JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 10 (2022), Issue 3 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Corrigenda and Addenda

Viewpoint

Including the Public in Public eHealth: The Need for Community Participation in the Development of State-Sponsored COVID-19–Related Mobile Apps

Muhammed Yassin Idris[1,2], PhD; Maya Korin[1], MSc, PhD; Faven Araya[3], MPH; Sayeeda Chowdhury[1], BA; Patty Medina[1], MPH; Larissa Cruz[1], MPH; Trey-Rashad Hawkins[1], BS; Humberto Brown[3], MA; Luz Claudio[1], PhD

[1]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]Department of Medicine, Morehouse School of Medicine, Atlanta, GA, United States

[3]Arthur Ashe Institute for Urban Health, New York, NY, United States

Corresponding Author:
Muhammed Yassin Idris, PhD
Department of Medicine
Morehouse School of Medicine
720 Westview Dr SW
Atlanta, GA, 30310
United States
Phone: 1 404 752 1500
Email: myidris@msm.edu

Abstract

The COVID-19 pandemic has overwhelmed health care systems worldwide, particularly in underresourced communities of color with a high prevalence of pre-existing health conditions. Many state governments and health care entities responded by increasing their capacity for telemedicine and disease tracking and creating mobile apps for dissemination of medical information. Our experiences with state-sponsored apps suggest that because many of these eHealth tools did not include community participation, they inadvertently contributed to widening digital health disparities. We propose that, as eHealth tools continue to expand as a form of health care, more attention needs to be given to their equitable distribution, accessibility, and usage. In this viewpoint collaboratively written by a minority-serving community-based organization and an eHealth academic research team, we present our experience participating in a community advisory board working on the dissemination of the COVID Alert NY mobile app to illustrate the importance of public participation in app development. We also provide practical recommendations on how to involve community representatives in the app development process. We propose that transparency and community involvement in the process of app development ultimately increases buy-in, trust, and usage of digital technology in communities where they are needed most.

(JMIR Mhealth Uhealth 2022;10(3):e30872) doi:10.2196/30872

KEYWORDS

mobile apps; COVID-19; CBPR; digital health; eHealth; community health; health disparities

Introduction

Since the start of the COVID-19 outbreak, eHealth tools have been rapidly deployed, including telemedicine, mobile health apps, and wearable technologies [1,2]. These eHealth tools can be used to reduce and monitor disease. However, they may inadvertently result in increasing underlying inequalities by unevenly benefiting individuals who are better able to access new information, adopt technologies early on, and have more resources to pay for these new innovations [3]. As Crawford and Serhal (2020) [2] highlight, "digital health technologies interact with social, cultural, and economic realities and with social determinants of health to indirectly contribute to health inequity." Barriers to using eHealth technologies in underserved communities also include a lack of perceived value, limited digital and health literacy, and a lack of relevance [4,5].

Community-based partnerships are key to addressing these social determinants that serve as barriers to closing the digital divide and using technology to promote health equity [6]. In this viewpoint, when we refer to community, we are distinguishing between a "user" community defined as the client or consumer of a particular technology, platform, or service and

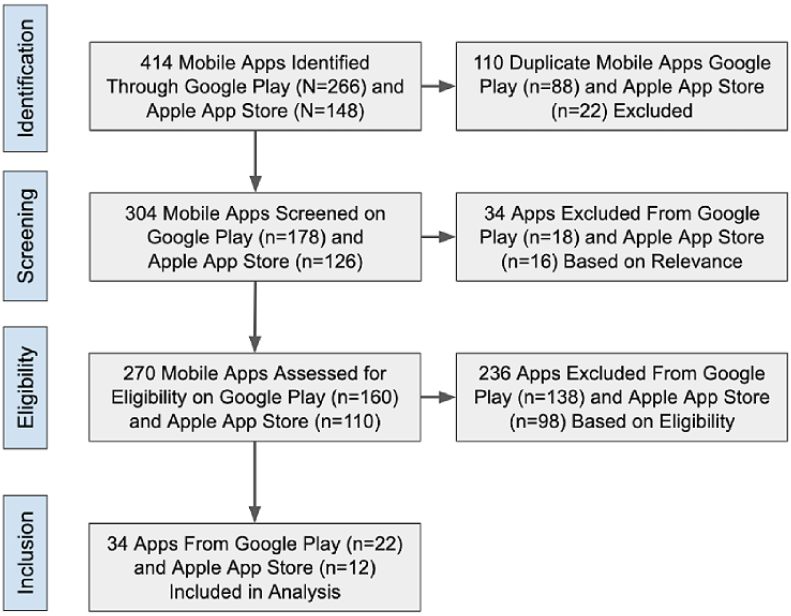
the communities (of color) as defined by the collective sector of the public that is disproportionately impacted by health disparities and that stands to benefit most by use of this technology. The latter is characterized by a shared sense of identity, understanding, or geographical distribution, but may or may not self-select into a user community [7]. As with the adoption of any new product or innovation, involvement of communities throughout the eHealth development process influences awareness of need in underserved communities as well as decisions around initial use, adoption, rejection, and continued use of eHealth innovations [8].

The goal of this viewpoint is to illustrate the importance of community involvement throughout the development of state-sponsored eHealth apps. We have chosen to focus on state-funded eHealth apps because these are funded by taxpayer dollars and should be responsive to public health needs. In what follows, we first present preliminary data on the proliferation of state-sponsored COVID-19–related contact tracing apps between February 26, 2020, and December 31, 2020. We then speak to our experience on the New York City (NYC) Health + Hospitals' community advisory board (CAB) during the rollout of the COVID Alert NY mobile app. Our experience shows how community involvement has practical implications on trust and buy-in from community-based organizations and, by extension, from communities disproportionately impacted by health inequities. We also provide some practical recommendations for developers on when and how to involve community representatives in their development process.

The Proliferation of State-Sponsored COVID-19 Contact Tracing Mobile Apps

We conducted a systematic search of state-sponsored COVID-19 apps on the Apple App Store and Google Play Store. We examined all COVID-19–related health apps released between February 26, 2020, and December 31, 2020. Following the methods outlined by Davalbhakta et al [9], we relied on the following keywords in the Apple App Store and Google Play Store: "Covid," "Corona," "Pandemic," "Covid-19," "SARSCOV2," "coronavirus," "2019-nCoV," "Covid-19 tracker," "Stop COVID," and "c-19." Two reviewers (TRH and SC) screened apps for relevance and eligibility for inclusion, including whether the apps were state-sponsored. Where there was ambiguity or disagreement around relevance and eligibility, review of the app was escalated to the lead investigator (MYI) for adjudication. Figure 1 provides a flowchart outlining our search and selection process.

Figure 1. Selection process of COVID-19 apps from Apple App Store and Google Play Store.



Following the 3 most common categories in schemas used in previous publications of COVID-19 apps [10-12], we categorized the apps in our search based on 3 distinct functionalities: (1) contact tracing/exposure notification, (2) symptom checking, and (3) information dissemination. Contact tracing and exposure notification functionality allows users to turn on exposure notifications and be alerted when users encounter (anonymous) others who tested positive in their location. Symptom checking allows users to enter symptoms along with some simple answers to questions and reveals options for next steps regarding the likelihood of COVID-19 infection. Information dissemination functionality in eHealth apps focuses on providing facts about COVID-19, good hygiene practices, and guidelines to follow, like social distancing and the importance of wearing face masks, how to access resources, and other types of relevant information.

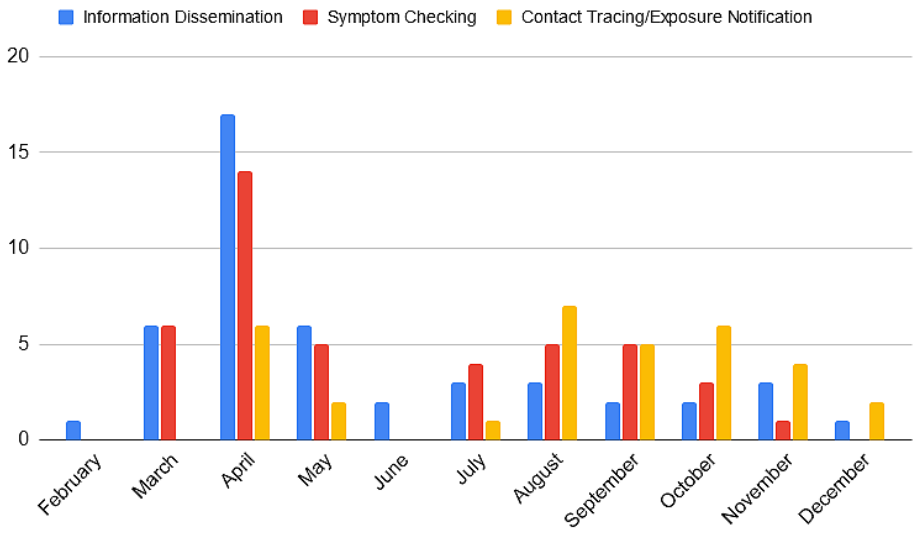
A total of 34 apps (12 Apple, 22 Google) met the inclusion criteria. Of them, 12% (n=4) only provided information to users (eg, suggested resources, provided updates, and delivered public service announcements through push notifications), 3% (n=1) only collected data from users (eg, symptom tracker that will determine whether a person may need further assessment or

testing for COVID-19), and 85% (n=29) of apps provided both information and collected data. In addition, 41% (n=14) of all apps reviewed included at least one feature for checking COVID-19–related symptoms, 77% (n=26) included functionality around contact tracing and exposure notification, and 50% (n=17) provided some information dissemination functionality. Table 1 shows the apps by category (eg, checking symptoms, contact tracing/exposure notification, and information dissemination) and Figure 2 provides a visualization of the number of COVID-19 apps released by category between February 2020 and December 2020.

Table 1. List of state-sponsored COVID-19 apps by category.

Name of app	Approximate date released	Operating system	Contact tracing and exposure notification	Symptom checking	Information dissemination
ABTrace Together	4/1/2020	iOS	✓		
AlohaSafe Alert	11/10/2020	iOS	✓		✓
ArriveCAN	4/21/2020	iOS		✓	
BC COVID-19 Support	5/21/2020	iOS		✓	✓
CA Notify	12/11/2020	Android	✓		
Canada COVID-19 (COVID Alert)	7/30/2020	iOS		✓	✓
Care19 Alert	8/10/2020	Both	✓		
Care19 Diary	4/19/2020	Android	✓	✓	✓
CO Exposure Notification	10/23/2020	Android	✓		
CombatCOVID MDC	8/20/2020	Both	✓		✓
COVID Alert	7/28/2020	Both	✓		
Covid Alert CT	10/30/2020	Android	✓		
Covid Alert DE	9/8/2020	Both	✓	✓	
Covid Alert NJ	9/30/2020	Both	✓	✓	✓
Covid Alert NY	9/30/2020	Both	✓	✓	
Covid Alert PA	9/11/2020	Both	✓	✓	✓
Covid Trace Nevada	8/17/2020	Both	✓	✓	✓
Covid View	12/29/2020	iOS			✓
Covid Watch Arizona	08/17/2020	Both	✓		
COVID-19 Virginia Resources	4/27/2020	Both			✓
Covid-19 Wisconsin Connect	5/14/2020	iOS			✓
Covidaware MN	11/16/2020	Both	✓		✓
CovidWise	8/7/2020	Both	✓		✓
Crush Covid RI	5/15/2020	Both	✓	✓	✓
DC Can	10/23/2020	Android	✓		
GuideSafe	8/11/2020	Both	✓	✓	
Healthy Together	4/1/2020	Both	✓	✓	✓
MD Covid Alert	11/5/2020	Android	✓		
MI Covid Alert	10/1/2020	Both	✓	✓	
NJ COVID 19	3/31/2020	iOS			✓
SlowCovidNC	9/16/2020	Both	✓		
Stronger than C19	4/25/2020	Both		✓	✓
WA Notify	11/25/2020	Android	✓		
WI Exposure Notification	12/20/2020	Android	✓		

Figure 2. Count of apps released over time by category.



A clear trend emerges where most apps released early in the pandemic were primarily focused on symptom checking and information dissemination. Starting in August 2020, however, we see fewer apps being released overall, but more activity around contact tracing apps. This trend corresponds with the release of many apps by local and national governments using Bluetooth-based exposure notification incorporating a system codeveloped by Google and Apple [13]. The technology allows users to turn on exposure notifications and be alerted when users encounter (anonymous) others who tested positive in their location. This collaboration between high-tech and public health organizations has been hailed as an exemplar of technology partnership for social good [14], yet little is publicly known about who is using these state-sponsored contact tracing apps.

COVID Alert NY Mobile App as a Case In Point

In NYC, there were 3 entities that played a significant role in the development and dissemination of the COVID Alert NY mobile app: the New York State Department of Health, NYC Department of Health and Mental Hygiene (DOHMH), and NYC Health + Hospitals, the largest municipal health care system in the United States, serving almost 500,000 uninsured NYC residents. Despite caution by community-based organizations and advocates early in the pandemic, there was not enough data to document health disparities related to COVID-19 morbidity and mortality [15]. It was clear on the ground that COVID-19 testing sites were not accessible to NYC communities of color [16]. Hesitancy around testing was further amplified by inadequate care where patients of color would, despite showing symptoms, be told that they were fine and be sent home on multiple occasions.

Once the data were collected and racial and ethnic disparities became apparent, an NYC Test and Trace CAB was organized to provide input on COVID-19–related efforts in NYC in May 2020. The CAB, which meets weekly with city health officials, was organized under the auspices of NYC Health + Hospitals and consisted of 71 members representing a broad range of organizations across all 5 boroughs [17]. The CAB was instrumental in directing where testing sites should be deployed and providing guidance on effective strategies for communicating critical information to community members while addressing language barriers and concerns around health literacy. The CAB also played a vital role in sharing lived experiences of community members to inform and reinforce community-based recommendations.

Although the DOHMH and NYC Health + Hospitals were responsive to the CAB's questions, concerns, and counsel around testing, there was no CAB input into the design, development, and dissemination of the COVID Alert NY mobile app for contact tracing and exposure notification, which was presented to the CAB in September 2020. Given the government's responsibility for public health, state and local authorities must be responsive to the best interest of their constituents and the public. Recognizing this responsibility, it is therefore essential to establish and implement community-driven processes that incorporate and prioritize the needs and concerns of disproportionately impacted and underserved communities. Our experience at the grassroots suggests that when there is proper community consultation, what follows is more engagement with, usage of, and penetration of government-led interventions.

As the epicenter of the pandemic in the United States in early 2020, NYC's response to the COVID-19 pandemic was positioned to model community-engaged practices that combat health disparities and promote health equity. Community-based organizations and advocates on the CAB expressed concerns around the digital divide and trust in state-sponsored apps within communities of color early in the pandemic. It is important to note here that the release of the COVID Alert NY mobile app was coming on the heel of Public Charge, in which public officials could deny applications for lawful immigration if they determined the applicant has used or will depend on public benefits [18]. The app was developed by the Department of Health and it was shared with the CAB to disseminate by

DOHMH and NYC Health + Hospitals. Despite expressing concerns around privacy and confidentiality, the CAB was given an app that was developed without community input and was being asked to disseminate it without having the ability to incorporate community feedback. As a result of this lack of community involvement, the CAB was reluctant to endorse or share the COVID Alert NY mobile app within our communities.

A Community-Based Participatory Research Approach to Health App Development in Underserved Communities

Developers of COVID-19 mobile apps must address impediments to eHealth tool utilization among underserved and disproportionately impacted communities, including access, privacy, and confidentiality, poor eHealth literacy, and language barriers [2,19]. There are numerous frameworks for incorporating community input into the research and development of programs to tackle COVID-19 health disparities. One relevant framework is participatory app design, which involves affected stakeholders from the inception of the project in designing solutions that identify and incorporate the community's unique needs into the app development [20]. Similarly, user-centered designs provide a framework to better understand who users are, as well as their goals, experiences, and expectations, to ensure users are kept at the center of the design process [21].

Our work focuses on using a community-based participatory research (CBPR) lens to involve community representatives in eHealth app development. CBPR is a collaborative approach that emphasizes long-term partnerships between communities and academics to ensure equity in each aspect of the research and development process [22,23]. For mobile apps to be able to address health disparities, development should include incorporating communities in the problem definition and design, technical and content development, deployment, evaluation, and dissemination of results to stakeholders.

Unlike participatory and user-centered design, a CBPR approach extends beyond incorporating a community perspective into technology design and aims to ensure equity throughout the entire process of design, development, and deployment [24-28]. These principles include understanding communities' resources and technical capacity, defining interactive processes that are responsive to community needs, equitable decision-making, and building collaboration in the design, development, deployment, and transparency around technology-related outputs, ownership, and maintenance [28]. We expand the principle around technology-related outputs to also include transparency around data collection and use. The goal of CBPR is not only to build something useful, but also to improve public health through an iterative and sustainable process where communities of color are kept front and center.

Practical Recommendations on When and How to Get Community Involved

Our experience serving on the NYC Test and Trace CAB illustrates how transparency and community involvement in the creation of COVID-19–related apps have practical implications for buy-in from community-based organizations and, by extension, from communities disproportionately impacted by COVID-19. Community-based organizations that have the public's trust due to years of work at the grassroots level, especially those representing and serving communities of color, rely on information about the extent to which the communities they serve were involved in the development of eHealth tools to assess whether to promote these solutions within the communities they serve. We recommend that developers work closely with community-based organizations who can serve as trusted public brokers and can help facilitate community involvement in all phases of app development. In the design phase of the development process, we recommend that app developers work alongside community-based organizations to (1) complete a need-based assessment before/while designing the app, (2) solicit regular community feedback on low- and high-fidelity prototypes, and (3) clearly identify and attribute where community feedback was incorporated. During the development phase, we recommend that developers (4) involve community members in the technical development/testing of the app, (5) involve community members in the development/review of content, and (6) conduct a focus group of community members and leaders to demo a prototype and discuss deployment of the app. Finally, for deployment, we recommend that developers (7) cocreate an evaluation plan (with a corresponding logic model) with a community-based organization partner before the app is deployed and (8) complete the evaluation with the involvement of the community-based organization partner and present results publicly (ie, through publications, presentations) once an app is deployed. We also recommend (9) incentivizing involvement by compensating community members for the opportunity cost of participating in needs-based assessments, focus groups, and program evaluation.

