, Brakenridge CL, Trost SG, Healy GN. Using Bluetooth proximity sensing to determine where office workers spend time at work. PLoS One 2018 Mar 7;13(3):e0193971 [FREE Full text] [doi: 10.1371/journal.pone.0193971] [Medline: 29513754]
12. Nordström E, Diot C, Gass R, Gunningberg P. Experiences from measuring human mobility using Bluetooth inquiring devices. In: MobiEval '07: Proceedings of the 1st international workshop on System evaluation for mobile platforms. 2007 Presented at: 1st international workshop on System evaluation for mobile platforms; June 11, 2007; San Juan, Puerto Rico p. 15-20. [doi: 10.1145/1247721.1247725]
13. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying Behavioral Phenotypes of Loneliness and Social Isolation with Passive Sensing: Statistical Analysis, Data Mining and Machine Learning of Smartphone and Fitbit Data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
14. Wu C, Barczyk AN, Craddock RC, Harari GM, Thomaz E, Shumake JD, et al. Improving prediction of real-time loneliness and companionship type using geosocial features of personal smartphone data. Smart Health 2021 Apr;20:100180. [doi: 10.1016/j.smhl.2021.100180]
15. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington p. 3-14. [doi: 10.1145/2632048.2632054]
16. Moturu S, Khayal I, Aharony N, Wei P, Pentland A. Sleep, mood and sociability in a healthy population. 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 3, 2011; Boston, MA, USA p. 5267-5270. [doi: 10.1109/iembs.2011.6091303]
17. Bogomolov A, Lepri B, Pianesi F. Happiness Recognition from Mobile Phone Data. 2013 Presented at: 2013 International Conference on Social Computing; September 8-14, 2013; Alexandria, VA, USA p. 790-795. [doi: 10.1109/socialcom.2013.118]
18. Bogomolov A, Lepri B, Ferron M, Pianesi F. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. 2014 Presented at: MM '14: Proceedings of the 22nd ACM international conference on Multimedia; November 3-7, 2014; Orlando, FL, USA p. 477-486. [doi: 10.1145/2647868.2654933]
19. Walker WH, Walton JC, DeVries AC, Nelson RJ. Circadian rhythm disruption and mental health. Transl Psychiatry 2020 Jan 23;10(1):28 [FREE Full text] [doi: 10.1038/s41398-020-0694-0] [Medline: 32066704]
20. Khoulji S, Garzón-Rey J, Aguilo J. Remote Assessment of Disease and Relapse – Central Nervous System- RADAR-CNS. TMLAI 2017 Aug 31;5(4):1 [FREE Full text] [doi: 10.14738/tmlai.54.3293]
21. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort

study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
22. Hersenonderzoek.nl. URL: https://hersenonderzoek.nl [accessed 2020-07-20]
23. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Hyve, RADAR-CNS Consortium. RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
24. Kroenke K, Strine TW, Spitzer RL, Williams JBW, 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]
25. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
26. Lee RM, Draper M, Lee S. Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. Journal of Counseling Psychology 2001;48(3):310-318. [doi: 10.1037/0022-0167.48.3.310]
27. Aluoja A, Leinsalu M, Shlik J, Vasar V, Luuk K. Symptoms of depression in the Estonian population: prevalence, sociodemographic correlates and social adjustment. Journal of Affective Disorders 2004 Jan;78(1):27-35. [doi: 10.1016/s0165-0327(02)00179-9]
28. Akhtar-Danesh N, Landeen J. Relation between depression and sociodemographic factors. Int J Ment Health Syst 2007 Sep 04;1(1):4 [FREE Full text] [doi: 10.1186/1752-4458-1-4] [Medline: 18271976]
29. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
30. Farhan AA, Yue C, Morillo R, Ware S, Lu J, Bi J, et al. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. 2016 Presented at: 2016 IEEE Wireless Health (WH); October 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764553]
31. Costa M, Goldberger AL, Peng C. Multiscale entropy analysis of biological signals. Phys Rev E Stat Nonlin Soft Matter Phys 2005 Feb 18;71(2 Pt 1):021906. [doi: 10.1103/PhysRevE.71.021906] [Medline: 15783351]
32. Broughton S, Bryan K. Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing. Hoboken, NJ, USA: Wiley; 2008.
33. Silva LEV, Cabella BCT, Neves UPDC, Murta Junior LO. Multiscale entropy-based methods for heart rate variability complexity analysis. Physica A: Statistical Mechanics and its Applications 2015 Mar;422:143-152. [doi: 10.1016/j.physa.2014.12.011]
34. Mizuno T, Takahashi T, Cho RY, Kikuchi M, Murata T, Takahashi K, et al. Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy. Clin Neurophysiol 2010 Sep;121(9):1438-1446 [FREE Full text] [doi: 10.1016/j.clinph.2010.03.025] [Medline: 20400371]
35. Costa M, Peng C, L. Goldberger A, Hausdorff JM. Multiscale entropy analysis of human gait dynamics. Physica A: Statistical Mechanics and its Applications 2003 Dec;330(1-2):53-60. [doi: 10.1016/j.physa.2003.08.022]
36. Busa MA, van Emmerik RE. Multiscale entropy: A tool for understanding the complexity of postural control. J Sport Health Sci 2016 Mar;5(1):44-51 [FREE Full text] [doi: 10.1016/j.jshs.2016.01.018] [Medline: 30356502]
37. Mizuno K. Human circadian rhythms and exercise: Significance and application in real-life situations. JPFSM 2014;3(3):307-315. [doi: 10.7600/jpfsm.3.307]
38. Shannon CE. A Mathematical Theory of Communication. The Bell System Technical Journal 1948;27(3):379-423. [doi: 10.1002/j.1538-7305.1948.tb01338.x]
39. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982 Dec;38(4):963-974. [Medline: 7168798]
40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
41. Glover S, Dixon P. Likelihood ratios: a simple and flexible statistic for empirical psychologists. Psychon Bull Rev 2004 Oct;11(5):791-806. [doi: 10.3758/bf03196706] [Medline: 15732688]
42. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
43. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian Data Analysis. New York, NY: Chapman and Hall/CRC; 2013.
44. Salvatier J, Wiecki T, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2016;2:e55 [FREE Full text] [doi: 10.7717/peerj-cs.55]
45. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
46. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA, USA p. 785-794. [doi: 10.1145/2939672.2939785]

47. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
48. Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Vedel Kessing L, Winther O. Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15028 [FREE Full text] [doi: 10.2196/15028] [Medline: 32234702]
49. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2021-07-20]
50. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using Mobile Sensing to Test Clinical Models of Depression, Social Anxiety, State Affect, and Social Isolation Among College Students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
51. Zhang Y, Folarin AA, Sun S, Cummins N, Bendayan R, Ranjan Y, RADAR-CNS Consortium. Relationship Between Major Depression Symptom Severity and Sleep Collected Using a Wristband Wearable Device: Multicenter Longitudinal Observational Study. JMIR Mhealth Uhealth 2021 Apr 12;9(4):e24604 [FREE Full text] [doi: 10.2196/24604] [Medline: 33843591]
52. Goldberg B, Brintnell ES, Goldberg J. The Relationship Between Engagement in Meaningful Activities and Quality of Life in Persons Disabled by Mental Illness. Occupational Therapy in Mental Health 2002 Mar;18(2):17-44. [doi: 10.1300/j004v18n02_03]
53. Dearmon J, Smith TE. Gaussian Process Regression and Bayesian Model Averaging: An Alternative Approach to Modeling Spatial Phenomena. Geogr Anal 2015 Aug 21;48(1):82-111. [doi: 10.1111/gean.12083]

Abbreviations

CIBER: Centro de Investigación Biomédican en Red
FD: frequency domain
HF: high frequency
KCL: King's College London
LAO: leave-all-out
LF: low frequency
LOO: leave-one-out
MAC: Media Access Control
MF: middle frequency
MSE: multiscale entropy
NBDC: nearby Bluetooth device count
PHQ-8: 8-item Patient Health Questionnaire
RADAR-CNS: Remote Assessment of Disease and Relapse - Central Nervous System
RADAR-MDD: Remote Assessment of Disease and Relapse - Major Depressive Disorder
RMSE: root mean squared error
RMT: remote measurement technology
VUmc: Vrije Universiteit Medisch Centrum

Edited by G Eysenbach; submitted 23.04.21; peer-reviewed by A Lamer; comments to author 14.05.21; revised version received 18.05.21; accepted 31.05.21; published 30.07.21.

Please cite as:
Zhang Y, Folarin AA, Sun S, Cummins N, Ranjan Y, Rashid Z, Conde P, Stewart C, Laiou P, Matcham F, Oetzmann C, Lamers F, Siddi S, Simblett S, Rintala A, Mohr DC, Myin-Germeys I, Wykes T, Haro JM, Penninx BWJH, Narayan VA, Annas P, Hotopf M, Dobson RJB, RADAR-CNS Consortium
Predicting Depressive Symptom Severity Through Individuals' Nearby Bluetooth Device Count Data Collected by Mobile Phones: Preliminary Longitudinal Study
JMIR Mhealth Uhealth 2021;9(7):e29840
URL: https://mhealth.jmir.org/2021/7/e29840
doi:10.2196/29840
PMID:34328441

Original Paper

Early Acceptability of a Mobile App for Contact Tracing During the COVID-19 Pandemic in France: National Web-Based Survey

Rajae Touzani[1,2], MSc; Emilien Schultz[1,3], PhD; Seth M Holmes[2,4,5,6,7], MD, PhD; Stéphanie Vandentorren[8], MD, PhD; Pierre Arwidson[8], MD; Francis Guillemin[9], PhD; Dominique Rey[1], MD; Alexandra Rouquette[10,11], PhD; Anne-Déborah Bouhnik[1], PhD; Julien Mancini[1,12], MD, PhD

[1]Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, ISSPAM, Equipe CANBIOS Labellisée Ligue Contre le Cancer, Marseille, France

[2]Institut Paoli-Calmettes, SESSTIM U1252, Marseille, France

[3]CEPED, Université de Paris, IRD, Paris, France

[4]IMERA Mediterranean Institute for Advanced Study, Marseille, France

[5]Society and Environment, Rausser College of Natural Resources, University of California, Berkeley, Berkeley, CA, United States

[6]Medical Anthropology, Department of Anthropology, University of California, Berkeley, Berkeley, CA, United States

[7]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[8]Santé Publique France, Saint-Maurice, France

[9]CIC 1433 Clinical Epidemiology, Inserm, CHRU, University of Lorraine, Nancy, France

[10]Public Health and Epidemiology Department, AP-HP, Bicêtre Hôpitaux Universitaires Paris Sud, Le Kremlin-Bicêtre, France

[11]Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM U1018, Villejuif, France

[12]APHM, Hop Timone, BioSTIC, Biostatistique et Technologies de l'Information et de la Communication, Marseille, France

Corresponding Author:
Rajae Touzani, MSc
Institut Paoli-Calmettes, SESSTIM U1252
232 Boulevard de Sainte-Marguerite
Marseille, 13009
France
Phone: 33 491223502
Email: rajae.touzani@inserm.fr

Abstract

Background: Several countries have implemented mobile apps in an attempt to trace close contacts of patients with COVID-19 and, in turn, reduce the spread of SARS-CoV-2. However, the effectiveness of this approach depends on the adherence of a large segment of the population.

Objective: The aims of this study were to evaluate the acceptability of a COVID-19 contact tracing mobile app among the French population and to investigate the barriers to its use.

Methods: The Health Literacy Survey 2019 questioned 1003 people in France during the COVID-19 pandemic on the basis of quota sampling. The survey collected sociodemographic characteristics and health literacy data, as well as information on participants' communication with caregivers, trust in institutions, and COVID-19 knowledge and preventive behaviors. The acceptability of a mobile app for contact tracing was measured by a single question, the responses to which were grouped into three modalities: app-supporting, app-willing, and app-reluctant. Multinomial logistic regression analysis was performed to identify the factors associated with the acceptability of a mobile app during the COVID-19 pandemic.

Results: Only 19.2% (193/1003) of all participants were app-supporting, whereas half of them (504/1003, 50.3%) were reluctant. The factors associated with willingness or support toward the contact tracing app included lower financial deprivation (app-willing: adjusted odds ratio [aOR] 0.8, 95% CI 0.69-0.93; app-supporting: aOR 0.7, 95% CI 0.58-0.84) and higher perceived usefulness of using a mobile app to send completed health questionnaires to doctors (app-willing: aOR 2.3, 95% CI 1.70-3.26; app-supporting: aOR 3.1, 95% CI 2.04-4.82). Furthermore, the likelihood of supporting the mobile app increased with age over 60 years (aOR 1.9, 95% CI 1.13-3.22), trust in political representatives (aOR 2.7, 95% CI 1.72-4.23), feeling concerned about the pandemic situation (aOR 2.2, 95% CI 1.47-3.32), and knowledge about the transmission of COVID-19 (aOR 2.0, 95% CI 1.39-2.96).

Conclusions: The most socioeconomically precarious people, who are at a higher risk of SARS-CoV-2 infection, are also the most reluctant to using a contact tracing mobile app. Therefore, optimal adherence can only be effective with a targeted discourse on public health benefits to adopt such an app, which should be combined with a reduction in inequalities by acting on structural determinants.