Discussion and Conclusions

Although other publications have reviewed COVID-19–related mobile apps, no reviews have considered community involvement in their design, development, and dissemination. Ming et al [29] provided an overview of features and functionality of 223 mobile health apps released in the early days of the pandemic on public app stores. Salehinejad et al [30] used the Mobile App Rating Scale to assess the acceptability of quality, content, and functionality of COVID-19–related apps found on the Google Play Store and Apple App Store. In addition, 2 other studies reviewed the literature to identify COVID-19 apps and reviewed apps with features ranging from information dissemination and risk/symptom assessment to contact tracing [10,31]. In line with the studies outlined above, we found that the majority of COVID-19 apps focused on symptom tracking, followed by

information dissemination, and then contact tracing and exposure notification. Almalki and Giannicchi [32] provided an overview and taxonomy of COVID-19 apps through September 2020 and found that the majority of apps they reviewed were contact tracing and exposure notification apps developed by governments or national authorities. Unlike previous studies, we focused on state-sponsored apps and provided the first comprehensive search of apps through December 2021.

In this viewpoint, we argue that COVID-19–related eHealth tools funded by taxpayer dollars should involve community-based organizations and advocates in the design, development, and dissemination given the government's responsibility for the entire public's health. Community-based organizations assess whether to promote eHealth tools to the communities they serve. In this way, our experiences suggest that transparency and community involvement in the process of app development increase buy-in, trust, and usage. As our personal experience illustrates, when affected communities are not included, this can lead to a lack of buy-in and trust, as well as a lack of community participation in the diffusion of eHealth innovations among constituents of those communities [21].

López et al [31] argue that to "harness its true potential and make the greatest difference, [health information technologies] need to be (1) designed with components that focus on the identification and eliminations of disparities...and (2) tailored to the needs of diverse populations." The authors point to user-centered design as a framework for incorporating communities of color into the design, development, and deployment of mobile health apps to achieve more equitable and better health outcomes. However, we encourage researchers and practitioners to use a CBPR lens to include communities of color, who are often disproportionately impacted by the COVID-19 pandemic, because this approach goes beyond involving individual "users." Rather, this CBPR approach includes the broader community, with the goal of improving public health and enhancing community capacity by supporting participation and establishing sustainable programs.

A lack of input from communities of color at every step of the eHealth tool development process contributes to bias. That is, small choices made throughout the design, development, and dissemination process have a large collective impact on the perceptions of adopters in underserved communities of the relative advantage, compatibility with values and experiences, and complexity of use, and the extent to which eHealth tools can be tested or provide tangible benefits [33]. When technology is not designed around the needs, expectations, values, and experiences of individual users, this can lead to a lack of adoption. Within the context of eHealth, this lack of inclusive design results in a lack of diffusion of these innovations within underserved communities, which ultimately exacerbates health disparities [34]. If we truly recognize that health disparities exist, we must ensure that underserved communities feel comfortable with eHealth apps to realize their full potential.

Acknowledgments

This work was supported by the Institute for Exposomic Research and pilot funding from the Transdisciplinary Center on Early Environmental Exposures (TCEEE) Pilot Project Program (grant P30ES023515). It was also funded by grants 5R25HL108857 and 5T37MD001452. Thank you also to Walter Ramos for expert administrative support.

Conflicts of Interest

None declared.

References

1. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
2. Crawford A, Serhal E. Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
3. Timmermans S, Kaufman R. Technologies and Health Inequities. Annu Rev Sociol 2020 Jul 30;46(1):583-602 [FREE Full text] [doi: 10.1146/annurev-soc-121919-054802]
4. Christopher Gibbons M. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011 Jan 01;8:1f [FREE Full text] [Medline: 21307989]
5. Polonsky WH, Fisher L. When does personalized feedback make a difference? A narrative review of recent findings and their implications for promoting better diabetes self-care. Curr Diab Rep 2015 Aug;15(8):50. [doi: 10.1007/s11892-015-0620-7] [Medline: 26077015]
6. Ginossar T, Nelson S. Reducing the Health and Digital Divides: A Model for Using Community-Based Participatory Research Approach to E-Health Interventions in Low-Income Hispanic Communities. Journal of Computer-Mediated Communication 2010;15(4):530-551. [doi: 10.1111/j.1083-6101.2009.01513.x]
7. Hiller ET. The Community as a Social Group. American Sociological Review 1941 Apr;6(2):189. [doi: 10.2307/2085549]
8. Robertson TS. The Process of Innovation and the Diffusion of Innovation. Journal of Marketing 1967 Jan;31(1):14. [doi: 10.2307/1249295]
9. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A Systematic Review of Smartphone Applications Available for Corona Virus Disease 2019 (COVID19) and the Assessment of their Quality Using the Mobile Application

Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
10. Kondylakis H, Katehakis D, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 Mobile Apps: A Systematic Review of the Literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
11. Lee B, Ibrahim SA, Zhang T. Mobile Apps Leveraged in the COVID-19 Pandemic in East and South-East Asia: Review and Content Analysis. JMIR mHealth uHealth 2021 Nov 11;9(11):e32093 [FREE Full text] [doi: 10.2196/32093] [Medline: 34748515]
12. John Leon Singh H, Couch D, Yap K. Mobile Health Apps That Help With COVID-19 Management: Scoping Review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
13. Privacy-Preserving Contact Tracing. Apple. URL: https://covid19.apple.com/contacttracing [accessed 2022-02-21]
14. Google and Apple Change Tactics on Contact Tracing Tech. WIRED. URL: https://www.wired.com/story/google-apple-change-tactics-contact-tracing-tech/ [accessed 2022-02-21]
15. The Pandemic's Missing Data. The New York Times. URL: https://www.nytimes.com/2020/04/07/opinion/coronavirus-blacks.html [accessed 2022-02-21]
16. Grigsby-Toussaint DS, Shin JC, Jones A. Disparities in the distribution of COVID-19 testing sites in black and Latino areas in new York City. Prev Med 2021 Jun;147:106463. [doi: 10.1016/j.ypmed.2021.106463] [Medline: 33647352]
17. NYC Health + Hospitals. URL: https://www.nychealthandhospitals.org/test-and-trace/community-partnerships/ [accessed 2021-08-24]
18. Perreira KM, Yoshikawa H, Oberlander J. A New Threat to Immigrants' Health - The Public-Charge Rule. N Engl J Med 2018 Sep 06;379(10):901-903. [doi: 10.1056/NEJMp1808020] [Medline: 30067441]
19. VonHoltz LAH, 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]
20. Spinuzzi C. The Methodology of Participatory Design. Society for Technical Communication 2005;52(2):163-174 [FREE Full text]
21. Abras C, Maloney-Krichmar D, Preece J. User-Centered Design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004:763.
22. Wallerstein N, Duran B, Oetzel JG, Minkler M, editors. Community-Based Participatory Research for Health: Advancing Social and Health Equity, 3rd Edition. Hoboken, NJ: John Wiley & Sons; 2017:3.
23. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. Health Promot Pract 2006 Jul 30;7(3):312-323. [doi: 10.1177/1524839906289376] [Medline: 16760238]
24. Israel B, Schulz A, Parker E, Becker A. Review of community-based research: assessing partnership approaches to improve public health. Annu Rev Public Health 1998;19:173-202. [doi: 10.1146/annurev.publhealth.19.1.173] [Medline: 9611617]
25. Doty J, Brady S, Monardez Popelka J, Rietveld L, Garcia-Huidobro D, Doty M, et al. Designing a Mobile App to Enhance Parenting Skills of Latinx Parents: A Community-Based Participatory Approach. JMIR Form Res 2020 Jan 24;4(1):e12618 [FREE Full text] [doi: 10.2196/12618] [Medline: 32012034]
26. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. Mhealth 2018 Aug;4:31 [FREE Full text] [doi: 10.21037/mhealth.2018.07.02] [Medline: 30221166]
27. Brewer L, Hayes S, Caron A, Derby D, Breutzman N, Wicks A, et al. Promoting cardiovascular health and wellness among African-Americans: Community participatory approach to design an innovative mobile-health intervention. PLoS One 2019;14(8):e0218724 [FREE Full text] [doi: 10.1371/journal.pone.0218724] [Medline: 31430294]
28. Unertl K, Schaefbauer C, Campbell T, Senteio C, Siek K, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]
29. Ming L, Untong N, Aliudin N, Osili N, Kifli N, Tan C, et al. Mobile Health Apps on COVID-19 Launched in the Early Days of the Pandemic: Content Analysis and Review. JMIR mHealth uHealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
30. Salehinejad S, Niakan Kalhori SR, Hajesmaeel Gohari S, Bahaadinbeigy K, Fatehi F. A review and content analysis of national apps for COVID-19 management using Mobile Application Rating Scale (MARS). Inform Health Soc Care 2021 Mar 02;46(1):42-55. [doi: 10.1080/17538157.2020.1837838] [Medline: 33164594]
31. John Leon Singh H, Couch D, Yap K. Mobile Health Apps That Help With COVID-19 Management: Scoping Review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
32. Almalki M, Giannicchi A. Health Apps for Combating COVID-19: Descriptive Review and Taxonomy. JMIR mHealth uHealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
33. Kaminski J. Diffusion of Innovation Theory. Canadian Journal of Nursing Informatics 2011;6(2):1444 [FREE Full text]
34. Groeneveld PW, Laufer SB, Garber AM. Technology diffusion, hospital variation, and racial disparities among elderly Medicare beneficiaries: 1989-2000. Med Care 2005 Apr;43(4):320-329. [doi: 10.1097/01.mlr.0000156849.15166.ec] [Medline: 15778635]

Abbreviations

CAB: community advisory board
CBPR: community-based participatory research
DOHMH: Department of Health and Mental Hygiene
NYC: New York City

Edited by L Buis; submitted 01.06.21; peer-reviewed by R Fullilove, G Darville, C Jonassaint, YM Huang, HM Tsai, S Sanchez; comments to author 17.07.21; revised version received 28.08.21; accepted 05.11.21; published 09.03.22.

Please cite as:
Idris MY, Korin M, Araya F, Chowdhury S, Medina P, Cruz L, Hawkins TR, Brown H, Claudio L
Including the Public in Public eHealth: The Need for Community Participation in the Development of State-Sponsored COVID-19–Related Mobile Apps
JMIR Mhealth Uhealth 2022;10(3):e30872
URL: https://mhealth.jmir.org/2022/3/e30872
doi:10.2196/30872
PMID:35113793

Review

mHealth Interventions for Self-management of Hypertension: Framework and Systematic Review on Engagement, Interactivity, and Tailoring

Weidan Cao[1], PhD; M Wesley Milks[2], MD; Xiaofu Liu[1], MSc; Megan E Gregory[1,3], PhD; Daniel Addison[2], MD; Ping Zhang[1], PhD; Lang Li[1], PhD

[1]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, United States

[2]Division of Cardiovascular Medicine, The Ohio State University College of Medicine, Columbus, OH, United States

[3]The Center for the Advancement of Team Science, Analytics, and Systems Thinking (CATALYST), The Ohio State University College of Medicine, Columbus, OH, United States

Corresponding Author:
Lang Li, PhD
Department of Biomedical Informatics
The Ohio State University
250 Lincoln Tower, 1800 Cannon Drive
Columbus, OH, 43210
United States
Phone: 1 614 292 4778
Email: lang.li@osumc.edu

Abstract

Background: Engagement is essential for the effectiveness of digital behavior change interventions. Existing systematic reviews examining hypertension self-management interventions via mobile apps have primarily focused on intervention efficacy and app usability. Engagement in the prevention or management of hypertension is largely unknown.

Objective: This systematic review explores the definition and role of engagement in hypertension-focused mobile health (mHealth) interventions, as well as how determinants of engagement (ie, tailoring and interactivity) have been implemented.

Methods: A systematic review of mobile app interventions for hypertension self-management targeting adults, published from 2013 to 2020, was conducted. A total of 21 studies were included in this systematic review.

Results: The engagement was defined or operationalized as a microlevel concept, operationalized as interaction with the interventions (ie, frequency of engagement, time or duration of engagement with the program, and intensity of engagement). For all 3 studies that tested the relationship, increased engagement was associated with better biomedical outcomes (eg, blood pressure change). Interactivity was limited in digital behavior change interventions, as only 7 studies provided 2-way communication between users and a health care professional, and 9 studies provided 1-way communication in possible critical conditions; that is, when abnormal blood pressure values were recorded, users or health care professionals were notified. The tailoring of interventions varied at different aspects, from the tailoring of intervention content (including goals, patient education, advice and feedback from health professionals, reminders, and motivational messages) to the tailoring of intervention dose and communication mode. Tailoring was carried out in a number of ways, considering patient characteristics such as goals, preferences, disease characteristics (eg, hypertension stage and medication list), disease self-management experience levels, medication adherence rate, and values and beliefs.

Conclusions: Available studies support the importance of engagement in intervention effectiveness as well as the essential roles of patient factors in tailoring, interactivity, and engagement. A patient-centered engagement framework for hypertension self-management using mHealth technology is proposed here, with the intent of facilitating intervention design and disease self-management using mHealth technology.

(JMIR Mhealth Uhealth 2022;10(3):e29415) doi:10.2196/29415

KEYWORDS

mHealth; mobile app; digital behavior change; interventions; systematic review; hypertension; engagement; interactivity; tailoring; mobile phone

Introduction

Background

Hypertension is an impactful risk factor for heart disease and stroke, both of which are leading causes of death in the United States [1]. Approximately 45% of American adults have a diagnosis of hypertension, and only 24% of those with hypertension have their condition under optimal control [1]. Effective treatment of hypertension requires patients to work with their health care providers and follow self-management guidelines, particularly relating to medication adherence.

Mobile health (mHealth) is defined as the use of mobile technologies (eg, smartphones) to provide medical and public health practice [2]. mHealth interventions used for disease self-management belong to digital behavior change interventions, defined as those involving digital technologies (eg, mHealth apps) to promote or support behavior change for improved health and self-management of chronic disease [3,4] for better health, which have been used to facilitate hypertension self-management. Potential benefits of mHealth interventions for disease self-management include (1) increasing medication adherence [5], (2) increasing knowledge, (3) empowering patients for self-care, (4) providing personalized self-care recommendations, and (5) facilitating patient–care provider communication and decision-making [6,7]. Hundreds of mHealth apps, often with features such as educational resources and monitoring reminders, have previously been developed to support hypertension self-management [8], and studies have demonstrated the effectiveness of using apps in blood pressure (BP) control and self-management behavior change such as medication adherence [6,9,10].

Existing systematic reviews examining the mHealth interventions for hypertension self-management or digital behavior change interventions using mHealth in the context of hypertension have focused on intervention efficacy and app usability [6,9,11]. However, studies examining engagement in the context of hypertension are limited. Engagement can be defined from a microlevel perspective (ie, engagement with digital behavior change interventions only, such as intervention use, or the subjective experience characterized by attention, affect, and interest) or from a macrolevel perspective (ie, engagement with the broader landscape of behavior change, such as medication adherence) [3,4]. Macrolevel engagement can be the result of the microlevel of engagement. For instance, engagement (eg, frequency, amount, and duration) with interventions can lead to behavior engagement or change (eg, medication adherence). Engagement is essential for the effectiveness of interventions involving digital technology [3,12,13], whereas lack of engagement with mHealth interventions would be expected to be associated with a lower rate of intervention success [14]. Subsequently, understanding the determinants of engagement can help design effective interventions. Theoretical frameworks [15,16] have proposed specific strategies to engage patients using mobile apps. Such strategies include but are not limited to providing educational information, reminding or alerting users, recording and tracking health information, providing guidance based on information entered by the user, enabling communication with clinicians, providing support through social networks, and supporting behavior change through rewards [15,16].

Although prior work has examined engagement in the context of chronic conditions [17,18], it has largely focused only on the microlevel of engagement [17,18], thus leaving the macrolevel predominantly unexamined. Intervention effectiveness measured by app use alone cannot be taken as a valid indicator of engagement because use metrics (microlevel engagement) do not indicate offline engagement indicators, and microlevel disengagement with the intervention or technology does not necessarily preclude macrolevel engagement (eg, users may take medications adherently but do not use the app to track medication taking behaviors) [4].

Therefore, it is of vital importance to examine both types of engagement as well as their determinants in mHealth behavior change interventions [19]. According to the motivational technology model [20], customization (ie, tailoring) and interactivity are the two key determinants of engagement. Tailoring refers to the extent to which users can customize the mHealth intervention to meet their needs [21]. For instance, an app may tailor the educational content delivered, messages, alerts, and reminders, and displays to users' specific needs and preferences; a patient on medication to manage severe hypertension likely requires different features and messaging than a patient who is managing mild hypertension through lifestyle modifications. Interactivity refers to the opportunities that the mHealth intervention affords for users to communicate with others, especially health care professionals [21]. For example, apps that have coaching from a trained professional tend to be more interactive. Along these lines, systematic reviews focusing on mHealth disease self-management interventions found that effective interventions integrated features of interactive communication [10] and tailored messages [22]. A systematic review of nutrition apps found that tailoring the apps to the needs of specific user groups can be beneficial in increasing engagement [23]. In addition, interventions can be more engaging if they are designed to be tailored to participants' health beliefs and needs. For instance, a systematic review of studies on health beliefs and medication adherence in patients with hypertension found that medication adherence was related to health beliefs that vary within and across countries, such as disease severity and susceptibility, medication necessity, or efficacy [24]. This implies that medication adherence interventions need to consider individual health beliefs about hypertension and BP medications [24].