(JMIR Mhealth Uhealth 2021;9(7):e27768) doi:10.2196/27768

KEYWORDS

COVID-19; mobile app; contact tracing; HLS19; health care disparities; public health

Introduction

Monitoring contacts of patients with COVID-19 is a key issue for long-term control of the pandemic. Several digital tools and eHealth applications have been deployed to effectively support health care systems in these efforts [1]. In particular, contact tracing apps have been designed to identify and inform people of likely exposure [2]. Since March, such contact tracing apps were installed by approximately 9.3% of the population around the world [3] and adopted by many countries, including China, South Korea, Australia, Turkey, Germany, Israel, and Singapore [3], with mandatory or voluntary setups. In France, the governmental decision—following the advice of its scientific board—was to develop a dedicated app called "StopCovid"; this app has been available for download on mobile phones since early June 2020 [4]. In October 2020, the StopCovid app was updated and renamed "TousAntiCovid" [5]. When app users indicate that they have been infected with COVID-19, the app uses Bluetooth technology to recover information of all close contacts (ie, all other TousAntiCovid users who have spent more than 15 minutes within a distance of 1 meter of the said user, as recorded over a period of 14 days) and alerts them with generic notifications, recommending them to quarantine themselves and to take a COVID-19 screening test. There have been numerous criticisms focused on data privacy [6,7] and technical limitations of the app software [8]. Because downloading and using the TousAntiCovid app is voluntary in France, its effectiveness is dependent on acceptability and adoption among users.

Beyond the technical considerations, such contact tracing apps require high adherence rates among the population to be effective. According to a previous study, at least 56% of a population must use a digital contact tracing app in order to control the pandemic [9]. However, another study reported that app-based tracing was more efficient than conventional tracing even with 20% coverage [2]. It is therefore necessary to convince a maximum number of citizens of their interest in using such an app and to remove potential barriers to app use. However, the factors that determine the acceptance of contact tracing apps remain largely unknown. Beyond the lack of information as a barrier, social deprivation that increases risk and severity of COVID-19 might also relate to potential barriers to the use of a contact tracing app [10]. For this reason, it is also important that the introduction of the apps does not inadvertently create or exacerbate social inequalities [11].

The objective of this study is to determine the acceptability of a contact tracing mobile app, as well as the potential barriers to its use.

Methods

The data analyzed here has been sourced from the French Health Literacy Survey 2019 (HLS19), which was set up as part of France's participation in the World Health Organization Action Network project Measuring Population and Organizational Health Literacy (M-POHL) [12]. HLS19 was administered to a sample of the adult French population 2 weeks after the end of the first lockdown in France (ie, between May 27 and June 5, 2020) at the time of the official launch of the TousAntiCovid app (June 2, 2020). After this complete lockdown, several other national or regional lockdown periods were announced in France in 2020-2021. The sample of 1000 internet users aged 18-75 years was drawn from an access panel, respecting the French population structure for sex, age, regions, and area of residence (urban or rural). After informed consent was obtained, the web-based survey collected information on respondents' sociodemographic characteristics (gender, age, level of education, and region of residence), perception of health, health literacy, and navigation in the health system. In addition, the survey collected additional national items such as communication with caregivers through new technologies, perception of medical research, trust in institutions, as well as data about the COVID-19 pandemic. The study methodology was reviewed and approved by the Ethics Evaluation Committee of the French National Health and Medical Research Institute (CEEI, IRB 00003888).

In this study, we focused on analyzing the sociodemographic data of the survey respondents. A financial deprivation score was calculated by combining 3 questions related to financial abilities: (1) to pay all bills at the end of the month, (2) to buy drugs if needed, and (3) to pay for medical examinations and treatments that are not covered by health insurance (response scale: very easy, easy, difficult, and very difficult). The mean score was transformed by multiplying it by (5/3) [13]. The scores ranged from 0 to 5, with higher scores indicating higher financial deprivation [13].

Health literacy level was calculated using the 16-item version of the European Health Literacy Survey Questionnaire (HLS-EU-Q16). Responses to the survey items were recorded using a 4-point Likert-type scale [14]. For the score, the modalities were dichotomized into easy and difficult categories, by merging "very easy" and "easy" responses, as well as "difficult" and "very difficult" responses. This allowed us to have 3 groups according to the health literacy score: inadequate (HLS-EU score: <9), problematic (HLS-EU score: 9-12), and adequate (HLS-EU score: >12), with a maximum possible score of 16 [14].

The level of knowledge on the transmission of COVID-19 (KT-COVID-19) was measured by combining the correct answers to the following two questions: (1) "In your opinion, can someone who shows no sign of the disease transmit the coronavirus?" and (2) "In your opinion, are protective behaviors effective in limiting the spread of the coronavirus?" A perfect level of KT-COVID-19 involved answering "Yes, definitely" to both questions.

Trust in institutions was measured using three questions asking about the participants' level of trust in doctors, scientists, and political representatives. For each question, the responses were grouped into two modalities: yes (very and rather trustworthy) and no (rather not and not at all trustworthy).

The participants' acceptability of a mobile app to follow close contact between people during an epidemic was measured by the following question: "In your opinion, is it acceptable to use mobile phones to study close contact between people during an epidemic?" with 5 possible modalities: "Yes, definitely," "Yes, probably," "No, probably not," "No, definitely not," and "I don't know." We grouped the last three modalities as *app-reluctant* and named the other two modalities *app-willing* ("Yes, probably") and *app-supporting* ("Yes, definitely").

Chi-square tests and analysis of variance (ANOVA) were used for descriptive analyses, depending on the type of variables. A multinomial logistic model was used to identify the factors associated with the acceptability of a mobile app to study close contact between people during an epidemic. After adjustment for age, a stepwise procedure was performed to select significant factors in the model (entry threshold; $P<.20$). The significance threshold in multivariate analyses was set at 5%. All analyses were performed using the STATA software program (version 14.0; StataCorp LLC).

Results

Description of the Study Sample

A description of our study sample is presented in Table 1. Overall, 1003 French adults responded to the survey. Half of the participants were women (n=506, 50.4%), and half (n=510, 50.9%) were 46 years old or younger. A little more than half of all participants (n=592, 59.0%) had a high level of education. A small proportion of participants reported they had been infected with COVID-19 themselves (n=50, 5.0%) or that someone in their household had been infected (n=58, 5.8%). In terms of health literacy, 39.8% (n=399) of the participants had an HLS-EU-Q16 score reflecting problematic (n=264, 26.3%) or inadequate (n=135, 13.5%) levels (ie, HLS-EU score: ≤12). Of note, most respondents (n=664, 66.2%) had an imperfect level of KT-COVID-19. With regard to our variable of interest, 50.3% (n=504) of all participants were app-reluctant ("No, probably not acceptable": n=183, 18.3%; "No, definitely not acceptable": n=216, 21.5%; and "Don't know": n=105, 10.5%) and only 19.2% (n=193) were app-supporting ("Yes, definitely acceptable").

Table 1. Description of the characteristics of the study sample (N=1003).

Characteristic	Participants
Age (years), n (%)	
18-35	291 (29.0)
36-46	219 (21.9)
47-59	253 (25.2)
60-75	240 (23.9)
Gender, n (%)	
Women	506 (50.4)
Men	497 (49.6)
Place of birth, n (%)	
France	951 (94.8)
Other country	52 (5.2)
Education level, n (%)	
Less than high-school degree	172 (17.2)
High-school degree	239 (23.8)
Higher education	592 (59.0)
Area of residence, n (%)	
Rural	212 (21.1)
Urban	791 (78.9)
Financial deprivation score (range: 0-5)[a], mean (SD)	1.7 (1.1)
Health literacy level (HLS-EU-Q16[b] score range: 0-16), n (%)	
Inadequate (<9)	135 (13.5)
Problematic (9-12)	264 (26.3)
Adequate (>12)	604 (60.2)
Infected with COVID-19, n (%)	
Yes	50 (5.0)
No	862 (85.9)
I don’t know	91 (9.1)
A household member infected with COVID-19, n (%)	
Yes	58 (5.8)
No	851 (84.8)
I don’t know	94 (9.4)
Significantly concerned about the situation caused by COVID-19, n (%)	
Yes	266 (26.5)
No	737 (73.5)
KT-COVID-19[c] level, n (%)	
Perfect	339 (33.8)
Imperfect	664 (66.2)
Acceptability of a mobile app to study close contact between people during an epidemic[d], n (%)	
Yes, definitely	193 (19.2)
Yes, probably	306 (30.5)
No, probably not	183 (18.3)

Characteristic	Participants
No, definitely not	216 (21.5)
Don't know	105 (10.5)

[a]A financial deprivation score was calculated by combining answers to 3 questions. Higher scores indicate higher financial deprivation.

[b]HLS-EU-Q16: European Health Literacy Survey Questionnaire, 16-item.

[c]KT-COVID-19: knowledge on the transmission of COVID-19 (measured by combining answers to 2 questions on the transmission of COVID-19 and on the effectiveness of barrier gestures; a perfect level of KT-COVID-19 corresponds to a correct answer to both questions).

[d]This variable has been grouped into 3 modalities: app-supporting ("Yes, definitely"), app-willing ("Yes, probably"), and app-reluctant ("No, probably not"; "No, definitely not"; and "Don't know").

Factors Associated With the Acceptability of Contact Tracing Using Mobile Phones

Results of the univariate analysis (Table 2) showed no significant association between app-reluctance and sociodemographic factors except for financial deprivation, with participants reluctant to use such apps reporting higher financial deprivation scores (P<.001).

App-supporting participants (75/193, 38.9%) felt significantly more concerned by the situation caused by COVID-19 than app-willing participants (80/306, 26.1%), who in turn felt more concerned than app-reluctant participants (111/504, 22.0%; P<.001). The positive gradient was also observed between the acceptability of contact tracing apps and avoiding close contact in the past week with people they did not live with (app-reluctant 334/504, 66.3%; app-willing: 229/306, 74.8%; app-supporting: 147/193, 76.2%; P=.006). The level of KT-COVID-19 was also significantly associated with participants' attitudes toward such apps; for instance, 53.4% (103/193) of app-supporting participants had a perfect KT-COVID-19 level compared to 29.4% (148/504) of app-reluctant participants (P<.001).

Moreover, trust in political representatives (P<.001), scientists (P=.02), and doctors (P=.006) was positively associated with the acceptability of a contact tracing app during a pandemic. We also observed a positive association concerning the perceived usefulness of digital technologies: during medical consultations (broadcasting and recording), to complete and send health assessment questionnaires, or to make a medical appointment (all P<.001).

After adjusting for age (Figure 1), the two groups not reluctant to use a contact tracing app were found to be associated with a lower level of financial deprivation and with higher perceived usefulness of a mobile app to send doctors answers to health questionnaires. The likelihood of a participant's willingness to use a contact tracing app increased among those who trusted doctors and those who had avoided close contact with other people in the past week. App-supporters were 60 years and older, felt more concerned about the situation of the COVID-19 pandemic, trusted political representatives, and had a perfect level of KT-COVID-19.

Table 2. Factors associated with the acceptability of a mobile app to study close contact between people during an epidemic: univariate analysis (N=1003).

Factor	App-reluctant (n=504)	App-willing (n=306)	App-supporting (n=193)	*P* value[a]
Age (years), n (%)				.11
18-35	146 (29.0)	95 (31.1)	50 (25.9)	
36-46	120 (23.8)	54 (17.7)	45 (23.3)	
47-59	134 (26.6)	76 (24.8)	43 (22.3)	
60-75	104 (20.6)	81 (26.5)	55 (28.5)	
Gender, n (%)				.35
Women	265 (52.6)	145 (47.4)	96 (49.7)	
Men	239 (47.4)	161 (52.6)	97 (50.3)	
Place of birth, n (%)				.16
France	476 (94.4)	287 (93.8)	188 (97.4)	
Other country	28 (5.6)	19 (6.2)	5 (2.6)	
Education level, n (%)				.73
Less than high-school degree	90 (17.9)	55 (18.0)	27 (14.0)	
High-school degree	122 (24.2)	72 (23.5)	45 (23.3)	
Higher education	292 (57.9)	179 (58.5)	121 (62.7)	
Area of residence, n (%)				.26
Rural	117 (23.2)	57 (18.6)	38 (19.7)	
Urban	387 (76.8)	249 (81.4)	155 (80.3)	
Financial deprivation score (range: 0-5)[b], mean (SD)	1.9 (1.1)	1.6 (0.9)	1.4 (1.1)	*<.001*
Health literacy level (HLS-EU-Q16[c] score range: 0-16), n (%)				*.046*
Inadequate (<9)	76 (15.1)	38 (12.4)	21 (10.9)	
Problematic (9-12)	140 (27.8)	65 (21.2)	59 (30.6)	
Adequate (>12)	288 (57.1)	203 (66.3)	113 (58.5)	
Infected with COVID-19, n (%)				.83
Yes	22 (4.4)	18 (5.9)	10 (5.2)	
No	438 (86.9)	261 (85.3)	163 (84.5)	
I don’t know	44 (8.7)	27 (8.8)	20 (10.4)	
A household member infected with COVID-19, n (%)				.49
Yes	23 (4.6)	23 (7.5)	12 (6.2)	
No	431 (85.5)	257 (84.0)	163 (84.5)	
I don’t know	50 (9.9)	26 (8.5)	18 (9.3)	
Significantly concerned about the situation caused by COVID-19, n (%)				*<.001*
Yes	111 (22.0)	80 (26.1)	75 (38.9)	
No	393 (78.0)	226 (73.9)	118 (61.1)	
Trust in political representatives, n (%)				*<.001*
Yes	62 (12.3)	56 (18.3)	61 (31.6)	
No	442 (87.7)	250 (81.7)	132 (68.4)	
Trust in scientists, n (%)				*.02*
Yes	438 (86.9)	278 (90.9)	181 (93.8)	
No	66 (13.1)	28 (9.1)	12 (6.2)	
Trust in doctors, n (%)				*.006*