Objective

Therefore, based on the importance of engagement, the research gap, the determinants of engagement, and the strategies to engage users, the following research questions were proposed:

1. What engagement strategies have been used in digital behavior change interventions for hypertension self-management?
2. How has engagement been defined or presented in the literature on digital behavior change interventions for hypertension self-management?
3. How has interactivity been implemented in digital behavior change interventions for hypertension self-management?
4. How has tailoring been implemented in digital behavior change interventions for hypertension self-management?

Methods

Overview

A systematic review of the literature was conducted to identify extant interventions and to investigate their key characteristics. Subsequently, content analyses of the studies were conducted for a deeper understanding of how engagement and its determinants were implemented in the interventions.

Search Strategy

The search focused on the identification of studies relating to mHealth interventions for hypertension self-management conducted worldwide. The search was conducted on the following databases: PubMed, PsycINFO, Embase, Communication and Mass Media Complete, CINAHL, MEDLINE, and MEDLINE Full Text. All articles indexed as of June 2020 were searched. A combination of search terms was used, including *hypertension*, *hypertensive*, *hypertensives*, or *blood pressure* (for the disease type); *self-management*, *self management*, *self-care*, *self care*, *management*, *coaching*, *control*, *monitor*, *adhere*, or *adherence* (for disease management); *mHealth*, *m-health*, *mobile*, *app*, *apps*, *application*, *applications*, *smart phone*, *smartphone*, *technology* (for mHealth); *intervention*, *trial*, *program*, *programme*, *experiment*, *pilot*, *study*, *effect*, *experience*, or *experiences* (for intervention). Please refer to Multimedia Appendix 1 for the search strategy and the corresponding justifications.

To ensure the comprehensiveness of the search, we also scanned relevant journals (eg, JMIR mHealth and uHealth) and the reference lists of review articles about these interventions.

Study Selection

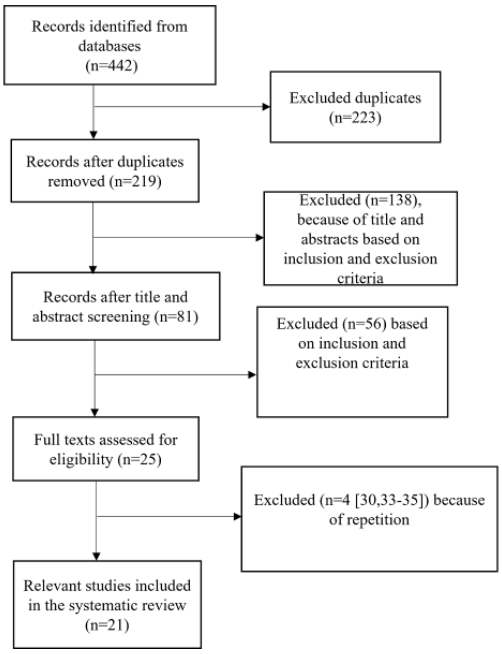
Included studies were those conducted among adults (aged ≥18 years), involving a mobile app to facilitate hypertension self-management, and with the aim of testing app or system experience. If an app was designed specifically for hypertension management for patients with hypertension, then the study was included (eg, the study by Kang and Park [25]). If an app was used for BP reduction or hypertension and another health condition (eg, weight management), the study was included (eg, the study by Mao et al [26]). If an app was designed for a specific purpose (eg, medication adherence) and could be applied to different health conditions or diseases, and was applied to patients with hypertension in the study, the study was included (eg, the study by Morawski et al [27]). If an app was used for >1 disease (eg, for both diabetes and hypertension) and if patients with hypertension or patients with both conditions were included as participants in the study, then the study was included [14]. If an app was used among patients with other diseases or conditions (eg, kidney transplant) and managing hypertension is crucial for that disease or condition, then the study was included (eg, the study by McGillicuddy et al [28]).

The exclusion criteria included articles that met one or more of the following characteristics: (1) use of apps for the purposes of disease screening or disease detection; (2) focus only on app design and development, without reporting any participants' app use experience; (3) primarily designed for healthcare providers or that reported professionals' user experience but did not focus on patients' user experience; (4) study of children; (5) based solely on non–smart phones, on the internet, or on text messages; (6) not written in English, and (7) contained only an abstract, without full publication. Covidence [29] was used to manage the review process.

A total of 442 records were imported to Covidence, and 223 duplicates were removed. After title and abstract screening, among the remaining 219 articles, 138 (63%) were removed based on not meeting the inclusion criteria. The remaining 37% (81/219) articles were screened, and 69% (56/81) of them not meeting the inclusion criteria were removed at this stage. A further review of the remaining 31% (25/81) of the articles indicated that 16% (4/25) of the articles [30-33] were based on the same app or intervention system. The study by Bengtsson et al [30] was not included in this review because it focused more on system development. The other study [33] was not included in this review because the patient participants were a subgroup of the patient participants in another study [32]. Furthermore, the studies by McGillicuddy et al [28,34] were based on the same system, and one of these studies [34] was not included because the other study [28] built on the 3-month randomized control trial conducted by McGillicuddy et al [34] and was a follow-up of that study. In addition, studies by Persell et al [35,36] were based on the same mobile app, so one of the studies [35] was removed because it focused more on the design of the app. Furthermore, the studies by Moore et al [37] and Thies et al [14] were based on the same app, and the studies by Chandler et al [38] and Davidson et al [39] were based on the same smartphone medication adherence stops hypertension program. For those studies that used the same system or app, if the participants, goals or outcomes, or methods (eg, surveys or interviews) were different, they were included in the review. For instance, although Moore et al [37] and Thies et al [14] used the same system, Moore et al [37] provided positive evidence that the system was effective in hypertension management, whereas Thies et al [14] analyzed the reasons why their intervention failed by using participant interviews. Including both articles would allow a full understanding of the effectiveness of the app or system. Therefore, the final sample size of the review is 21. A total of 2 authors (ie, WC and XL) worked independently during the screening and selection process first and then compared their results. Discrepancies were resolved through one round of discussion. Figure 1 presents the

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

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



Quality Assessment and Data Analysis

For quality assessment, 2 coders (WC and XL) independently evaluated the quality of the included studies using four sets of risk of bias evaluation tools: Cochrane Collaboration's Risk of Bias Tool for randomized control trials [40], 2 tools from the National Institutes of Health for observational studies and pre-post studies without a control group [41], and the critical appraisal skills program for qualitative studies [42]. For the study [43] that focused on both qualitative and quantitative data, we used two sets of criteria (ie, pre-post study with no control and qualitative study) to evaluate the risk of bias. These tools were chosen because they have been applied to previously published systematic reviews [9]. Disagreements between the coders were resolved after rounds of discussion and consultation with a third researcher (PZ). The results of the evaluation are in Multimedia Appendix 2 [27,36-39,44-47], Multimedia Appendix 3 [26,28,48,49], Multimedia Appendix 4 [25,31,43,50-52], and Multimedia Appendix 5 [14,32,43].

For the systematic review, 2 coders (WC and XL) independently coded the 21 studies. After independent coding, discrepancies were identified and resolved through multiple rounds of discussion and recoding. Through this iterative process, full agreement was reached for all variables of the systematic review. For coding, the investigators relied on the reporting in the article and referred to related articles listed in the references when applicable (eg, when coding an app that was published in multiple papers). If studies included both patients' and providers' perspectives, only the patients' perspectives were coded. In the event that the article presented no relevant information or the description was general or vague, we coded it as *unknown*.

Results

Study Characteristics

There were 21 studies included in the final analysis, with publication years ranging from 2013 to 2020. Most studies (14/21, 67%) were conducted in the United States. Moreover, of the 21 studies, 2 (10%) were conducted in China, and 2 (10%) other studies were conducted in Sweden. The remaining 14% (3/21) of the studies were conducted in Canada, South Korea, and Spain. The sample size ranged from 17 to 5115 participants, with mean age of 42.44 to 60 years.

The interventions were either developed for general audiences (eg, those who were overweight (BMI >25 kg/m^2; [26]), for specific audiences with hypertension (eg, patients with poorly controlled hypertension [27], or for patients with diabetes, hypertension, or both [14]. Of the 21 studies, 9 (43%) were randomized control studies [27,36-39,44-47], 4 (19%) were observational studies [26,28,48,49], 6 (29%) were pre-post studies without a control group [25,31,43,50-52], and 2 (10%) were qualitative studies [14,32]. Engagement or self-management behaviors (eg, medication adherence) were not the focus or outcome of longitudinal studies (eg, the study by McGillicuddy et al [28]). In terms of the intervention content, most studies [14,25,27,28,31,32,36-39,43-45,47,48,51] involved medication tracking or medication adherence.

In addition, 24% (5/21) of the studies used a theoretical framework in their interventions. Specifically, some studies [28,38,39] used the self-determination theory. Other theories, including the health belief model and technology acceptance model [43], and the technology-supported apprenticeship model [37] were also applied. No other studies reported any theoretical models. Table 1 presents a summary of the intervention characteristics.

Table 1. Intervention characteristics.

Study	Country	Sample size	Participants' demographic and hypertension characteristics	Duration	Outcomes	Theory
Bengtsson et al [31]	Sweden	50	Mean age 59.5 years; being currently treated for hypertension; mean SBP[a] 142, mean DBP[b] 84	56 days or 8 weeks	SBP and DBP, identification of subsets or classes of patients who differed from each other with respect to level of BP[c] at baseline	No
Chandler et al [38]	United States	54 (IG[d]=26; enhanced standard care=28)	I mean age 44.4 years; enhanced standard group mean age 46.8 years; Hispanic or Latino participants; diagnosed with and prescribed medication(s) for essential hypertension; uncontrolled hypertension	9 months	PO[e]: change in resting SBP from baseline to the 6-mo time point; SO[f]: resting DBP and MA[g]	Self-determination theory
Ciemins et al [49]	United States	IG=131; CG[h]=353	Mean age 60 years; patients with newly diagnosed or persistently uncontrolled BP[h] (ie≥140/90 mm Hg)	32 weeks	Patients' compliance with study protocol of taking 3 BPs per week	No
Davidson et al [39]	United States	38 (IG=18; CG=20)	I mean age 47.5 years; 47% (18/38) African Americans and 53% (20/38) Hispanics; uncontrolled hypertension	6 months	Changes in clinic SBP and changes in clinic DBP; changes in SBP control; changes in DBP control; recruitment and retention rates; MA; BP adherence	Self-determination theory
Duan et al [43]	China	143	Aged >18 years, hypertension diagnosis with no other serious complications	2 months	Patient compliance with hypertension self-management	Health belief model and the technology acceptance model
Gong et al [44]	China	480 (IG=225; CG=218)	I mean age 58.2 years; C mean age 59.27 years; patients aged 18 to 79 years diagnosed with primary hypertension	6 months	PO: SBP and DBP changes in patients; change in percentage of participants in the 2 groups with controlled BP. SO: MA	No
Hallberg et al [32]	Sweden	49	Female median age 58 years; male median age 62.5 years; female years with hypertension median 8; male years with hypertension median 6.6	8 weeks	Understanding of the interplay between BP and daily life; motivation to follow treatment	No
Kang and Park [25]	South Korea	38	Mean age 56 years; patients with hypertension who take antihypertensive medications (taking 1 or more antihypertensive drugs)	4 weeks	MA, perceived usefulness, user satisfaction	No
Kaplan et al [48]	United States	5115	Mean age 49 years; mean SBP130 mm Hg; participants who recorded ≥2 BP measurements were included in the study	22 weeks	Use pattern (engagement), efficacy of the app in BP reduction	No
Mao et al [26]	United States	IG=763; CG=73	I mean age 44.78 years; overweight (defined as BMI >25 kg/m^2; 14.3% (109/763) participants self-reported hypertension	First 4 months of intensive active coaching and 8 months of maintenance coaching	PO: weight loss at 4 months as defined by percent change in total body weight. SO: change in SBP after 4 months of intensive health coaching, as well as the change in number of participants in each hypertensive category from the beginning of enrollment to after 4 months of coaching	No
Márquez Contreras et al [45]	Spain	148 (IG=73; CG=75)	Mean age 57.5 years; patients with mild to moderate arterial hypertension	18 months (with an inclusion period of 6 months and a follow-up of 12 months)	Pharmacological adherence and control of BP in patients with mild to moderate arterial hypertension	No

Study	Country	Sample size	Participants' demographic and hypertension characteristics	Duration	Outcomes	Theory
McGillicuddy et al [28]	United States	IG=8; CG=9	I mean age 42.44 years, C mean age 57.89 years; renal transplant recipients with hypertension with documented medication nonadherence	12 months after the completion of a 3-month randomized control trial	SBP	Self-determination theory
Moore et al [37]	United States	42 (IG=20; CG=22)	Mean age 50.0 years; patients with essential hypertension (average BP≥140/90 and ≤180/120) who were taking 0 or 1 medications	12 weeks	PO: absolute decrease in SBP and DBP and the number of participants who reached the BP goal of ≤130/80 mm Hg. SO: the number of participants who reached the BP goal of ≤140/90 mm Hg, the number of participants who achieved >10 mm Hg decreases in SBP and >5 mm Hg decreases in DBP, the change in medication load, the absolute decrease in weight, the number of patients who lost at least 2.3 kg, hypertension knowledge, satisfaction in care, and the amount of clinician time required in the care	A technology-supported apprenticeship
Morawski et al [27]	United States	IG=209; CG=202	I mean age 51.7 years; C age mean=52.4 years; patients with poorly controlled hypertension	12 weeks	PO: change in self-reported MA and SBP. SO: whether participants had well-controlled BP, defined as 140/90 mm Hg or less	No
Ovbiagele et al [47]	United States	24 (IG=8; CG=16)	Patients with hypertension after stroke	3 months	SBP; emergency department use reduction	No
Patel et al [51]	United States	48	Mean age 53 years; African American 96% (46/48); established essential hypertension; prescribed at least two antihypertensive medications	12-week activation (intervention) phase	PO: MA; SO: MA, level of BP control by clinic measures, pill phone use, patient satisfaction, hypertension medication number and changes during the study period, office visits, emergency room visits, and hospitalization	No
Persell et al [36]	United States	IG=144; CG=153	I mean age 59.6 years; C mean age 58.3 years; adults with uncontrolled hypertension (defined as at least 145 mm Hg systolic or 95 mm Hg diastolic)	6 months	PO: SBP at 6 months. SO: self-reported antihypertensive MA, home monitoring and self-management practices, measures of self-efficacy associated with BP, weight, and health behaviors	No
Petrella et al [46]	Canada	IG=67; CG=60	I mean age 56.7 years; C mean age 59.1 years; participants with at least two metabolic syndrome risk factors	52 weeks including 12 weeks of intervention	PO: SBP and other cardiometabolic risk factors. SO: DBP, waist circumference, lipids (with the exception of high-density lipoprotein cholesterol, which was expected to increase) and markers for blood glucose and inflammation	No
Thies et al [14]	United States	15 out of 22 downloaded the app	Mean age 50 years (22 participants); 27% (6/22) of the patients with diabetes, 18% (4/22) with hypertension, and 55% (12/22) with both	Trial suspended, owing to low enrollment and inconsistent use of the app	The original aim of this study was to evaluate the effectiveness of a commercial mHealth app in improving clinical outcomes for adult patients with uncontrolled diabetes or hypertension, or both. Because of low enrollment and low app use, the project aim was changed to understanding why the trial was unsuccessful	No
Toro-Ramos et al [50]	United States	50	Starters mean age 40.40 years; completers mean age 47.68 years; adults with prehypertension or hypertension	24 weeks	Weight change, BMI change, DBP change, SBP change, hypertension category change	No

Study	Country	Sample size	Participants' demographic and hypertension characteristics	Duration	Outcomes	Theory
Weerahandi et al [52]	United States	17	Mean age 59 years; adults currently taking hypertension medication and had a diagnosis of prehypertension or stage 1 hypertension	13 weeks or 120 days	Engagement and acceptability: the number of blood pressure measurements, weight measurements, and daily steps were logged; the number of coaching phone calls attempted and completed, servings documented in the dietary assessment, and goals set were also assessed. Physiological parameters: BP, heart rate, weight, and steps changes	No

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]BP: blood pressure.

[d]IG: intervention group.

[e]PO: primary outcome.

[f]SO: secondary outcome.

[g]MA: medication adherence.

[h]CG: control group.

Intervention Strategies

Overview

All of the studies used at least two strategies to engage patients. The number of strategies used in the interventions varied from 2 to 6, with a possible maximum of 8. Table 2 provides further details.

Table 2. Engagement strategies used in the interventions.