Factor	App-reluctant (n=504)	App-willing (n=306)	App-supporting (n=193)	*P* value[a]
Yes	457 (90.7)	293 (95.8)	185 (95.9)	
No	47 (9.3)	13 (4.2)	8 (4.1)	
Having avoided close contact (<1 m) in the last week with people not living with you, n (%)				*.006*
Yes	334 (66.3)	229 (74.8)	147 (76.2)	
No	170 (33.7)	77 (25.2)	46 (23.8)	
Broadcasting of the consultation by video so that relatives who are not present can participate is useful, n (%)				*<.001*
Yes	162 (32.2)	136 (44.4)	95 (49.2)	
No	342 (67.9)	170 (55.6)	98 (50.8)	
Video recording of consultations to remember its content is useful, n (%)				*<.001*
Yes	201 (39.9)	164 (53.6)	118 (61.1)	
No	303 (60.1)	142 (46.4)	75 (38.9)	
Mobile app for scheduling medical appointments and reminders is useful, n (%)				*<.001*
Yes	393 (78.0)	269 (87.9)	170 (88.1)	
No	111 (22.0)	37 (12.1)	23 (11.9)	
Mobile app for sending your doctor answers to questionnaires assessing your health is useful, n (%)				*<.001*
Yes	284 (56.3)	225 (73.5)	153 (79.3)	
No	220 (43.7)	81 (26.5)	40 (20.7)	
KT-COVID-19[d] level, n (%)				*<.001*
Perfect	148 (29.4)	88 (28.8)	103 (53.4)	
Imperfect	356 (70.6)	218 (71.2)	90 (46.6)	

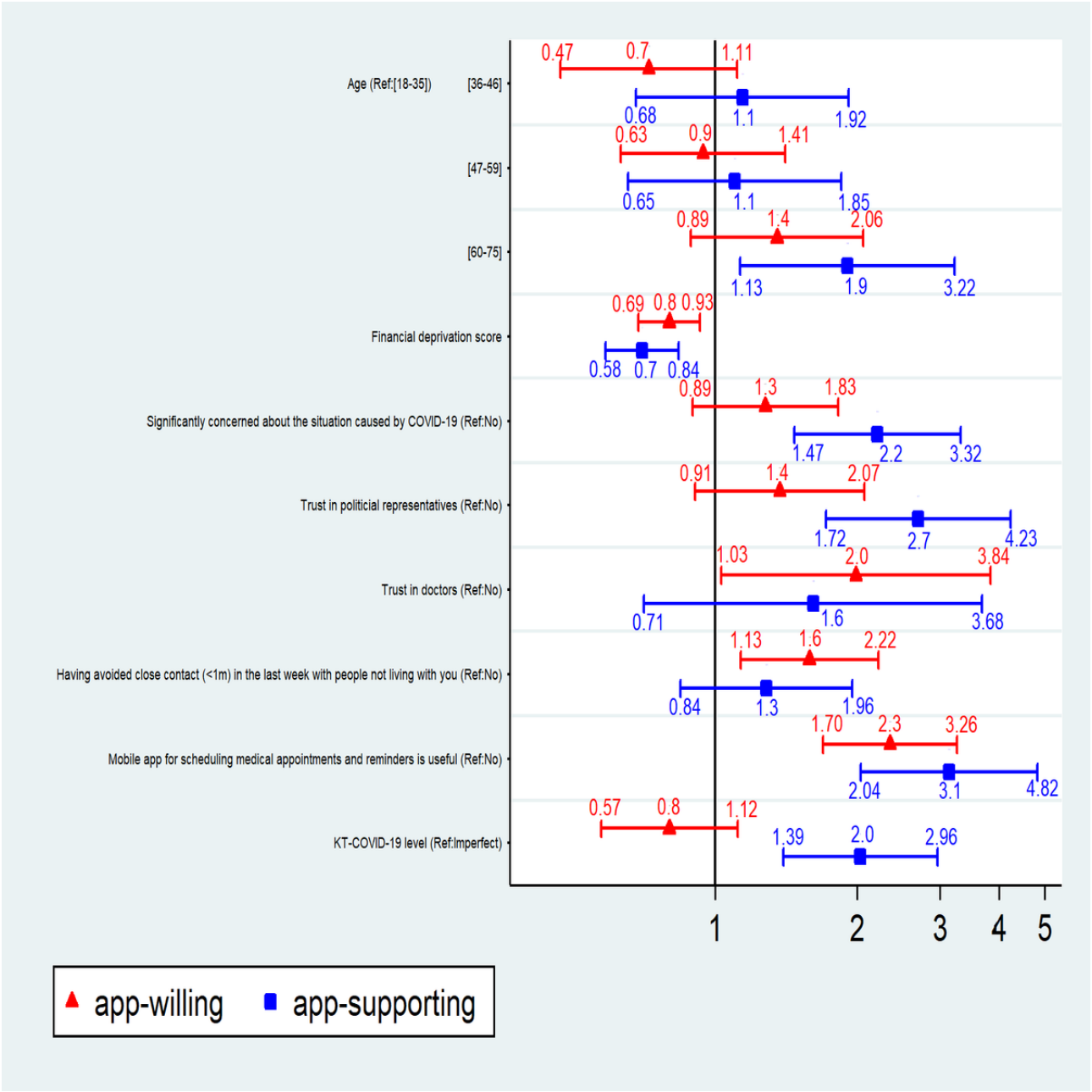
[a]Italicized values indicate statistical significance.

[b]A financial deprivation score calculated by combining answers to 3 questions. Higher scores indicate higher financial deprivation.

[c]HLS-EU-Q16: European Health Literacy Survey Questionnaire, 16-item.

[d]KT-COVID-19: knowledge on the transmission of COVID-19 (measured by combining answers to 2 questions on the transmission of COVID-19 and on the effectiveness of barrier gestures; a perfect level of KT-COVID-19 corresponds to a correct answer to both questions).

Figure 1. Factors associated with acceptability of a mobile app to study close contact between people during an epidemic (N=1003). Dots and whiskers represent adjusted odds ratio and 95% CI levels from multivariate analyses for definite ("app-supporting," blue) or probable ("app-willing," red) acceptability of a contact tracing app. KT-COVID-19: knowledge on the transmission of COVID-19; Ref: referent category in the model.