Study	Providing health-related educational information	Reminding or alerting users	Motivational messages or encouragement	Recording and tracking health information	Providing guidance based on information entered by the user	Enabling 2-way communication with clinicians	Providing support through social networks	Supporting behavior change through rewards
Bengtsson et al [31]	No	Yes	Yes	Yes	No	No	No	No
Chandler et al [38]	No	Yes	Yes	Yes	Yes	No	No	No
Ciemins et al [49]	Yes	Unknown	No	Yes	Yes	No	No	No
Davidson et al [39]	No	Yes	Yes	No	No	No	No	No
Duan et al [43]	Yes	Yes	No	Yes	Yes	No	Yes (leaderboard module, version 4)	No
Gong et al [44]	No	Yes	No	Yes	Yes	Yes	No	No
Hallberg et al [32]	No	Yes	Yes	Yes	No	No	No	No
Kang and Park [25]	Yes	Yes	No	Yes	Yes	No	No	No
Kaplan et al [48]	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Mao et al [26]	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Márquez Contreras [45]	No	Yes	No	Yes	No	No	No	No
McGillicuddy et al [28]	No	Yes	No	Yes	Yes	No	No	No
Moore et al [37]	No	No	No	Yes	No	Yes	No	No
Morawski et al [27]	No	Yes	No	Yes	No	No	Yes	No
Ovbiagele et al [47]	No	Yes	Yes	Yes	Yes	No	No	No
Patel et al [51]	Yes	Yes	No	Yes	No	No	No	No
Persell et al [36]	Yes	Yes	Yes	Yes	Yes	Artificial intelligence coaching	No	No
Petrella et al [46]	No	No	No	Yes	Yes	No	No	No
Thies et al [14]	No	No	No	Yes	Yes	Yes	No	No
Toro-Ramos et al [50]	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Weerahandi et al [52]	Yes	No	No	Yes	Yes	Yes	No	No

Providing Health-Related Educational Information

A total of 9 studies [25,26,36,43,48-52] provided health-related educational information. The content of education varied, with some studies [36,43,48,49] focusing on hypertension, some studies [25,51] focusing on hypertensive medications, and some studies (eg, the studies by Mao et al [26], Toro-Ramos et al [50], and Weerahandi et al [52]) including dietary approaches to reducing hypertension. However, of the 9 studies, only 1 (11%) [36] specified in the educational materials the reason why self-monitoring is important in BP management and how their control through healthy behavior change is important for lowering the risk of complications.

Recording and Tracking Health Information

Some studies [14,25,27,28,31,32,36-38,43,44,47] included features to record both BP and medication adherence or intake. Some other studies included features to record either BP [45,46,48-50,52] or medication intake or adherence [51]. Some studies also recorded other information such as medication side effects (eg, the studies by Bengtsson et al [31] and Hallberg et al [32]), symptom logging (eg, the studies by Bengtsson et al [31], Hallberg et al [32], and Duan et al [43]), and the tracking of diet, heart rate, weight, and steps (eg, the study by Weerahandi et al [52]).

Reminding or Alerting Users

Most studies [25,27,28,36,38,39,43-45,47,48,50,51] included reminders for medication intake or BP monitoring, or both. Various reminders focusing on other topics or other types were exercise [44]; weight, diet, exercise, and discomfort [43]; weight, meals and snacks, and physical activity [36]; hospital visit date and input of lifestyle data [25]; clients' personal goals [26]; appointments [45]; or alerting a *Medfriend* who provides peer support when doses are missed [27]. However, some studies [31,32] included reminders but did not specify the content of reminders.

Motivational Messages or Encouragement

There are studies that included motivational messages or encouragement [26,31,32,36,38,39,47,48,50]. For instance, the motivational messages of 1 intervention [38] were designed based on participants' previous medication adherence levels (ie, nonadherence, partial adherence, and complete adherence) and on their values, beliefs, and long-term or short-term life goals.

Providing Support Through Social Networks

A total of 2 studies included the feature to provide support through social networks. In version 4 of 1 app [43], a leaderboard module presenting and comparing the scores generated based on each patients' self-management behaviors was provided for those users who wanted to enhance their self-management motivation. In another study [27], users were able to designate a *Medfriend*, who was someone else who was granted access to the patient's medication taking history, received alerts when the patient missed doses, and was able to provide peer support.

Supporting Behavior Change Through Rewards

Only 1 intervention [48] included gamification features with a reward system to maximize user interaction. In this app, enthusiastic amination appeared on the screen after each BP recording event [48].

Providing Guidance Based on Information Entered by the Users

If cutoffs of BP were exceeded or out-of-range values were observed, patients were contacted [47] or were recommended to take additional BP measurements [28,36,44] or to seek medical attention [44,50]. Health care providers were notified [49] or were called when extreme values were recorded [36] or contacted to follow-up with participants [43,46] and asked to determine the course of action to take with the participant [38] or to make an adjustment to medical regimen as warranted [28,47]. In addition to guidance on out-of-range values, the interventions also helped with solving problems [52]; providing personalized or tailored recommendations or advice [26,36,44]; providing encouragement and suggestions and answering nonpressing questions [14]; providing personalized explanations regarding the stages of hypertension and translation into cardiovascular risk [48]; providing strategies to address behavior change related to calorie reduction, diet improvement, nutrient intake, physical activity increase, and sodium intake reduction [50]; or providing tailored recommendations to users' questions on lifestyle management (ie, sodium intake, body weight, waist circumference, exercise, alcohol, smoking, and stress) [25].

Interactivity

Interactivity was analyzed based on providing guidance related to information entered by the users (eg, when abnormal BP values were recorded) and based on whether or not the intervention enabled 2-way regular communication (outside of just specific situations such as when abnormal BP values are observed) between users and a health care professional. A total of 9 studies [28,36,38,43,44,46,47,49,50] included 1-way communication under possible critical conditions. Patients or their health care providers were notified or contacted when out-of-range BP values were reported. Interactivity was limited in the interventions; only 33% (7/21) of the studies provided the possibility of interaction with health care providers or health coaches. In terms of 2-way communication, studies included the possibility of communicating with physicians [14,44], health coaches [26,37,50,52], or an artificial intelligence coach [36]. Users were able to have remote consultations with professional doctors [44] or members of their care team [14]. For instance, a trained coach [50], professionals (licensed nutritionists, physical therapists, and social workers) [26], or master clinicians [37] provided human coaching. For a summary of interactivity, please see Table 3.

Table 3. Interactivity, tailoring, and engagement.

Study	Interactivity	Tailoring	Microlevel engagement	Macrolevel engagement
Bengtsson et al [31]	No	Yes	No	No
Chandler et al [38]	Yes	Yes	No	No
Ciemins et al [49]	Yes	No	No	No
Davidson et al [39]	No	Yes	No	No
Duan et al [43]	Yes	Yes	No	No
Gong et al [44]	Yes	Yes	No	No
Hallberg et al [32]	No	Yes	No	No
Kang and Park [25]	No	Yes	No	No
Kaplan et al [48]	No	Yes	Yes	No
Mao et al [26]	Yes	Yes	Yes	No
Márquez Contreras [45]	No	No	No	No
McGillicuddy et al [28]	Yes	Yes	No	No
Moore et al [37]	Yes	Yes	No	No
Morawski et al [27]	No	Yes	No	No
Ovbiagele et al [47]	Yes	Yes	No	No
Patel et al [51]	No	Yes	No	No
Persell et al [36]	Yes	Yes	No	No
Petrella et al [46]	Yes	Yes	No	No
Thies et al [14]	Yes	No	No	No
Toro-Ramos et al [50]	Yes	Yes	Yes	No
Weerahandi et al [52]	Yes	Yes	Yes	No

Tailoring, Customization, or Personalization

Some level of tailoring was achieved in many studies. All 21 studies included only 1 app version, except 1 (5%) study [43]. Moreover, 4 versions of the app were developed based on users' disease cognition, self-management experience, and self-management motivation, wherein version 1 had three functional modules (ie, management plan, reminder service, and health checkup), version 2 had four modules (health education was added), version 3 had five modules (health education and health report were added), and version 4 had all six modules (health education, health report, and health report were added) [43].

Some studies [26,46,52] included personalized health goals, such as individualized exercise prescription [46]. The content, information, or features of some interventions was or were customized, based on goals or individual preferences [27,43,50,51], based on antihypertensive medication prescription [31], and based on users' values and inputs [36].

Some studies [43,44,48] provided personalized feedback or advice in the intervention. For instance, in 1 study, physician's advice was based on patient's hypertension self-management experience level [43], and in another study, a personalized explanation of the relationship between stages of hypertension and cardiovascular risk was provided [48]. Interventions with tailored target management recommendations included the studies by Kang and Park [25] and Persell et al [36]. In the study by Moore et al [37], users could make shared decisions about diet, exercise, stress management, and medication with the coach. In some studies, the motivational messages were tailored based on users' personal preferences [31] and on users' medication adherence rates, goals, and values and beliefs [38,39]. A total of 2 studies [28,32] included tailored reminders. In some studies [28,47,52], the communication model or channel was customized so that patients were contacted via the preferred mode: SMS text messaging, email, or phone. In 1 study [26], the intervention dose (eg, coaching frequencies) was based on the participants' needs and availability. For a summary of tailoring, please see Table 3.

Engagement

In addition to describing the intervention strategies used to engage users, we also explored how engagement has been defined, reflected, and related to biomedical outcomes.

How Engagement Has Been Defined

Engagement was defined or operationalized as microlevel interactions with the interventions, but from different dimensions, such as frequency of engagement, time or duration of engagement with the app, and intensity of engagement. A total of 4 studies [26,48,50,52] clearly defined or operationalized engagement (Table 3). Moreover, 1 study [48] defined low engagement (ie, "recording BP for less than 4 weeks"), medium engagement (ie, "recording BP for 4-8 weeks"), and high

engagement (ie, "recording BP for longer than 8 weeks"). Another study [26] also defined low (ie, "at the bottom quartile of number of messages and video consults"), medium (ie, "participants in the 25th-75th engagement percentiles"), and high engagement (ie, "top quartile of messages sent per month or number of coaching consults in the 4-month coaching period"). Toro-Ramos et al [50] defined different levels of engagement as starters (ie, "participants who completed at least one lesson per week during the first month, as well as engaged with the health coach (at least once through in app one-on-one messages or through phone calls)"), and completers (ie, "participants who completed at least nine core lessons of 22"). Weerahandi et al [52] defined engagement as "Messages sent to the coach per person, messages sent from the coach per person," "number of times blood pressure was logged," "number of times weight was logged," "number of times steps were logged," "logged food entries," and "goals recorded."

None of the studies we reviewed directly examined or measured users' subjective experience of engagement, focusing on attention, interest, and affect. Some studies explored users' subjective experience with a focus on user satisfaction or usability in general using interviews or surveys [38,49], usually conducted at the end of the intervention, which may not objectively capture attention or affect during the intervention, given the broad focus and retrospective nature [53].

Behaviors Reflecting Engagement

Although some studies did not clearly state in the articles that they measured engagement with digital behavior change interventions (microlevel) or engagement with behavior change (macrolevel), those behavior-related outcome variables, to some extent, reflected users' macro- or microlevel of engagement, or both. Two commonly measured behaviors in the outcome variables of the studies were medication adherence and BP self-monitoring. Studies measuring medication adherence used different methods: using technology or devices [38,39,45,51], using self-report or surveys [25,27,36,38,44,51], or using a pharmacy refill rate [51]. Studies [36,38,39,43,46,48,49,52] also used the app or a Bluetooth BP device to measure BP self-monitoring behavior.

In addition to medication adherence and BP self-monitoring, other behaviors were also measured: food or meals logged [36,50,52], messages sent or conversations with the app [14,26,36,52], steps taken [46,52], body weight logging [46,52], frequency of users accessing their weekly BP report [48], number of coaching consults [26], and lessons completed [50].

The Relationship Between Engagement and Biomedical Outcomes

Of the 4 studies that clearly defined engagement, 3 (75%) studies [26,48,50] tested and demonstrated the statistical relationship between engagement and biomedical outcomes (ie, weight or BP change), indicating that higher engagement was associated with significantly better biomedical outcomes. However, in 1 study [52], the relationship between levels of engagement and biomedical outcomes was not tested. Among studies that did not explicitly define engagement but included behaviors reflecting engagement, none statistically tested the relationship between the behaviors and biomedical outcomes. However, using patient interviews, 1 study [32] explained the mechanism between engagement and the motivation for macrolevel behavior change: as patients became engaged in graphs or through answering questions and measuring their BP, they were motivated to follow their treatment and understood the interplay between lifestyle and BP. Persell et al [36] did not test engagement but tested the factor crucial to macrolevel engagement or behavior change, self-efficacy [54], or engagement self-efficacy [55], that is, the self-confidence in using the app, controlling BP, knowing when medication changes were needed, and performing nonpharmacologic behaviors to control BP. The study by Persell et al [36] found that self-efficacy in controlling BP was greater in the intervention group.

Discussion

Principal Findings

Overview

To the authors' knowledge, this is the first review examining the interactivity, customization, and engagement factors of mHealth interventions for hypertension self-management. This review included 21 studies.

Participants

On the basis of the results of the participant inclusion criteria, some studies had very specific criteria, whereas others had very broad inclusion criteria. As participant characteristics (eg, disease type and severity of disease) were quite diverse in some studies, more research is needed to explore the goals, needs, and characteristics of users.

Design of Interventions

Engagement or self-management behaviors were not included in the outcomes of the limited longitudinal studies. No studies tested the engagement or self-management behaviors after the mHealth technology was no longer provided. Lack of longitudinal design leads to inability to elucidate behavior change or engagement patterns over time. Without testing macrolevel engagement or self-management behaviors (eg, medication adherence) when interventions or apps are no longer available, it cannot be confirmed that digital behavior change interventions are effective in changing behaviors in the long run.

Theoretical Frameworks Applied

Self-determination theory, health belief model, technology acceptance model, and technology-supported apprenticeship models have been applied to a limited number of studies. Some of these theories (eg, self-determination theory) have also been applied to diabetes self-management interventions using mobile apps [56] and applied to mHealth interventions in improving medication adherence among people with hypertension [11]. More interventions should adopt a theoretical framework (eg, behavior change theories) to guide work in this area.

Optimal Combination of Engagement Strategies

Many studies have used a combination of features that are likely to engage users. What specific combination of features works best for engagement is yet to be determined, but should consider patients' characteristics. For instance, patient motivation for self-management may be a factor to consider. Providing patient education content in an app can be a way of engaging patients, especially those who are motivated. However, for those who are not motivated (eg, "motivating participants to read the educational materials remained a challenge," as indicated by Weerahandi et al [52]), a patient education section may be less beneficial, and other strategies should be considered to motivate patients. Similarly, reminders and motivational messages could be more effective when refined according to patient characteristics and interactions with the app [32,51].

Interactivity

Interactivity was limited in digital behavior change interventions, as only 7 interventions provided 2-way communication between users and a health care professional or a coach. Moreover, 9 interventions provided 1-way communication in possible critical conditions; that is, when abnormal BP values were recorded, users or health care professionals were notified. On the basis of the results, the levels of interactivity between users and health care professionals can be characterized into four major categories: no interaction, limited interaction, regular interaction, or focused interaction. Limited interaction includes providing support under possible critical conditions. Regular interaction includes providing the possibility of 2-way communication between users and health care providers (eg, questions and answers and receiving regular feedback and recommendations), along with other strategies or app features. Focused interaction includes providing patient coaching by a clinician or a trained coach, which is the dominant feature, goal of the app, or intervention.

In our review, we found that some apps contained interactivity functions, whereas others did not. Authors did not describe the decision to exclude interactive features; we posit that the decision of including or excluding interactivity might be based on a variety of factors, including patient factors (eg, needs), intervention goals, and health care providers' availability. For instance, if an app's aim is medication adherence, interactivity is not a very important feature, whereas if an app's aim is logging symptoms, especially alarming symptoms, then interactivity (eg, health care providers' feedback) would be a critical function. In addition, provider-related factors are also worth considering. Studies have demonstrated health care providers' barriers of using apps to communicate with patients, including time constraints, increased workload, lack of interest, and lack of investment in app development [57].

Some level of potential interactivity should be included in the interventions using mobile apps for hypertension self-management. As 1 article examining patient perspectives indicated, ambiguity and anxiety could be provoked with BP readings, especially when readings are high [7]. Interaction with health care professionals, at least during possible critical points perceived by patients, can be an essential feature to provide professional guidance and ease the concerns and promote engagement with the interventions and the behavior change process [58].

Tailoring

The tailoring of interventions varied at different aspects, from tailoring of intervention content (including goals, patient education, advice and feedback from health professionals, reminders, and motivational messages) to tailoring intervention dose and communication mode. Tailoring was carried out in several ways, including consideration of patient characteristics such as goals, preferences, disease characteristics (eg, hypertension stage and medication list), disease self-management experience level, medication adherence rate, values, and beliefs. Although multiple studies included reminders for medication administration, only two of them provided tailored reminders. Medication nonadherence can be due to many factors: medication side effects, cost, forgetfulness, or perceived lack of need to take medications. In addition, personal and cultural values and beliefs (eg, perception of illness, illness knowledge, health literacy, cultural beliefs, self-efficacy, and spiritual and religious beliefs [59]) can impact medication adherence. For instance, higher perceived benefits of herbs and lower perceived benefits of Western medications were predictors of antihypertensive medication nonadherence among Chinese immigrants [60]. A medication reminder feature may therefore work well for those who forget to take medications but may be ineffective for those who do not take medications as prescribed because of costs or potential side effects [61] or those who do not believe in the benefits of Western medications. For those who have high medication adherence rates, reminder features may be redundant or perceived as annoying. To promote medication adherence, intervention content and approach should be tailored according to such personal characteristics. Providing an app with various modules or features and giving the users the ability to select among them may represent an optimal solution for potentially diverse population.

Engagement

Engagement was defined or operationalized as a microlevel interaction with the interventions but from different dimensions, such as frequency of engagement, time or duration of engagement with the program, and intensity of engagement. There are a couple of possible reasons why subjective experience of engagement such as attention, affect, and interest were not examined in the studies. Self-report of engagement (subjective measure) at multiple points during the intervention or app use may be disrupting [53] and may create excessive burden for users, especially when they are asked to periodically perform different tasks (eg, logging symptoms and BP and reporting medication adherence) using the app. In addition, objective measures (eg, app use data) can be used to capture attention, interest, or affect during the app use or intervention [55]. For instance, mouse cursor tracking or eye tracking can be used to measure attention [62]. Similarly, digital health intervention use (eg, time spent on education content page and messages sent to the health coach) can be used to measure attention or interest in a nonobstructive way or, to some extent, reflect attention or interest. These could be the reasons why studies

used data use to measure engagement but seldom used subjective measure to directly measure attention, affect, or interest.