Discussion

Principal Findings

Even amidst vaccine rollout and with a significant risk of multiple epidemic waves or a possible shift toward a long-term pandemic, contact tracing is an important public health strategy to control the spread of the COVID-19 pandemic [15], and it can be effective if it is adopted by a large segment of the population. In our study, the proportion of people who would agree to use such a mobile app for tracing and follow-up of close contacts in the context of COVID-19 was found to be 49.7% (499/1003), including 306 participants responding "Yes, probably" and 193, "Yes, definitely".

Similar figures were noted before the implementation of TousAntiCovid app in a sample of the French population aged over 15 years (N=1051), wherein 49% of the study population indicated that they intended to install the app, of which only 15% were certain and 34% probably claimed to do so [16]. An acceptability rate of 38.4% was also reported in a recent study (May 07, 2020, N=1849) [17]. Despite relatively high theoretical acceptability, after the launch of the mobile app, only 3.1% of the people in France downloaded it (as of mid-July 2020) [3], as confirmed by 73.5% (737/1003) of our study participants reporting that they did not feel very concerned about the situation caused by COVID-19. Additional recent data further confirm the low adoption of TousAntiCovid [18], downloaded by approximately 20% of the French population in March 2021 [19]. Moreover, we do not know if this population indeed used

the app after downloading it, as no data are available regarding use and uninstallation. Recent results from the literature confirm this difference between the actual and theoretical rate of app use in the case of voluntary installation in various countries, including Australia (70% agreed to use the app vs 44% downloaded it) [20], United States (55% accepted vs 21% downloaded it) [21], and Germany (81% accepted vs 36% downloaded it) [22]. In countries where the installation of the app was mandatory, we observed a very high download rate (eg, 91% in Qatar) [23]. Several reasons might explain this large gap between theoretical acceptability and actual use. For instance, technical and financial considerations might limit access to the app. Therefore, special efforts are needed to help people who are more financially precarious; it may also be important to better inform people about the availability of such a tool and its usefulness in controlling the pandemic. Indeed, apart from the content and availability of the messages, many factors can be associated with app reluctance, such as the hierarchy of needs (food or administrative insecurity), discrimination and fear of losing one's job, worries about being isolated, and losing social support. To improve the acceptability of such a tool, special attention must be paid to protect people against the risks and difficulties they might be facing.

Several factors were identified to play a role in the potential acceptability of a contact tracing app. Our results show that people aged 60-75 years would support such an app during the COVID-19 pandemic, contrary to what has been demonstrated in the literature with regard to technology acceptance, where advanced age is associated with a lower level of acceptability of mobile apps [24]. However, this finding is consistent with early evidence on the acceptability of contact tracing apps reported by a German study [25]. Indeed, in the context of the COVID-19 pandemic, the population at higher risk is typically over 60 years old [26], and people who perceive themselves at risk might be more inclined to adopt such an app. Our results confirm that agreeing to use the contact tracing app requires people to be concerned about the situation created by the pandemic and/or to avoid close contacts outside their household. At the start of the pandemic, the communication by the French government and the scientific community sometimes included contradictory messages that led to confusion (eg, whether or not to wear a mask, COVID-19 is not an alarming virus) [27]. Because contact tracing has been framed as a long-term solution, it could have also been seen as less essential than treatments or wearing masks. Contradictory communication likely played a role in the disinterest on the part of the population and the lack of collective awareness. Indeed, knowledge about COVID-19 transmissibility through asymptomatic individuals and about the effectiveness of barrier gestures was low in our sample (339/1003, 33.8%).

Another aspect to consider here is people's trust in institutions. Contact app tracing is intrinsically linked to the state response to the pandemic. As such, attitudes toward institutions should influence the acceptability of an app. Our analysis shows that trusting politicians and doctors has a positive effect on people's intention to use the contact tracing mobile app. Those who trust political representatives' express absolute certainty, unlike those who trust doctors who are almost certain. A study conducted on the general population during the COVID-19 lockdown showed that one of the reasons for nonadherence with the app was the concern of the use of this technology for the purposes of government oversight [28]. In this study, we do not know if this is the reason people mistrust political representatives, but we assume that the issue of personal data privacy plays a role in the negative relationship between trust in the government and likelihood of installing the app [29], particularly in a context where individual rights are more generally restricted. Moreover, the conflicting messages from the government (including the fact that the Prime Minister and two other ministers of the French government had not downloaded the TousAntiCovid app [30]) and the medical controversies about drug treatment during the COVID-19 pandemic have further weakened the population's trust in institutions. These factors seem to have served to enrich conspiracy theories related to COVID-19. Marinthe et al [31] demonstrated that people with a conspiracy mentality are less willing to comply with government-driven preventive behaviors during the COVID-19 outbreak. These results are in line with those obtained in a French study [17], unlike an American study where trust in politicians was not correlated with app use [21]. These authors explained this difference by the fact that at the time of their study, various American political parties supported the use of the app.

Lack of familiarity with technologies is also an obstacle to using a contact tracing app. Our results show that app-reluctant individuals are those who consider that electronic patient-reported outcomes (ePRO) are not useful. One study indicated that patients with cancer who did not use ePROs were those who expressed lower acceptability for mobile technology and, therefore, lower adherence to mobile health [32]. The positive link between attitude toward technologies and its acceptance has previously been demonstrated in the context of the COVID-19 pandemic [33]. There is a need to better understand the link between the public understanding of science and technology, their overall acceptability, and the consequences for health practices, especially in an emergency context where uncertainty prevails.

Our results also show that the likelihood of agreeing to use mobile tracking technology during the COVID-19 pandemic increases with financial resources and health literacy. At the same time, the people most vulnerable to COVID-19 are also those whose precariousness is the most marked by their socioeconomic situation (income, professional activity, and origin) with frequent inadequate levels of health literacy [34]. We then assume that people with lower health literacy might have different risk perceptions and are probably not reached by adequate and understandable preventive recommendations. Several studies have shown that people with low health literacy have more difficulty finding information and understanding COVID-19–related messages [35,36]. Our results are in line with these previous findings. Indeed, our findings show clearly that adoption of a behavior of social utility—in this case, using a contact tracing app—was associated with a perfect level of knowledge of the modes of transmission of COVID-19. Studies have also shown that digital solutions are often less used by people with low levels of health literacy or those who do not have access to the internet [37,38]. More socioeconomically

precarious people often have poor internet access and may face difficult living conditions (housing conditions, employment, access to running water, etc) that could affect their ability to follow the recommended guidelines. Hence, unequal access to information and its understanding, as well as the level of risk perception, could play a major role in citizens' adherence to new COVID-19 prevention technologies. It is therefore clear that COVID-19 highlights pre-existing health inequalities and can even accentuate them through the deployment of the contact tracing app. This is why specific interventions are needed to fight against these inequalities. Targeting this group of the population with understandable messages on the usefulness of the app and addressing the specific reasons for their reluctance might help increase its use, but it is clearly not enough. More should be done to decrease inequities overall and specifically in relation to the implementation of such apps. Broader approaches that intervene not only on individual and interpersonal factors but also upon structural determinants are needed [39]. In particular, structural interventions should at longer term seek to change the global context that fosters social and health inequalities. If such contact tracing apps are needed to fight against pandemics, every willing person should be able to access such technology and to use it without potential negative consequences.