For all studies that tested this relationship, higher engagement was associated with better biomedical outcomes. However, this result should be interpreted with caution because, first, the result is from 3 applicable studies and, second, although the studies are valuable to the engagement literature, they might have some risk of bias.

Although most studies did not define or focus on engagement per se, their outcome variables reflected engagement with digital behavior change interventions. When reviewing studies of engagement in mHealth, authors should consider seeking out both studies that explicitly mention engagement, as well as studies that do not explicitly define engagement but examine behaviors that reflect engagement.

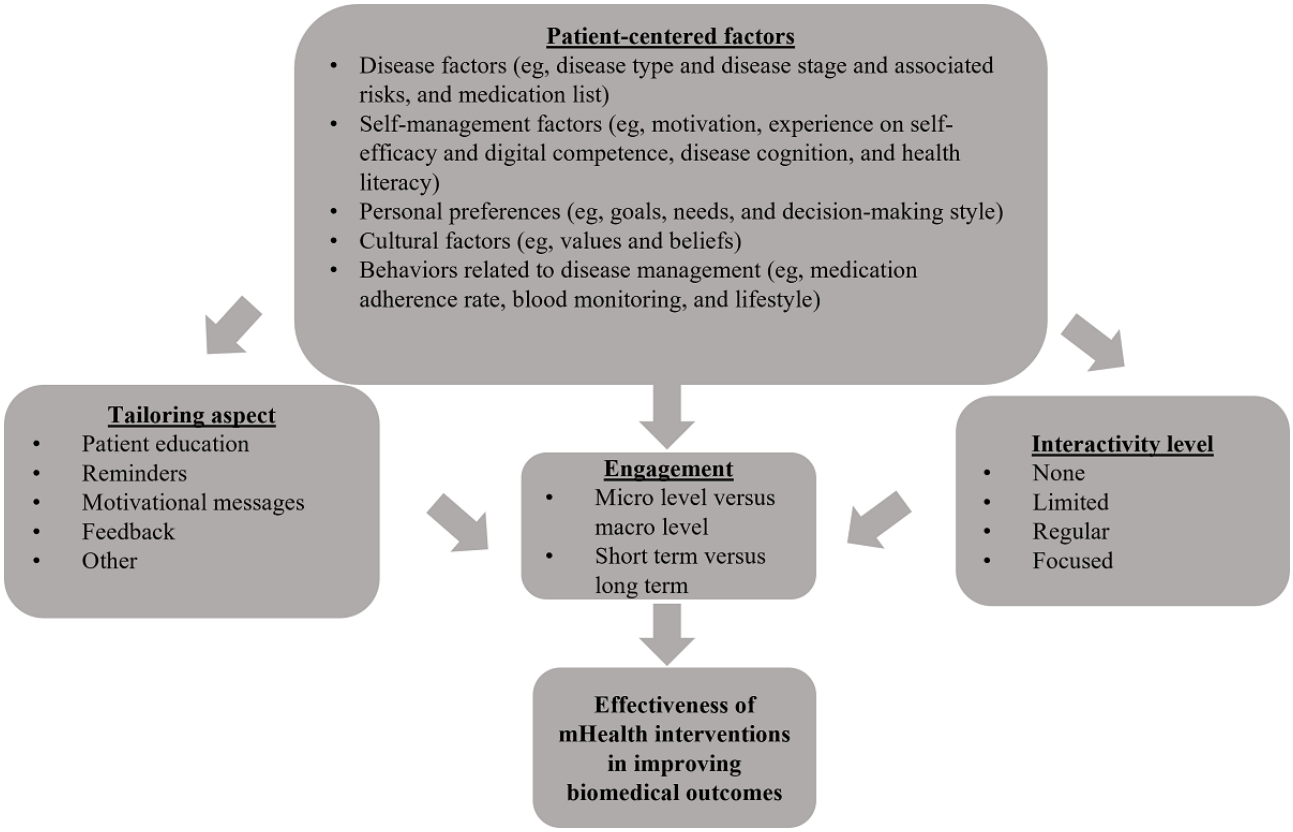
No studies tested macrolevel engagement directly, but 1 study [36] tested a factor crucial to macrolevel and microlevel engagement or behavior change: self-efficacy or self-confidence (eg, using the app, performing nonpharmacologic behaviors to control BP). In 1 study [7], researchers found that patients with hypertension have various levels of digital competence (defined as "becoming familiar and comfortable with using technology to manage hypertension") in that some were not interested in using apps for hypertension management and others were digitally competent to use apps. In another study focusing on African American older adults, "participants expressed concerns about not being informed or trained sufficiently to integrate technology for hypertension self-management" [63]. These findings imply that self-efficacy, especially self-efficacy in using technology for disease self-management, can be another important patient factor to be considered when designing interventions [54,55,64].

Our Patient-Centered Framework

Overall, the results of this study agreed with those of a prior systematic review [65] of mHealth for self-management of cardiometabolic risk factors, in that some studies are theoretically driven, while direct measurement and evaluation of engagement was limited.

Considering the definition of engagement [4], the motivational technology model [20], and the results of the review, a framework for the use of mHealth technology for hypertension self-management is proposed (Figure 2). This patient-centered engagement framework emphasizes the important role of patient-centered factors, including but not limited to disease factors, self-management factors, users' personal preferences, cultural factors, and behaviors related to disease self-management. These factors determine the aspects of an intervention to be tailored and determine the level of interactivity. Patient-centered factors, together with tailoring and interactivity level, determine engagement and subsequently intervention efficacy in improving biomedical outcomes. For instance, Thies et al [14] concluded that their failed intervention was due to lack of attention to patients' eHealth literacy (defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem") [66] and lack of proficiency regarding the chronic disease type. Whereas other engagement frameworks [15] focus on strategies (eg, providing medical information, sending reminders, and tracking health data) to engage patients, our framework highlights categories of patient factors to be considered and how those factors are crucial to customization, interactivity, and engagement.

Figure 2. Patient-centered engagement framework for hypertension self-management using mobile health (mHealth) technology.



Limitations

There are some limitations of this review. Only studies published in English were included in the review, and therefore, there remains potential neglect of important studies published in other languages. Although the authors used a systematic search strategy, other studies meeting the inclusion criteria may have been missed. For instance, those studies not including the key search terms used in the systematic search might be excluded. We were unable to conduct a meta-analysis, owing to the heterogeneity of the studies and outcomes. Further, other factors (eg, navigability) that are also important for engagement and efficacy of interventions were not examined in this review. Given that the proposed patient-centered framework was based on the results of the studies included in the review, it is possible that there are other patient-centered factors (eg, outcome expectation, a significant factor of engagement [55]) that could be important for engagement in the context of hypertension self-management that are not included in the framework.

Conclusions

Among mHealth app interventions focused on hypertensive management, engagement, interactivity, and tailoring have been implemented in various ways, as demonstrated by the 21 studies included in this review. The authors examined several strategies used to facilitate engagement. The results support the essential roles of engagement in intervention effectiveness and the essential roles of patient factors in tailoring, interactivity, and engagement. A patient-centered engagement framework for hypertension self-management using mHealth technology was proposed, with the intent to help facilitate intervention design and disease self-management using mHealth technology in the future.

Acknowledgments

This work was supported by National Cancer Institute of National Institutes of Health (NIH; grant U01CA248240 to LL), by Eunice Kennedy Shriver National Institute of Child Health and Human Development of NIH (grant P30HD106451 to LL), by National Heart, Lung, and Blood Institute of NIH (grant K23HL155890 to DA), and by an American Heart Association–Robert Wood Johnson Foundation Faculty Development Program Grant (to DA).

Authors' Contributions

WC, MWM, and LL conceived the study. WC performed the systematic search. WC and XL screened and coded the studies included for the review. WC analyzed the data and drafted the manuscript. All authors interpreted the results and revised the manuscript critically. All authors gave their final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 14 KB - mhealth_v10i3e29415_app1.docx]

Multimedia Appendix 2
Risk of bias assessment for randomized control trails.
[DOCX File , 15 KB - mhealth_v10i3e29415_app2.docx]

Multimedia Appendix 3
Risk of bias assessment for observational studies.
[DOCX File , 15 KB - mhealth_v10i3e29415_app3.docx]

Multimedia Appendix 4
Risk of bias assessment for pre-post studies without a control group.
[DOCX File , 15 KB - mhealth_v10i3e29415_app4.docx]

Multimedia Appendix 5
Risk of bias assessment for qualitative studies.
[DOCX File , 14 KB - mhealth_v10i3e29415_app5.docx]

References

1. Facts about hypertension. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/bloodpressure/facts.htm [accessed 2021-03-15]
2. Tuberculosis: global task force on digital health: frequently asked questions. World Health Organization. 2015. URL: https://www.who.int/news-room/q-a-detail/tuberculosis-global-task-force-on-digital-health [accessed 2020-12-30]
3. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
4. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
5. Andre N, Wibawanti R, Siswanto BB. Mobile phone-based intervention in hypertension management. Int J Hypertens 2019;2019:9021017 [FREE Full text] [doi: 10.1155/2019/9021017] [Medline: 31080670]
6. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020;8(8):e15779 [FREE Full text] [doi: 10.2196/15779] [Medline: 32459654]
7. Morrissey EC, Casey M, Glynn LG, Walsh JC, Molloy GJ. Smartphone apps for improving medication adherence in hypertension: patients' perspectives. Patient Prefer Adherence 2018;12:813-822 [FREE Full text] [doi: 10.2147/PPA.S145647] [Medline: 29785096]
8. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
9. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
10. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
11. Xiong S, Berkhouse H, Schooler M, Pu W, Sun A, Gong E, et al. Effectiveness of mHealth interventions in improving medication adherence among people with hypertension: a systematic review. Curr Hypertens Rep 2018;20(10):86. [doi: 10.1007/s11906-018-0886-7] [Medline: 30088110]
12. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

13. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
14. Thies K, Anderson D, Cramer B. Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a federally qualified health center: interview analysis of staff and patients in a failed randomized trial. JMIR Hum Factors 2017;4(4):e24 [FREE Full text] [doi: 10.2196/humanfactors.7709] [Medline: 28974481]
15. Shapiro-Mathews E, Barton AJ. Using the patient engagement framework to develop an institutional mobile health strategy. Clin Nurse Spec 2013;27(5):221-223. [doi: 10.1097/NUR.0b013e3182a0b9e2] [Medline: 23942098]
16. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016;5:1-11. [Medline: 26934758]
17. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
18. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
19. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
20. Sundar SS, Bellur SS, Jia H. Motivational technologies: a theoretical framework for designing preventive health applications. In: Proceedings of 7th International Conference on Persuasive Technology. 2012 Presented at: PERSUASIVE '12; June 6-8, 2012; Linköping, Sweden p. 112-122. [doi: 10.1007/978-3-642-31037-9_10]
21. Sundar SS. The handbook of the psychology of communication technology. Hoboken, NJ: Wiley; 2015.
22. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
23. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
24. Al-Noumani H, Wu JR, Barksdale D, Sherwood G, AlKhasawneh E, Knafl G. Health beliefs and medication adherence in patients with hypertension: a systematic review of quantitative studies. Patient Educ Couns 2019;102(6):1045-1056. [doi: 10.1016/j.pec.2019.02.022] [Medline: 30846205]
25. Kang H, Park HA. A mobile app for hypertension management based on clinical practice guidelines: development and deployment. JMIR Mhealth Uhealth 2016;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
26. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A mobile phone-based health coaching intervention for weight loss and blood pressure reduction in a national payer population: a retrospective study. JMIR Mhealth Uhealth 2017;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
27. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
28. McGillicuddy JW, Taber DJ, Mueller M, Patel S, Baliga PK, Chavin KD, et al. Sustainability of improvements in medication adherence through a mobile health intervention. Prog Transplant 2015;25(3):217-223 [FREE Full text] [doi: 10.7182/pit2015975] [Medline: 26308780]
29. Covidence systematic review software. Melbourne, Australia: Veritas Health Innovation URL: https://www.covidence.org/ [accessed 2021-07-30]
30. Bengtsson U, Kjellgren K, Höfer S, Taft C, Ring L. Developing an interactive mobile phone self-report system for self-management of hypertension. Part 2: content validity and usability. Blood Press 2014;23(5):296-306 [FREE Full text] [doi: 10.3109/08037051.2014.901009] [Medline: 24786778]
31. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved blood pressure control using an interactive mobile phone support system. J Clin Hypertens (Greenwich) 2016;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
32. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
33. Hallberg I, Ranerup A, Bengtsson U, Kjellgren K. Experiences, expectations and challenges of an interactive mobile phone-based system to support self-management of hypertension: patients' and professionals' perspectives. Patient Prefer Adherence 2018;12:467-476 [FREE Full text] [doi: 10.2147/PPA.S157658] [Medline: 29643739]

34. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: a proof-of-concept randomized controlled trial. JMIR Res Protoc 2013;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
35. Persell SD, Karmali KN, Stein N, Li J, Peprah YA, Lipiszko D, et al. Design of a randomized controlled trial comparing a mobile phone-based hypertension health coaching application to home blood pressure monitoring alone: the smart hypertension control study. Contemp Clin Trials 2018;73:92-97. [doi: 10.1016/j.cct.2018.08.013] [Medline: 30172038]
36. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open 2020;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]
37. Moore JO, Marshall MA, Judge DC, Moss FH, Gilroy SJ, Crocker JB, et al. Technology-supported apprenticeship in the management of hypertension: a randomized controlled trial. J Clin Outcomes Manag 2014;21(3):110-122.
38. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019;16(7):1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
39. Davidson TM, McGillicuddy J, Mueller M, Brunner-Jackson B, Favella A, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
40. Table 8.5.d: Criteria for judging risk of bias in the 'Risk of bias' assessment tool. The Cochrane collaboration's tool for assessing risk of bias. 2011. URL: https://handbook-5-1.cochrane.org/chapter_8/table_8_5_d_criteria_for_judging_risk_of_bias_in_the_risk_of.htm [accessed 2021-07-10]
41. Study quality assessment tools. National Institutes of Health. 2021. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2021-07-10]
42. Critical appraisal skills programme qualitative checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2021-07-10]
43. 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;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
44. Gong K, Yan YL, Li Y, Du J, Wang J, Han Y, APP Study Group. Mobile health applications for the management of primary hypertension: a multicenter, randomized, controlled trial. Medicine (Baltimore) 2020;99(16):e19715 [FREE Full text] [doi: 10.1097/MD.0000000000019715] [Medline: 32311957]
45. Márquez Contreras E, Márquez Rivero S, Rodríguez García E, López-García-Ramos L, Carlos Pastoriza Vilas J, Baldonedo Suárez A, Compliance Group of Spanish Society of Hypertension (SEH-LELHA). Specific hypertension smartphone application to improve medication adherence in hypertension: a cluster-randomized trial. Curr Med Res Opin 2019;35(1):167-173. [doi: 10.1080/03007995.2018.1549026] [Medline: 30431384]
46. Petrella RJ, Stuckey MI, Shapiro S, Gill DP. Mobile health, exercise and metabolic risk: a randomized controlled trial. BMC Public Health 2014;14:1082 [FREE Full text] [doi: 10.1186/1471-2458-14-1082] [Medline: 25326074]
47. Ovbiagele B, Jenkins C, Patel S, Brunner-Jackson B, Anderson A, Saulson R, et al. Mobile health medication adherence and blood pressure control in recent stroke patients. J Neurol Sci 2015;358(1-2):535-537. [doi: 10.1016/j.jns.2015.10.008] [Medline: 26463572]
48. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. Health Inf Sci Syst 2017;5(1):4 [FREE Full text] [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]
49. Ciemins EL, Arora A, Coombs NC, Holloway B, Mullette EJ, Garland R, et al. Improving blood pressure control using smart technology. Telemed J E Health 2018;24(3):222-228. [doi: 10.1089/tmj.2017.0028] [Medline: 28930497]
50. Toro-Ramos T, Kim Y, Wood M, Rajda J, Niejadlik K, Honcz J, et al. Efficacy of a mobile hypertension prevention delivery platform with human coaching. J Hum Hypertens 2017;31(12):795-800. [doi: 10.1038/jhh.2017.69] [Medline: 28972573]
51. Patel S, Jacobus-Kantor L, Marshall L, Ritchie C, Kaplinski M, Khurana PS, et al. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Technol 2013;7(3):630-639 [FREE Full text] [doi: 10.1177/193229681300700307] [Medline: 23759395]
52. Weerahandi H, Paul S, Quintiliani LM, Chokshi S, Mann DM. A mobile health coaching intervention for controlling hypertension: single-arm pilot pre-post study. JMIR Form Res 2020;4(5):e13989 [FREE Full text] [doi: 10.2196/13989] [Medline: 32379049]
53. Sharek D, Wiebe E. Measuring video game engagement through the cognitive and affective dimensions. Simul Gaming 2014;45(4-5):569-592. [doi: 10.1177/1046878114554176]
54. Bandura A. Self-efficacy: the exercise of control. New York, NY: W. H. Freeman; 1997.
55. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]

56. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the development of mobile phone apps to support diabetes self-management: systematic review. JMIR Mhealth Uhealth 2018;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
57. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: a mixed method study. Health Policy Technol 2017;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
58. Fletcher BR, Hinton L, Hartmann-Boyce J, Roberts NW, Bobrovitz N, McManus RJ. Self-monitoring blood pressure in hypertension, patient and provider perspectives: a systematic review and thematic synthesis. Patient Educ Couns 2016;99(2):210-219. [doi: 10.1016/j.pec.2015.08.026] [Medline: 26341941]
59. Shahin W, Kennedy GA, Stupans I. The impact of personal and cultural beliefs on medication adherence of patients with chronic illnesses: a systematic review. Patient Prefer Adherence 2019;13:1019-1035 [FREE Full text] [doi: 10.2147/PPA.S212046] [Medline: 31303749]
60. Li WW, Stewart AL, Stotts N, Froelicher ES. Cultural factors associated with antihypertensive medication adherence in Chinese immigrants. J Cardiovasc Nurs 2006;21(5):354-362. [doi: 10.1097/00005082-200609000-00005] [Medline: 16966912]
61. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
62. Lagun D, Lalmas M. Understanding user attention and engagement in online news reading. In: Proceedings of the 9th ACM International Conference on Web Search and Data Mining. 2016 Presented at: WSDM '16; February 22-25, 2016; San Francisco, CA p. 113-122. [doi: 10.1145/2835776.2835833]
63. Still CH, Jones LM, Moss KO, Variath M, Wright KD. African American older adults' perceived use of technology for hypertension self-management. Res Gerontol Nurs 2018;11(5):249-256 [FREE Full text] [doi: 10.3928/19404921-20180809-02] [Medline: 30230518]
64. Sieck CJ, Walker DM, Gregory M, Fareed N, Hefner JL. Assessing capacity to engage in healthcare to improve the patient experience through health information technology. Patient Exp J 2019;6(2):28-34. [doi: 10.35680/2372-0247.1366]
65. Delva S, Waligora Mendez KJ, Cajita M, Koirala B, Shan R, Wongvibulsin S, et al. Efficacy of mobile health for self-management of cardiometabolic risk factors: a theory-guided systematic review. J Cardiovasc Nurs 2021;36(1):34-55 [FREE Full text] [doi: 10.1097/JCN.0000000000000659] [Medline: 32040072]
66. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

Abbreviations

BP: blood pressure
mHealth: mobile health
NIH: National Institutes of Health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 09.04.21; peer-reviewed by Y Chen, J Lauffenburger, C Yeager; comments to author 07.06.21; revised version received 01.08.21; accepted 15.12.21; published 02.03.22.

Please cite as:
Cao W, Milks MW, Liu X, Gregory ME, Addison D, Zhang P, Li L
mHealth Interventions for Self-management of Hypertension: Framework and Systematic Review on Engagement, Interactivity, and Tailoring
JMIR Mhealth Uhealth 2022;10(3):e29415
URL: https://mhealth.jmir.org/2022/3/e29415
doi:10.2196/29415
PMID:35234655

Review

Technologies for Medication Adherence Monitoring and Technology Assessment Criteria: Narrative Review

Madilyn Mason[1]; Youmin Cho[1], MSN; Jessica Rayo[1], MSc; Yang Gong[2], MD, PhD; Marcelline Harris[1], PhD; Yun Jiang[1], MS, PhD

[1]School of Nursing, University of Michigan, Ann Arbor, MI, United States

[2]School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Yun Jiang, MS, PhD
School of Nursing
University of Michigan
400 North Ingalls
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 3705
Email: jiangyu@umich.edu

Abstract

Background: Accurate measurement and monitoring of patient medication adherence is a global challenge because of the absence of *gold standard* methods for adherence measurement. Recent attention has been directed toward the adoption of technologies for medication adherence monitoring, as they provide the opportunity for continuous tracking of individual medication adherence behavior. However, current medication adherence monitoring technologies vary according to their technical features and data capture methods, leading to differences in their respective advantages and limitations. Overall, appropriate criteria to guide the assessment of medication adherence monitoring technologies for optimal adoption and use are lacking.

Objective: This study aims to provide a narrative review of current medication adherence monitoring technologies and propose a set of technology assessment criteria to support technology development and adoption.

Methods: A literature search was conducted on PubMed, Scopus, CINAHL, and ProQuest Technology Collection (2010-present) using the combination of keywords *medication adherence*, *measurement technology*, and *monitoring technology*. The selection focused on studies related to medication adherence monitoring technology and its development and use. The technological features, data capture methods, and potential advantages and limitations of the identified technology applications were extracted. Methods for using data for adherence monitoring were also identified. Common recurring elements were synthesized as potential technology assessment criteria.

Results: Of the 3865 articles retrieved, 98 (2.54%) were included in the final review, which reported a variety of technology applications for monitoring medication adherence, including electronic pill bottles or boxes, ingestible sensors, electronic medication management systems, blister pack technology, patient self-report technology, video-based technology, and motion sensor technology. Technical features varied by technology type, with common expectations for using these technologies to accurately monitor medication adherence and increase adoption in patients' daily lives owing to their unobtrusiveness and convenience of use. Most technologies were able to provide real-time monitoring of medication-taking behaviors but relied on proxy measures of medication adherence. Successful implementation of these technologies in clinical settings has rarely been reported. In all, 28 technology assessment criteria were identified and organized into the following five categories: *development information*, *technology features*, *adherence to data collection and management*, *feasibility and implementation*, *and acceptability and usability*.

Conclusions: This narrative review summarizes the technical features, data capture methods, and various advantages and limitations of medication adherence monitoring technology reported in the literature and the proposed criteria for assessing medication adherence monitoring technologies. This collection of assessment criteria can be a useful tool to guide the development and selection of relevant technologies, facilitating the optimal adoption and effective use of technology to improve medication adherence outcomes. Future studies are needed to further validate the medication adherence monitoring technology assessment criteria and construct an appropriate technology assessment framework.

(JMIR Mhealth Uhealth 2022;10(3):e35157) doi:10.2196/35157

KEYWORDS

medication adherence; technology assessment; remote sensing technology; telemedicine

Introduction

Background

Accurately measuring and monitoring patient medication adherence is critical in clinical practice and research settings but continues to be a challenging task globally [1]. Various methods are used to measure medication adherence, such as patient self-reports, pill counts, pharmacy refill records, drug metabolites or biomarker testing, and directly observed therapy (DOT) [1]. However, none of these methods have been accepted as a standard measure of medication adherence across a variety of settings [2]. More recently, sensor technologies have been increasingly used to track the medication-taking behaviors of patients [1]. For example, the Medication Event Monitoring System (MEMS) can record every time the patient opens the pill bottle via a sensor embedded in the pill cap [3,4]. Such technologies provide a unique opportunity to measure and monitor patient medication adherence over time [1]. The notion that medication adherence monitoring technology represents the *gold standard* of measurement of patient medication adherence has been voiced by some researchers [3-9] but continues to be disregarded by others [10-15]. There is limited consensus on how to determine or select the appropriate medication adherence monitoring technology for use, which may be due to the lack of appropriate technology assessment criteria in this field.

The advantages and limitations of the commonly used methods for measuring medication adherence have been described in the literature. For example, DOT allows for direct observation of patient medication-taking actions [16-18], but it is expensive to sustain and produces a constrictive time strain on both health care providers (HCPs) and patients' daily schedules [1,12,18,19]. As a common way to measure medication adherence, patient self-reporting respects patient autonomy but carries the potential risk of patient overestimation or underestimation of their adherence abilities [20-22]. Medication adherence monitoring technologies with various types and features are being continuously developed and upgraded [1]. Some newly developed technologies may possess unique features that are unfamiliar to users [23]. Despite this literature, there is no summary or synthesis that reflects a clear understanding of the characteristics and values of a variety of medication adherence monitoring technologies. There is a growing need for technology assessment criteria to guide the development and selection of appropriate technologies for monitoring medication adherence to improve patient outcomes [24].

Stakeholders' expectations regarding the use of health information technology for monitoring medication adherence also vary. From a clinical practice perspective, a user-friendly interface and the accurate monitoring of adherence are considered when selecting appropriate monitoring technologies [1]. From the technological development perspective, although system accuracy and data fidelity remain high priorities, developers also need to consider the feasibility of technical engineering of the system, such as energy consumption and battery lifetime [25]. Advanced medication adherence monitoring technologies may not be limited to a single method to gather patient medication adherence information [1]. In addition, human interactions with these technologies can be complicated owing to the comprehensive medical and pharmacological contexts, as well as multidimensional patient medication adherence behaviors [22]. A compiled summary and assessment of currently available applications of medication adherence monitoring technologies is important for a better understanding of their capacities and performance when making decisions for their adoption and use.

Objectives

The purpose of this narrative review is to summarize literature reports on the current applications of medication adherence monitoring technologies and identify potential assessment criteria to support decisions related to technology development and adoption.

Methods

Literature Search

PubMed, Scopus, CINAHL, and ProQuest Technology Collection databases were searched because of their broad collection of literature focusing on health, health care, and technological domains. A combination of search terms was included as follows: (*medication adherence*) AND (*measurement technology* OR *monitoring technology*). A full list of search strategies used for each database is included in Multimedia Appendix 1. To gather the most recent collection of medication adherence monitoring technologies, searches were focused on scholarly articles published between January 2010 and June 2021 and written in English.

Eligibility Criteria

Studies were included in this narrative review if they met the following criteria: (1) described the development of medication adherence monitoring technologies, (2) assessed the characteristics of medication adherence monitoring technologies, or (3) tested the application of technologies for monitoring medication adherence. All study methods were included. Only articles published in English with their full text available were included. Considering the ease of dispensing medication and self-administration of pill form of medications, we focused on medication adherence technologies suited for pills. Medication adherence monitoring technologies that suited nonpill forms of medications, such as inhalers, eye drops, and injectable medications, were excluded. Studies that did not provide adequate descriptions of technology characteristics or used technologies that did not monitor patient medication adherence were also excluded. Study selection was performed manually using this set of eligibility criteria.

Data Extraction and Information Synthesis

Data concerning medication adherence monitoring technologies were manually extracted from the reviewed articles by the first author (MM) and discussed with the research team. These elements included the following: (1) type of technology, (2) name of technology, (3) technical features, (4) data capture and applications, (5) perceived advantages, and (6) limitations of the identified technologies. Data directly related to publications, such as the country and publication year, were also gathered. Information regarding adherence monitoring technologies was extracted and organized into a table for further synthesis (Multimedia Appendix 2 [1-6,9-13,15-18,20,26-99]).

A descriptive analysis of the characteristics of the selected studies was conducted. Key characteristics, including the technical features, data capture methods, advantages, and limitations of each technology type, were assessed and summarized. Common and recurring elements were coded and categorized as potential assessment criteria. All identified potential criteria were discussed and evaluated among the team members until a consensus was reached. The final criteria were organized into categories and subcategories and presented as a matrix.

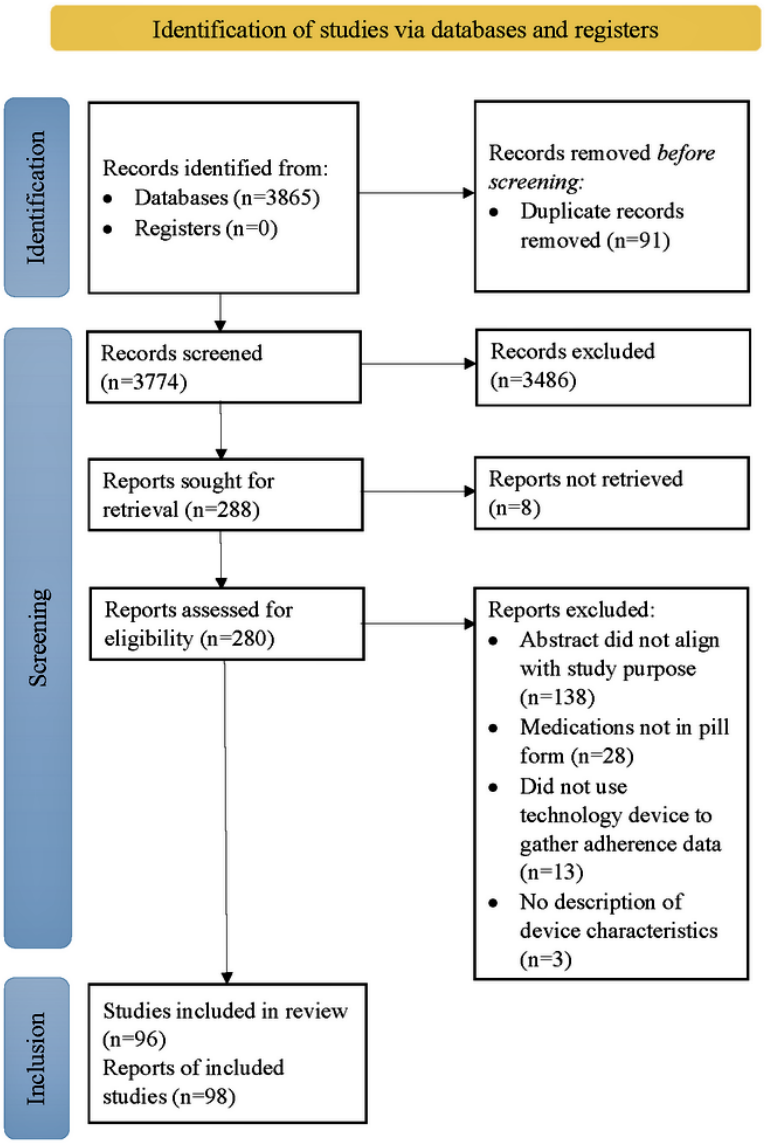
Results

Study Selection and Characteristics

A total of 3865 records were retrieved from the database search. Of these 3865, the removal of duplicates left 3774 (97.65%) articles for title screening. After reviewing the titles and abstracts for relevance, 7.63% (288/3774) of the articles were identified for retrieval, of which 97.2% (280/288) were successfully gathered. Of the 288 articles, 8 (2.8%) articles were not retrievable because their full text was not available on the web. Following the assessment for eligibility via full-text review, 35% (98/280) of the articles were included in the final analysis. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart describing the overall search and selection process.

Among the reviewed articles, the vast majority (72/98, 73%) were published between January 2015 and June 2021. Over half (50/98, 51%) of the identified studies were published in the United States, followed by Canada (8/98, 8%), and Japan (4/98, 4%). The study types and designs varied greatly among the 98 reviewed articles. Most studies (41/98, 42%) were pilot tests of feasibility, acceptability, usability, or proof of concept. Only a few studies were randomized controlled trials, including pilot randomized controlled trials (5/98, 5%), retrospective cohort studies or secondary data analyses (6/98, 6%), or qualitative studies (8/98, 8%). Literature review articles (8/98, 8%), study protocols (4/98, 4%), and commentary and editorial comments (2/98, 2%) were included in the review analysis. The most common medications studied were tuberculosis treatment regimens (19/98, 19%) and antiretroviral therapy for HIV (16/98, 16%).

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



Technology Types and Characteristics

Overview

A total of 81% (79/98) of applications of medication adherence monitoring technology were reported in the 98 reviewed articles. The identified technology types were categorized into eight major groups based on their technical designs and adherence monitoring functions*: electronic pillboxes or bags*, *electronic pill bottles*, *ingestible sensors*, *blister pack technology*, *electronic medication management systems*, *patient self-report–based technology*, *video-based technology*, and *motion sensor technology*. Table 1 shows the number of articles for each technology type group. As noted, some articles have reported more than one type of technology. The following sections outline the common defining technical features, data capture methods, and advantages and limitations gathered from the existing literature for each technology type.

Table 1. The number of publications by technology type (n=98).

Technology type	Publications, n (%)
Electronic pill boxes or bags	32 (33)
Electronic pill bottles	25 (26)
Ingestible sensors	22 (22)
Electronic medication management systems	12 (12)
Patient self-report technology	12 (12)
Blister pack technology	10 (10)
Video-based technology	7 (7)
Motion sensor technology	3 (3)

Electronic Pill Bottles

Consisting of a standard size pill bottle and an electronic cap that contains a microchip, this type of technology records a date-and-time stamp once the cap has been removed during an opening event. The transfer of adherence data depends on the version of the electronic pill bottle device. Some old versions of the MEMS often require manual download of the stored patient medication adherence data from the MEMS cap into the MEMS software through a small reader device [8-10,27,28]. Some newer versions of electronic pill bottle technologies, such as the GlowCap and AdhereTech devices, possess the ability to wirelessly transmit patient medication adherence data, providing opportunities to assess and monitor patient medication adherence in real time [1,15,18,29-31]. Electronic pill bottle technologies are commonly reported to have advantages including their discrete design and small size [10,16,32,33], objective medication adherence monitoring ability [1,8,11,14,34,35], and acceptance among patients [1,30,31]. However, as the pill bottle design is only capable of storing 1 type of medication at a time, these devices are not suitable for patients with complex multidrug regimens [1,4,8,9,18,30,32]. In addition, because the opening of the pill bottle is used as a proxy measure for adherence, patient actions such as failing to ingest removed medications, *pocket dosing* (taking out multiple doses of medications at one time), and *curiosity openings* (opening the device but not removing medications) can lead to inaccurate estimates of patient medication adherence [1,2,4,5,9-11,14-16,27,28,30,33,36,37].