Limitations and Strengths

Our study has some limitations. The use of the close contact tracing app was studied through theoretical acceptability. This type of study upstream of the implementation of a new technology can help adapt the actions but should be followed, after the implementation of the app, by additional studies on real acceptability and usage. Moreover, we do not know whether participants have a smartphone with adequate data allowance and battery performance to use the app; despite these reasons, our web-based survey was conducted among a highly educated sample (59% in this study compared to 37% in the French population [40]) of internet users, which is not representative of the entire French population because there are still numerous households that do not have internet access.

Finally, one of the strengths of this study is the rather large size of the sample and the national representativeness with regard to age group, gender, area of residence, and population density, which was achieved by quota sampling. This large sample size gave us enough statistical power to detect the detrimental effect of financial deprivation and imperfect health literacy despite surveying a socioeconomically privileged sample.

Conclusions

To our knowledge, this is the first study in France to evaluate the impact of financial deprivation on the acceptability of a close contact tracing mobile app used during the COVID-19 outbreak on a representative quota sample. Knowing the characteristics of the people who do not adhere to the new tracking technology in the context of a pandemic is essential to adopting effective strategies. Combining the tracing tool with testing and isolation can significantly facilitate the fight against the virus. Strong adherence to this technology would not be possible if public authorities do not conduct extensive public awareness campaigns to foster trust in institutions and to clarify what the app does, and importantly, what it does not do, particularly among more precarious people. Globally, the current COVID-19 health crisis reinforces the need to fight against social inequities generally and specifically to provide to all people not only masks or vaccines but also technologies that are useful to control the spread of transmissible diseases.

Acknowledgments

We thank the Réseau Francophone de la Littératie en Santé (REFLIS) network for their collaboration in the European HLS19 survey. This work was funded by Santé Publique France and La Ligue Contre le Cancer (Equipe CANBIOS Labellisée). SH's collaboration with SESSTIM was supported through Institut Paoli-Calmettes/IMéRA (Aix-Marseille Univ) Chair.

Conflicts of Interest

None declared.

References

1. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
2. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health 2020 Aug;5(8):e452-e459. [doi: 10.1016/s2468-2667(20)30157-2]
3. Chan S. COVID-19 contact tracing apps reach 9% adoption in most populous countries. SensorTower. 2020 Jul 14. URL: https://sensortower.com/blog/contact-tracing-app-adoption [accessed 2021-06-09]
4. TousAnitCovid app. Webpage in French. 2020 Oct 22. URL: https://www.economie.gouv.fr/stopcovid [accessed 2021-06-09]
5. "TousAntiCovid", new version of the StopCovid app: what's changing? Webpage in French. Live Well Digital. 2020 Oct 22. URL: https://bienvivreledigital.orange.fr/sante/tousanticovid-nouvelle-version-de-lappli-stopcovid-quest-ce-qui-change/ [accessed 2021-06-09]

6. Bahrain, Kuwait and Norway contact tracing apps among most dangerous for privacy. Amnesty. 2020 Jun 16. URL: https://www.amnesty.org/en/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy/ [accessed 2021-06-09]
7. Saw YE, Tan EY, Liu JS, Liu JC. Predicting public uptake of digital contact tracing during the COVID-19 pandemic: results from a nationwide survey in Singapore. J Med Internet Res 2021 Feb 03;23(2):e24730 [FREE Full text] [doi: 10.2196/24730] [Medline: 33465034]
8. Horton O. French app StopCovid still facing hurdles amid EU concerns about data access. Radio France Internationale (RFI). 2020 Jun 17. URL: https://tinyurl.com/uzdjkhjv [accessed 2021-06-09]
9. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: a report to NHSX. 2020 Apr 16. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-06-09]
10. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436. [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
11. Thorneloe R, Epton T, Fynn W, Daly M, Stanulewicz N, Kassianos A, et al. Scoping review of mobile phone app uptake and engagement to inform digital contact tracing tools for COVID-19. PsyArXiv. Preprint posted online on April 25, 2020. [doi: 10.31234/osf.io/qe9b6]
12. M-POHL - Action Network on Measuring Population and Organizational Health Literacy of EHII - WHO-Europe. URL: https://m-pohl.net/ [accessed 2021-06-09]
13. Dietscher C, Pelikan J. The action network for measuring population and organizational health literacy (M-POHL) and its Health Literacy Survey 2019 (HLS19). Eur J Public Health 2019 Nov 13;29(4):ckz185.556. [doi: 10.1093/eurpub/ckz185.556]
14. Rouquette A, Nadot T, Labitrie P, Van den Broucke S, Mancini J, Rigal L, et al. Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. PLoS One 2018 Dec;13(12):e0208091 [FREE Full text] [doi: 10.1371/journal.pone.0208091] [Medline: 30521552]
15. Atlani-Duault L, Chauvin F, Yazdanpanah Y, Lina B, Benamouzig D, Bouadma L, et al. France's COVID-19 response: balancing conflicting public health traditions. The Lancet 2020 Jul;396(10246):219-221. [doi: 10.1016/s0140-6736(20)31599-3]
16. Martelli-Banégas D, Desreumaux M, Favré T. Questions spécifiques : perceptions de l'application STOPCOVID et regards sur l'enjeu du partage des données personnelles. Rapport de résultats – Mai 2020. Report in French. Harris Interactive. 2020 May. URL: https://harris-interactive.fr/wp-content/uploads/sites/6/2020/05/Observatoire_DATA_PUBLICA-Sondage_HI_mai_2020_vDEF.pdf [accessed 2021-06-09]
17. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
18. Lévèque L. Checknews - How many times has the StopCovid app been downloaded? How much does it cost? For what results? Webpage in French. Libération. 2020 Sep 18. URL: https://www.liberation.fr/checknews/2020/09/18/combien-de-fois-l-application-stopcovid-a-t-elle-ete-telechargee-combien-coute-t-elle-pour-quels-res_1799913 [accessed 2021-06-09]
19. Grably R. Since its launch, the TousAntiCovid application has alerted 100,000 contact cases. Webpage in French. BFM Business. 2021 Mar 10. URL: https://www.bfmtv.com/tech/depuis-son-lancement-l-application-tous-anti-covid-a-alerte-100-000-cas-contacts_AN-202103100026.html [accessed 2021-06-09]
20. Garrett PM, White JP, Lewandowsky S, Kashima Y, Perfors A, Little DR, et al. The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLoS One 2021;16(1):e0244827 [FREE Full text] [doi: 10.1371/journal.pone.0244827] [Medline: 33481841]
21. Hargittai E, Redmiles EM, Vitak J, Zimmer M. Americans' willingness to adopt a COVID-19 tracking app. First Monday 2020 Oct 06. [doi: 10.5210/fm.v25i11.11095]
22. Blom AG, Wenz A, Cornesse C, Rettig T, Fikel M, Friedel S, et al. Barriers to the large-scale adoption of a COVID-19 contact tracing app in Germany: survey study. J Med Internet Res 2021 Mar 02;23(3):e23362 [FREE Full text] [doi: 10.2196/23362] [Medline: 33577466]
23. Jacob S, Lawarée J. The adoption of contact tracing applications of COVID-19 by European governments. Policy Design and Practice 2020 Nov 28:1-15. [doi: 10.1080/25741292.2020.1850404]
24. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
25. Grill E, Eitze S, DeBock F, Dragano N, Huebl L, Schmich P, et al. Sociodemographic characteristics determine download and use of a corona contact tracing app in Germany - results of the COSMO surveys. PsychArchives. Preprint posted online on March 18, 2021. [doi: 10.23668/psycharchives.4713]