Electronic Pill Boxes or Bags

Similar to the electronic pill bottle technology, electronic pill boxes or bags record a date-and-time stamp whenever they are opened. However, unlike electronic pill bottles, these technologies can often store multiple types or strengths of medication in various compartments within the device. The size and storage capacity vary among the different types of available electronic pill boxes or bags. Most of the identified electronic pill boxes or bags possessed the ability to transmit patient medication adherence data in real time via existing cellular networks [1,9,26-28,38-43,100], wireless Bluetooth [1,44,45], or general packet radio service [27,46,101,102]. One device required manual uploading of patient adherence data during clinic visits [47]. Although the capability of these devices to store multiple medications makes them better suited for complex multidrug regimens, this advantage is dependent on the device, as they can vary drastically in size and pill storage capacity. This was evident when examining the Wisepill device's storage capacity of 60 small-sized pills [1,27,39,41,46] compared with MedTracker's storage capacity of a week's worth of medication [44]. However, larger-sized devices are often described as obtrusive [10,40,48] and have increased risks to patient privacy [1,39,40,49], thus limiting the acceptability of the device for patient populations, particularly for those who do not wish to disclose their health status (eg, HIV positive) to others [39]. Furthermore, these devices cannot directly confirm ingestion of medications, raising concerns toward their medication adherence monitoring accuracy because potential patient behaviors, including *pocket dosing* and *curiosity events*, may impede medication adherence rate estimates [1,10,15,16,26,27,30,38-40,44,48-50,101]

Blister Pack Technologies

All but 3 blister pack technology applications identified included an attachable adhesive label that contained a microchip and conductive wire pattern [4,36,51]. Removing medication from the blister pack created a break in the label circuit and was recorded by the microchip with a date-and-time stamp. Patient medication adherence data are wirelessly transmitted to central servers and are often accessible to HCPs, allowing for real-time adherence monitoring [1,6,51-55]. As the design of blister packs stores the appropriate dose of medication in singular pockets, each removed dose is registered as an individual event, thereby eliminating the potential for patient *pocket dosing* and *curiosity openings* [56]. However, currently within these devices, the action of removing a dose has been found to break the conductive tracks of the surrounding doses occasionally and accidentally, leading to the registration of multiple removal events, which decreases the accuracy of monitoring with these technologies [56,57]. Moreover, this method of medication adherence monitoring is a proxy measure and cannot confirm patient ingestion of medication, further limiting the accuracy of patient medication adherence estimates [1,4,52,55].

Ingestible Sensors

Ingestible sensors, otherwise known as digital pills [12] or digital ingestion monitoring [50], consist of a technological system that includes microsensors, an adhesive external monitor worn on the abdomen, and a mobile app. The microingestible sensors are coencapsulated with medication and ingested into

the body, where stomach gastric fluids dissolve the capsule containing the medication and sensor. Activation of the sensor upon contact with gastric fluid transmits a unique signal to the external monitor. The detected ingestion event is transferred to a mobile app that uploads the event's date-and-time stamp, along with other recorded physiological measures (eg, heartbeat), to a central server. These technological systems possess the advantage of direct observation of medication ingestion [3,13,15,16,30,37,58-65], as well as real-time adherence monitoring [1,3,9,12,30,37,49,58,61-66,103]. By directly identifying individual ingestion events of medication, these technologies can detect multiple ingestion events at a given time, thereby improving the accuracy of measuring patient medication adherence rates [60,61,65,103]. In addition, the ingestion event detection accuracy of ingestible sensors is high, with rates of 95% to 99.1% observed experimentally [3,58,65,103]. However, owing to the direct ingestion of technological sensors, concerns over patient privacy and autonomy are prominent because of the invasive nature of these devices [9,13,20]. Patient reports of skin irritation caused by the external monitor [1,12,40,59,61,63,64] and the possibility of sensor retention within the body [15,60,64] are considerable limitations of these technologies, as well as potential risks to patient health and safety.

Electronic Medication Management Systems

The devices identified within the category of Electronic Medication Management System (EMMS) vary in their functionalities, with reported advantages and limitations; however, all systems possess similar features that focus on aiding patients in their medication management and documenting their medication adherence patterns. Three novel EMMS devices that presented interesting functionality characteristics included the radio frequency identification (RFID)-based medication adherence intelligence system [44,67], ReX (DosentRx Ltd) [68], and the Medication Behavior Monitoring System (MBMS) [69]. The RFID-based medication adherence intelligence system (RMAIS) is composed of an RFID reader, scale, microcontroller, liquid crystal display panel, and a motorized rotation platform [44,67]. The patient's pill bottles are labeled with an RFID tag that stores the medication's information, such as the medication name and appropriate dose [67]. At a scheduled medication administration time, the RMAIS generates audio medication reminders and rotates the correct pill bottle in front of the patient [44,67]. The scale underneath the rotation platform weighs the pill bottle, and the medication information is displayed using an RFID reader [44,67]. After the patient has removed the medication from the pill bottle, the scale measures the weight of the bottle and uses the difference in weight to determine the number of doses removed [44,67]. If the system detects events of nonadherence, an HCP is alerted [44,67]. An advantage of this system is that it provides guidance to patients who must navigate complex multidrug regimens by eliminating the need for patient decision-making concerning what medication to take, how much, and at what time [67]. However, because this system is also a proxy monitor of medication adherence and cannot confirm the actual ingestion of medication, its accuracy is consequently limited [44].

ReX is a recently developed device composed of a reusable drug dispensing unit, disposable cassette, mobile app, and a Dose-E Analytics cloud system [68]. The patient's medication is stored inside the device and can only be released at the appropriate time, at the correct dose, and directly into the patient's mouth [68]. The mobile app transfers patient medication adherence data from the drug dispensing unit to the Dose-E Analytics cloud system, which is accessible to HCPs, allowing real-time medication adherence monitoring [68]. A critical advantage of the device is the dispenser mechanism that prevents patient medication overadherence and administration of medication at incorrect time intervals [68]. However, even though the device can monitor the medication up until delivery into the patient's mouth, it cannot confirm the actual ingestion of the medication, thereby inhibiting the accuracy of its medication adherence estimates.

Finally, MBMS devices use newly emerging technologies such as the Internet of Things, deep learning, and artificial intelligence [67]. The MBMS is unique in that it combines the following three categories of medication adherence monitoring technologies: electronic pillboxes, motion sensor technology, and video-based monitoring technology [67]. The device uses a set alarm to remind patients to take their medication [67]. As the patient approaches the device, motion sensors placed around the patient's home detect the movement and signal the MBMS device to begin recording a video of the patient's medication behavior [67]. Once the device recognizes the patient's act of raising an arm to drink water, the internal pillbox that stores the medication releases the appropriate medication and quantity onto a platform with a scale [67]. The MBMS determines whether the patient takes the dispensed medication based on whether the scale converges to zero [67]. HCPs receive weekly adherence reports from the MBMS. Roh et al [69] found that when an MBMS device was used, medication adherence was higher than in patients who did not use the device. However, similar to RMAIS and ReX, the inability of the system to detect actual medication ingestion inhibits its potential accuracy in monitoring patient medication adherence.

Video-Based Monitoring Technology

Similar to DOT, where patients administer their medication in the presence of an HCP, most video-based adherence monitoring technologies use video cameras for patients to self-record medication ingestion event videos for retrospective analysis by HCPs or, in 2 unique cases, by artificial intelligence [70-72]. Video-DOT (VDOT) was the most common technological method for this category of technology. Patients either ingest their medication during a synchronous video call with their HCP or upload an asynchronous video for the HCP to review [16,18,48,73,74]. Real-time medication adherence monitoring is facilitated by the direct and continuous use of medication ingestion event observation by HCPs [49,74]. An additional advantage of VDOT is that, compared with DOT, VDOT is considered more flexible, cheaper, and less intrusive to HCPs and patients [16,49,73]. However, several potential limitations to VDOT include technical difficulties, such as poor video quality [73], trouble uploading ingestion event videos [16], and complications with video camera devices [74]. There is also a potential risk of patients forgetting to self-record as they ingest

their medication, which may lead to inaccurate reports of medication adherence [16].

Motion Sensor Technology

Currently, the medication adherence monitoring motion sensor technologies that we have identified are still under development. Three individual adherence-monitoring motion sensor devices were found, yet all their functionalities were similar. These devices were worn on wrists and resembled the size of a wristwatch [13,75,76]. The wrist-worn devices were triaxial accelerometers that identified the medication administration movements of patients [13,75,76]. Patient medication adherence data were then stored and uploaded to an HCP-accessible database in real time [75]. Wang et al [13] reported a correct ingestion event detection rate of 84.17%. Given that the action of administering medication closely resembles other everyday actions such as eating, drinking, or wiping one's mouth, the accuracy of these technological systems is currently limited [75]. Despite these limitations, motion sensor technologies possess the advantages of being noninvasive [75] and nonintrusive [76] methods of medication adherence monitoring.

Patient Self-report Technology

Similar to EMMS, patient self-report technologies vary in their specific functionalities, yet they all gather subjective medication adherence data by interacting with the patient via phone calls [16,18,26,38,49,53,77-80], smart buttons [55], eDiaries [81], web-based platforms [82,83], and mobile apps [84]. Patient adherence is available in real time for most self-reported devices [18,49,53,78-80,82-84]. Compared with objective adherence monitoring technologies, patient self-report technologies are lower in cost [26,53] and less stigmatizing [16]. Nevertheless, because this technological method of adherence monitoring is subjective, there is a high potential for inaccurate medication adherence reporting by patients, negatively impacting the accuracy of these technologies [16,38,78,79].

A summary of the defining characteristics, data capture methods, and use of data in patient adherence monitoring for each technology type is presented in Table 2. The full details are included in Multimedia Appendix 2.

Table 2. Summary of the defining characteristics, data capture methods, and use of data for patient medication adherence monitoring for each technology category.

Technology category	Defining characteristics	Data capture methods	Use of data for adherence monitoring
Electronic pill bottles	Standard size pill bottles with electronic caps that contain microchips to detect opening events	Opening events of the pill bottle are date-and-time stamped	Recorded opening events act as a proxy measure for medication ingestion
Electronic pill boxes and bags	Devices shaped as pill boxes or bags. Sizes of devices vary. Within each device is a microchip that detects opening events	Opening events of the device are date-and-time stamped	Recorded opening events act as a proxy measure for medication ingestion
Blister pack technologies	Most of these devices are attachable adhesive labels containing a microchip and conductive wire pattern applied to standard blister packs[a]	Breakages in the conductive wire track are recorded as *opening* events and date-and-time stamped	Recorded opening events act as a proxy measure for medication ingestion
Ingestible sensors	Pills embedded with ingestible microsensors that are paired with an external wearable sensor and mobile app	Contact with gastric environment activates microsensor which transmits a signal to the external monitor and is recorded with a date-and-time stamp	Direct measure of medication ingestion events
EMMS[b]	Devices that aim to aid patients in managing medication administration by controlling the type of medication, dosage, or timeframe that medications are accessible[a]	Systems dispense medications and record date-and-time stamps of these events. For example, using scales to detect differences in the device's weight and calculating the amount of medication removed by the patient[a]	Most systems used technologies such as scales and medication dispensing events as proxy measures for medication ingestion
Video-based monitoring technology	Systems that used video cameras to capture patients' medication ingestion events	Video recording of medication-taking events which are later verified by reviewers	Substitute for DOT[c]
Motion sensor technology	Devices are worn on the wrists and contain motion sensing gyrometers and accelerometers to detect patient medication-taking behaviors	Wearable gyrometers and accelerometers identify and record patient motions that match previously programmed medication-taking movements	Physical motions of patients used as a proxy for medication ingestion
Patient self-report technology	Devices that gather adherence data via patient reporting[a]	Patients report medication-taking events via phone calls or other electronic means, such as mobile apps or web-based platforms[a]	Patient reports act as subjective indicators of medication ingestion events

[a]More examples and the full list of features and functions is provided in Multimedia Appendix 2.

[b]EMMS: Electronic Medication Management System.

[c]DOT: directly observed therapy.

Medication Adherence Monitoring Technology Assessment Criteria

Categories Identified

During the data extraction process, common characteristics, recurring elements, and the reported advantages and limitations of all medication adherence monitoring technologies were synthesized and categorized into a set of adherence monitoring technology assessment criteria. These assessment criteria were not categorized by technology type, as various potential assessment criteria were commonly expressed across technologies, suggesting the plausibility of general assessment criteria for all medication adherence monitoring technologies. All 28 specific criteria were included under the following five assessment categories: *development information*, *technology features*, *adherence data collection and management*, *feasibility and implementation*, *and acceptability and usability*. Each category possesses the main feature of interest that allows and supports medication adherence monitoring or measurement. A brief description of each assessment category is provided in the following sections.

Development Information

The development information category contains components related to the general development information of the medication adherence monitoring technology of interest. This category should include information regarding the developer, development stage, commercial availability, and regulatory approval status of organizations such as the Food and Drug Administration.

Technology Features

They contain criteria directly related to the technological setup of medication adherence monitoring technologies. This category includes the following two subcategories: device or hardware and system or software features. The assessment elements of device size, battery life, medication storage capacity, installation or software needs, and the need for wireless connection are considered device or hardware features. System or software feature assessment includes reminder and alert functions, device

accommodation for complex medication regimens, and information technology support availability.

Adherence Data Collection and Management

This category pertains to methods for the capture of medication adherence data and the use of such data. This assessment category was subdivided into data collection and management categories. In data collection, the assessment focuses on subjective versus objective data collection, proxy data collection, date-and-time stamps, and the potential for data entry errors. Data management pertains to the assessment of transmission and upload methods, data display and summary, real-time monitoring capabilities, data accessibility by HCPs, and data security.

Feasibility and Implementation

This category focuses on the components necessary or related to the use of the technology in real-world settings. In addition to device cost efficiency, the interoperability of the technology with current clinical systems should also be considered.

Acceptability and Usability

This is the last category, examining the interaction and relationship between the technology of interest and technology users. These elements include ease of learning and use, device portability, potential risks to patient privacy, and technology-related harms, such as risks to patient health or safety.

All assessment categories and criteria are listed within an organized matrix structured to support technology development and adoption (Textbox 1).

Textbox 1. Medication adherence monitoring technology assessment criteria.

Development information

- Developer
- Development stage
- Regulatory approval status
- Commercial availability

Technology features

- Device or hardware
  - Size
  - Battery life
  - Storage capacity
  - Installation or additional software needed
  - Wireless connection needed
- System or software
  - Reminder and alert function
  - Accommodation for complex medication regimens
  - Information technology support availability

Adherence data collection and management

- Data collection
  - Subjective vs objective data collection
  - Proxy data collection
  - Date-and-time stamps
  - Data entry error (eg, curiosity opening and sensor retention)
- Data management
  - Data transmission and upload methods
  - Data display and summary
  - Real-time monitoring
  - Data accessibility by health care providers
  - Data security

Feasibility and implementation

- Cost efficiency
- Interoperability with current clinical systems

Acceptability and usability

- Ease of learning
- Ease of use
- Portability
- Risks to patient privacy
- Risks to patient health or safety (eg, skin rashes)

Discussion

Principal Findings

As the adoption and development of medication adherence monitoring technologies continue to increase, understanding their key characteristics is vital. This narrative review provides an overview of the technical features, data capture methods, and advantages and limitations of current medication adherence monitoring technologies reported in the literature and synthesizes 28 technology assessment criteria that can be used to guide the development and selection of relevant technologies. Overall, there were 8 types of medication adherence monitoring technologies, dominated by electronic pill bottles, electronic pill boxes or bags, and ingestible sensors. Although technical features varied by technology type, there were common expectations regarding the advantages of using these technologies for accurately monitoring medication adherence and increasing the adoption of these technologies in patients' daily lives.

Current Technology Characteristics

All current medication adherence monitoring technologies have varying degrees of technological restriction. The most commonly reported technology types, electronic pill boxes or bags and electronic pill bottles that use opening events as a proxy for medication ingestion, face undesired patient behaviors such as *pocket dosing* and *curiosity openings*, which are obstacles to the device's accuracy for patient adherence estimates [2,9,11,14,16,27,30,33,37,41,48-50]. Despite this limitation, the popularity of developing and using pill monitoring devices remains, which may be due to their unobtrusiveness and convenience of use in patients' everyday routines, suggesting an increasing adoption of objective measurement and monitoring of medication adherence through technological approaches.

Although electronic pill bottles possess extensive histories of being used in both clinical and research settings, the presence of many other medication adherence monitoring technology studies in the pilot and feasibility phase implies that the integration of newer technologies, such as motion sensor–based technologies and ingestible sensors, is still relatively new and ongoing [1,4,15,18,60,64,76]. Overall, technologies capable of monitoring patient medication adherence provide significant advantages, including real-time medication adherence data reporting, yet questions concerning the accuracy of these devices prohibit them from becoming a *gold standard* in clinical and research standings. Thus, until further developments in medication adherence monitoring technologies occur, multiple methods for patient medication adherence assessment must be used to evaluate patient medication adherence rates and behaviors [1].

Many medication adherence monitoring technologies possess software to organize patient medication adherence data to an extent; however, most of these devices require separate analysis and quantification of the data by HCPs or researchers [1], creating a significant burden of time consumption and the concern of further data integration with other technology applications. The development of a patient medication adherence data management software that can construct automatic visualizations of patient medication adherence estimates should be considered to provide an easy interpretation of patient medication adherence patterns. Moreover, the use of advanced software for adherence data processing and presentation may improve the adoption and integration of medication adherence monitoring technologies in clinical settings.

Medication Adherence Data Capture and Use

In addition to variances in technical features, current medication monitoring technologies differ in their data capture methods and the subsequent use of such data in relation to patients' medication adherence assessment. The ability of most medication adherence monitoring technologies to provide real-time observations of patient medication adherence behaviors is beneficial to HCPs and researchers to prevent nonadherence and facilitate appropriate interventions [1,15,36,37,79]. However, most of these technologies rely on proxy measures of medication adherence, such as device opening events, thereby limiting their data accuracy [2,9,11,14,16,27,30,33,37,41,48-50]. Furthermore, successful implementation of these technologies in clinical settings or the integration of patient medication adherence monitoring data into clinical practice has rarely been reported. One of the major barriers is the interoperability of these monitoring technology systems with established clinical information systems and workflow. To facilitate the adoption of medication adherence monitoring technologies in clinical systems to improve patient care, the method of adherence data capture must be feasible for targeted patients and the acquired data must be easily integrated into standard electronic health record systems. The medication adherence data capture methods and data use presented in this review can help guide HCPs and researchers toward the appropriate selection of medication adherence monitoring technology. Developers must also consider the implications of medication adherence data capture within clinical and research settings to ensure greater ease of use for both patients and providers.