26. COVID-19: epidemiological update of August 27, 2020. Santé Publique France. Webpage in French. 2020. URL: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-27-aout-2020 [accessed 2021-06-17]
27. Schultz E, Ward JK, Atlani-Duault L, Holmes SM, Mancini J. French public familiarity and attitudes toward clinical research during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Mar 05;18(5):2611 [FREE Full text] [doi: 10.3390/ijerph18052611] [Medline: 33807787]
28. Milsom L, Abeler J, Altmann S, Toussaert S, Zillessen H, Blasone R. Survey of acceptability of app-based contact tracing in the UK, US, France, Germany and Italy. OSF. 2020 Mar 26. URL: https://osf.io/7vgq9/ [accessed 2021-06-09]
29. Why is the StopCovid app debated? Webpage in French. Le Journal du Net (JDN). 2020 May 13. URL: https://www.journaldunet.com/media/publishers/1491317-pourquoi-l-application-stopcovid-fait-elle-debat/ [accessed 2021-06-09]
30. Lausson J. Jean Castex estime qu'il n'a pas besoin d'utiliser l'application StopCovid. Webpage in French. Numerama. 2020 Sep 25. URL: https://www.numerama.com/politique/651121-jean-castex-estime-quil-na-pas-besoin-dutiliser-lapplication-stopcovid.html [accessed 2021-06-09]
31. Marinthe G, Brown G, Delouvée S, Jolley D. Looking out for myself: Exploring the relationship between conspiracy mentality, perceived personal risk, and COVID-19 prevention measures. Br J Health Psychol 2020 Nov;25(4):957-980 [FREE Full text] [doi: 10.1111/bjhp.12449] [Medline: 32583540]
32. Blake H. Mobile phone technology in chronic disease management. Nurs Stand 2008;23(12):43-46. [doi: 10.7748/ns2008.11.23.12.43.c6728] [Medline: 19093357]
33. Jansen S, Hurmuz M, Den Ouden M, van Velsen L. Predictors to use mobile apps for monitoring COVID-19 symptoms and contact tracing: a survey among Dutch citizens. medRxiv. Preprint posted online on June 02, 2020. [doi: 10.1101/2020.06.02.20113423]
34. Dubost CL, Pollak C, Rey S. Social inequalities in the face of the Covid-19 epidemic - State of play and perspectives. Webpage in French. DREES (Directorate of Research, Studies, Evaluation and Statistics). 2020 Jul. URL: https://drees.solidarites-sante.gouv.fr/publications/les-dossiers-de-la-drees/les-inegalites-sociales-face-lepidemie-de-covid-19-etat-des [accessed 2021-06-09]
35. McCaffery K, Dodd R, Cvejic E, Ayrek J, Batcup C, Isautier J, et al. Health literacy and disparities in COVID-19-related knowledge, attitudes, beliefs and behaviours in Australia. Public Health Res Pract 2020 Dec 09;30(4):30342012 [FREE Full text] [doi: 10.17061/phrp30342012] [Medline: 33294907]
36. Wolf MS, Serper M, Opsasnick L, O'Conor RM, Curtis LM, Benavente JY, et al. Awareness, attitudes, and actions related to COVID-19 among adults with chronic conditions at the onset of the U.S. outbreak: a cross-sectional survey. Ann Intern Med 2020 Jul 21;173(2):100-109 [FREE Full text] [doi: 10.7326/M20-1239] [Medline: 32271861]
37. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the internet as a resource for health information: results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
38. Triana AJ, Gusdorf RE, Shah KP, Horst SN. Technology literacy as a barrier to telehealth during COVID-19. Telemed J E Health 2020 Sep;26(9):1118-1119. [doi: 10.1089/tmj.2020.0155] [Medline: 32429770]
39. Brown AF, Ma GX, Miranda J, Eng E, Castille D, Brockie T, et al. Structural interventions to reduce and eliminate health disparities. Am J Public Health 2019 Jan;109(S1):S72-S78. [doi: 10.2105/ajph.2018.304844]
40. Teachers - Students per teaching staff. OECD (Organisation for Economic Co-operation and Development) Data. URL: http://data.oecd.org/teachers/students-per-teaching-staff.htm [accessed 2021-06-09]

Abbreviations

ANOVA: analysis of variance
ePRO: electronic patient-reported outcomes
HLS19: Health Literacy Survey 2019
HLS-EU-Q16: European Health Literacy Survey Questionnaire, 16-item
KT-COVID-19: knowledge on the transmission of COVID-19
M-POHL: Measuring Population and Organizational Health Literacy

Edited by L Buis; submitted 05.02.21; peer-reviewed by D Vollmer Dahlke, MDG Pimentel; comments to author 23.03.21; revised version received 12.04.21; accepted 27.05.21; published 19.07.21.

Please cite as:
Touzani R, Schultz E, Holmes SM, Vandentorren S, Arwidson P, Guillemin F, Rey D, Rouquette A, Bouhnik AD, Mancini J
Early Acceptability of a Mobile App for Contact Tracing During the COVID-19 Pandemic in France: National Web-Based Survey
JMIR Mhealth Uhealth 2021;9(7):e27768
URL: https://mhealth.jmir.org/2021/7/e27768
doi:10.2196/27768
PMID:34086589

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org