Technology Assessment Criteria

To the best of our knowledge, this is the first collection of assessment criteria focused on technologies to monitor patient medication adherence. The proposed assessment criteria include five major categories as follows: development information, technology features, adherence to data collection and management, feasibility and implementation, and acceptability and usability. The identified criteria highlight significant aspects of medication adherence monitoring technologies that must be considered during technology development and adoption. For example, an important component of medication adherence monitoring technology implementation is cost; however, a common limitation of these technologies is their expensive price tags [1,2,5,7,9,12,18,27,38,54,78,85,102]. The proposed criteria emphasize the cost efficiency of medication adherence monitoring technologies within the feasibility and implementation category. The high cost of devices restricts their adoption in clinical and research settings because other methods, such as patient self-reports, are significantly cheaper [1,78].

As this compilation of assessment criteria was formed by reviewing the current literature, other existing challenges within

technology acceptance or technology design features were also addressed, such as risks to patient privacy or the effect of large device size on user adoption owing to daily life inconveniences [25]. Given the multifaceted nature of the proposed assessment criteria, they can be used to guide the improvement of these technologies for better medication adherence measures and monitoring.

In addition, our set of proposed assessment criteria possessed a structure similar to that of other validated mobile health assessment frameworks. For example, a pyramid for app evaluation framework, proposed by Henson et al [104] and adapted by the American Psychiatric Association as the App Evaluation Model, introducing a similar 5-level structure of evaluation categories, including access and background, privacy and security, clinical foundation, usability, and data integration toward therapeutic goals [104,105]. Similarly, each category covers a few specific evaluation criteria; for example, ease of use is assessed under the usability category [105]. Certain general technology assessment criteria can be applied to both mobile health apps and medication adherence monitoring technology, such as usability, privacy and security, and data integration. However, adherence monitoring technology possesses technical features to support medication storage and management, which results in its unique assessment criteria, such as the medication storage capacity of the device or date-and-time stamps indicating medication-taking actions. The collection of medication adherence monitoring technology assessment criteria was generated from an extensive literature review and information synthesis, which demonstrates its solid evidence foundation but also suggests that further empirical tests and validation are needed in the future.

Limitations

This narrative review has some limitations. First, our database selection and search strategies might not have been sufficiently extensive to capture all published literature. Moreover, we limited the studies to those published in English, potentially excluding other existing medication adherence monitoring technologies from non-English sources. The proposed medication adherence monitoring technology assessment criteria are representative of the elements identified in our literature review and synthesis, which are subject to further validation and evaluation. We did not review detailed information published by specific manufacturers. Finally, given that the scope of this review was focused on medication adherence technologies used for the monitoring of pill form medications, the assessment criteria and the rest of our findings may not be generalizable to all types of medication. It is noteworthy that a large proportion of the identified articles were pilot or feasibility studies. Consequently, our assessment domain of the criteria may also be limited to the early stages of technology development.

Conclusions

Overall, this narrative review presents a summary of the current technological features and data capture methods, reports the advantages and limitations of medication adherence monitoring technologies for pill form medications, and proposes a potential technology assessment criteria. Our constructed assessment criteria are crucial for the development and adoption of these technologies. Specifically, further technological development is required to expand the interoperability of medication adherence monitoring technology systems in clinical settings. The increased implementation of technologies that monitor patient medication adherence has demonstrated the potential to improve patient medication adherence behaviors. Although this technological method of patient medication adherence monitoring cannot be defined as the *gold standard* method for medication adherence monitoring, the functionalities that they possess may improve patient medication adherence and support greater patient health outcomes over time.

Acknowledgments

This study was supported by the Agency for Healthcare Research and Quality (grant R01HS027846).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOC File , 31 KB - mhealth_v10i3e35157_app1.doc]

Multimedia Appendix 2
Summary of medication adherence monitoring technologies.
[DOCX File , 57 KB - mhealth_v10i3e35157_app2.docx]

References

1. Park LG, Howie-Esquivel J, Dracup K. Electronic measurement of medication adherence. West J Nurs Res 2015 Jan;37(1):28-49. [doi: 10.1177/0193945914524492] [Medline: 24577868]

2. Walter T, Wang L, Chuk K, Ng P, Tannock IF, Krzyzanowska MK. Assessing adherence to oral chemotherapy using different measurement methods: lessons learned from capecitabine. J Oncol Pharm Pract 2014 Aug;20(4):249-256. [doi: 10.1177/1078155213501100] [Medline: 24022406]
3. Profit D, Rohatagi S, Zhao C, Hatch A, Docherty JP, Peters-Strickland TS. Developing a digital medicine system in psychiatry: ingestion detection rate and latency period. J Clin Psychiatry 2016 Dec;77(9):1095-1100 [FREE Full text] [doi: 10.4088/JCP.16m10643] [Medline: 27379966]
4. De Bleser L, De Geest S, Vandenbroeck S, Vanhaecke J, Dobbels F. How accurate are electronic monitoring devices? A laboratory study testing two devices to measure medication adherence. Sensors (Basel) 2010;10(3):1652-1660 [FREE Full text] [doi: 10.3390/s100301652] [Medline: 22294891]
5. Nerini E, Bruno F, Citterio F, Schena FP. Nonadherence to immunosuppressive therapy in kidney transplant recipients: can technology help? J Nephrol 2016 Oct;29(5):627-636. [doi: 10.1007/s40620-016-0273-x] [Medline: 26885659]
6. Arnet I, Walter PN, Hersberger KE. Polymedication Electronic Monitoring System (POEMS) - a new technology for measuring adherence. Front Pharmacol 2013;4:26 [FREE Full text] [doi: 10.3389/fphar.2013.00026] [Medline: 23493650]
7. Miguel-Cruz A, Bohórquez AF, Parra PA. What does the literature say about using electronic pillboxes for older adults? A systematic literature review. Disabil Rehabil Assist Technol 2019 Nov;14(8):776-787. [doi: 10.1080/17483107.2018.1508514] [Medline: 30451543]
8. Williams A, Low JK, Manias E, Dooley M, Crawford K. Trials and tribulations with electronic medication adherence monitoring in kidney transplantation. Res Social Adm Pharm 2016;12(5):794-800. [doi: 10.1016/j.sapharm.2015.10.010] [Medline: 26616159]
9. Levin JB, Sams J, Tatsuoka C, Cassidy KA, Sajatovic M. Use of automated medication adherence monitoring in bipolar disorder research: pitfalls, pragmatics, and possibilities. Ther Adv Psychopharmacol 2015 Apr;5(2):76-87 [FREE Full text] [doi: 10.1177/2045125314566807] [Medline: 26240747]
10. Davis A, Sarsembayeva L, Gulyaev V, Primbetova S, Terlikbayeva A, Mergenova G, et al. If you build it, will they use it? Preferences for antiretroviral therapy (art) adherence monitoring among people who inject drugs (PWID) in Kazakhstan. AIDS Behav 2019 Dec;23(12):3294-3305 [FREE Full text] [doi: 10.1007/s10461-019-02421-y] [Medline: 30741397]
11. Musinguzi N, Muganzi CD, Boum Y, Ronald A, Marzinke MA, Hendrix CW, Partners PrEP Ancillary Adherence Study Team. Comparison of subjective and objective adherence measures for preexposure prophylaxis against HIV infection among serodiscordant couples in East Africa. AIDS 2016 Apr 24;30(7):1121-1129. [doi: 10.1097/QAD.0000000000001024] [Medline: 26785125]
12. Vallejos X, Wu C. Digital medicine: innovative drug-device combination as new measure of medication adherence. J Pharm Technol 2017 Aug 28;33(4):137-139 [FREE Full text] [doi: 10.1177/8755122517704212] [Medline: 34860983]
13. Wang R, Sitova Z, Jia X, He X, Abramson T, Gasti P, et al. Automatic identification of solid-phase medication intake using wireless wearable accelerometers. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:4168-4171. [doi: 10.1109/EMBC.2014.6944542] [Medline: 25570910]
14. Lyimo RA, van den Boogaard J, Msoka E, Hospers HJ, van der Ven A, Mushi D, et al. Measuring adherence to antiretroviral therapy in northern Tanzania: feasibility and acceptability of the Medication Event Monitoring System. BMC Public Health 2011 Feb 09;11:92 [FREE Full text] [doi: 10.1186/1471-2458-11-92] [Medline: 21306643]
15. Chai PR, Castillo-Mancilla J, Buffkin E, Darling C, Rosen RK, Horvath KJ, et al. Utilizing an ingestible biosensor to assess real-time medication adherence. J Med Toxicol 2015 Dec;11(4):439-444. [doi: 10.1007/s13181-015-0494-8] [Medline: 26245878]
16. DiStefano MJ, Schmidt H. mHealth for tuberculosis treatment adherence: a framework to guide ethical planning, implementation, and evaluation. Glob Health Sci Pract 2016 Jun 20;4(2):211-221 [FREE Full text] [doi: 10.9745/GHSP-D-16-00018] [Medline: 27353615]
17. Manyazewal T, Woldeamanuel Y, Holland DP, Fekadu A, Blumberg HM, Marconi VC. Electronic pillbox-enabled self-administered therapy versus standard directly observed therapy for tuberculosis medication adherence and treatment outcomes in Ethiopia (SELFTB): protocol for a multicenter randomized controlled trial. Trials 2020 May 05;21(1):383 [FREE Full text] [doi: 10.1186/s13063-020-04324-z] [Medline: 32370774]
18. Cross A, Gupta N, Liu B, Nair V, Kumar A, Kuttan R, et al. 99DOTS: a low-cost approach to monitoring and improving medication adherence. In: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development. USA: Association for Computing Machinery; 2019 Presented at: ICTD '19: Tenth International Conference on Information and Communication Technologies and Development; January 4 - 7, 2019; Ahmedabad India p. 1-12. [doi: 10.1145/3287098.3287102]
19. Liu X, Blaschke T, Thomas B, De Geest S, Jiang S, Gao Y, et al. Usability of a medication event reminder monitor system (MERM) by providers and patients to improve adherence in the management of tuberculosis. Int J Environ Res Public Health 2017 Sep 25;14(10) [FREE Full text] [doi: 10.3390/ijerph14101115] [Medline: 28946683]
20. Silva DS, Snyder J. The ethics of new technologies to track drug adherence. CMAJ 2018 Oct 09;190(40):1209-1210 [FREE Full text] [doi: 10.1503/cmaj.171514] [Medline: 30998499]

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;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
22. Kronish IM, Thorpe CT, Voils CI. Measuring the multiple domains of medication nonadherence: findings from a Delphi survey of adherence experts. Transl Behav Med 2021 Feb 11;11(1):104-113 [FREE Full text] [doi: 10.1093/tbm/ibz133] [Medline: 31580451]
23. Faisal S, Ivo J, Patel T. A review of features and characteristics of smart medication adherence products. Can Pharm J (Ott) 2021;154(5):312-323 [FREE Full text] [doi: 10.1177/17151635211034198] [Medline: 34484481]
24. Basu S, Garg S, Sharma N, Singh MM. Improving the assessment of medication adherence: challenges and considerations with a focus on low-resource settings. Ci Ji Yi Xue Za Zhi 2019;31(2):73-80 [FREE Full text] [doi: 10.4103/tcmj.tcmj_177_18] [Medline: 31007485]
25. Aldeer M, Javanmard M, Martin R. A review of medication adherence monitoring technologies. Appl Syst Innovat 2018 May 06;1(2):14. [doi: 10.3390/asi1020014]
26. Nsengiyumva NP, Mappin-Kasirer B, Oxlade O, Bastos M, Trajman A, Falzon D, et al. Evaluating the potential costs and impact of digital health technologies for tuberculosis treatment support. Eur Respir J 2018 Nov;52(5) [FREE Full text] [doi: 10.1183/13993003.01363-2018] [Medline: 30166325]
27. Campbell JI, Haberer JE. Cell phone-based and adherence device technologies for HIV care and treatment in resource-limited settings: recent advances. Curr HIV/AIDS Rep 2015;12(4):523-531. [doi: 10.1007/s11904-015-0282-8] [Medline: 26439917]
28. Haberer JE, Musinguzi N, Tsai AC, Boum Y, Bwana BM, Muzoora C, et al. Real-time electronic adherence monitoring plus follow-up improves adherence compared with standard electronic adherence monitoring. AIDS 2017 Jan 02;31(1):169-171 [FREE Full text] [doi: 10.1097/QAD.0000000000001310] [Medline: 27835622]
29. Mehta SJ, Asch DA, Troxel AB, Lim R, Lewey J, Wang W, et al. Comparison of pharmacy claims and electronic pill bottles for measurement of medication adherence among myocardial infarction patients. Med Care 2019 Feb;57(2):9-14. [doi: 10.1097/MLR.0000000000000950] [Medline: 30045159]
30. Garrison LE, Haberer JE. Technological methods to measure adherence to antiretroviral therapy and preexposure prophylaxis. Curr Opin HIV AIDS 2017 Sep;12(5):467-474. [doi: 10.1097/COH.0000000000000393] [Medline: 28590335]
31. Zijp TR, Touw DJ, van Boven JF. User acceptability and technical robustness evaluation of a novel smart pill bottle prototype designed to support medication adherence. Patient Prefer Adher 2020;14:625-634 [FREE Full text] [doi: 10.2147/PPA.S240443] [Medline: 32256053]
32. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: a qualitative study. Int J Med Inform 2019 Nov;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]
33. Robiner WN, Flaherty N, Fossum TA, Nevins TE. Desirability and feasibility of wireless electronic monitoring of medications in clinical trials. Transl Behav Med 2015 Sep;5(3):285-293 [FREE Full text] [doi: 10.1007/s13142-015-0316-1] [Medline: 26327934]
34. Van Diest AM, Ramsey R, Aylward B, Kroner JW, Sullivan SM, Nause K, et al. Adherence to biobehavioral recommendations in pediatric migraine as measured by electronic monitoring: the adherence in migraine (AIM) study. Headache 2016 Jul;56(7):1137-1146 [FREE Full text] [doi: 10.1111/head.12836] [Medline: 27167502]
35. Park S, Sentissi I, Gil SJ, Park W, Oh B, Son AR, et al. Medication event monitoring system for infectious tuberculosis treatment in morocco: a retrospective cohort study. Int J Environ Res Public Health 2019 Jan 31;16(3) [FREE Full text] [doi: 10.3390/ijerph16030412] [Medline: 30709029]
36. De Bleser L, Vincke B, Dobbels F, Happ MB, Maes B, Vanhaecke J, et al. A new electronic monitoring device to measure medication adherence: usability of the Helping Hand™. Sensors (Basel) 2010;10(3):1535-1552 [FREE Full text] [doi: 10.3390/s100301535] [Medline: 22294885]
37. Bell KM, Haberer JE. Actionable adherence monitoring: technological methods to monitor and support adherence to antiretroviral therapy. Curr HIV/AIDS Rep 2018 Oct;15(5):388-396 [FREE Full text] [doi: 10.1007/s11904-018-0413-0] [Medline: 30232578]
38. Haberer JE, Kahane J, Kigozi I, Emenyonu N, Hunt P, Martin J, et al. Real-time adherence monitoring for HIV antiretroviral therapy. AIDS Behav 2010 Dec;14(6):1340-1346 [FREE Full text] [doi: 10.1007/s10461-010-9799-4] [Medline: 20809380]
39. Musiimenta A, Atukunda EC, Tumuhimbise W, Pisarski EE, Tam M, Wyatt MA, et al. Acceptability and feasibility of real-time antiretroviral therapy adherence interventions in rural Uganda: mixed-method pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 May 17;6(5):e122 [FREE Full text] [doi: 10.2196/mhealth.9031] [Medline: 29773527]
40. Sumari-de Boer M, Pima FM, Ngowi KM, Chelangwa GM, Mtesha BA, Minja LM, et al. Implementation and effectiveness of evriMED with short messages service (SMS) reminders and tailored feedback compared to standard care on adherence to treatment among tuberculosis patients in Kilimanjaro, Tanzania: proposal for a cluster randomized controlled trial. Trials 2019 Jul 12;20(1):426 [FREE Full text] [doi: 10.1186/s13063-019-3483-4] [Medline: 31300028]
41. Ingerski LM, Loew M, Porter JS, Su Y, Zhang H, Hankins JS, et al. Use of wise device technology to measure adherence to hydroxyurea therapy in youth with sickle cell disease. J Pediatr Hematol Oncol 2021 Jan;43(1):19-25. [doi: 10.1097/MPH.0000000000001997] [Medline: 33235145]

42. Theng Y, Fernando O, Deshan C, Goh L, Hong J, Sen A, et al. CuePBox: an integrated physical and virtual pillbox for patient care. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 27 - May 2, 2013; Paris France p. 433-438. [doi: 10.1145/2468356.2468433]
43. McGillicuddy JW, Weiland AK, Frenzel RM, Mueller M, Brunner-Jackson BM, Taber DJ, et al. Patient attitudes toward mobile phone-based health monitoring: questionnaire study among kidney transplant recipients. J Med Internet Res 2013;15(1):e6 [FREE Full text] [doi: 10.2196/jmir.2284] [Medline: 23305649]
44. Aldeer M, Martin R. Medication adherence monitoring using modern technology. In: Proceedings of the IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON). 2017 Presented at: IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON); Oct. 19-21, 2017; New York, NY, USA. [doi: 10.1109/uemcon.2017.8249101]
45. Jung H, Seong SJ, Choi J, Cho J, Park S, Kim C, et al. The efficacy and stability of an information and communication technology-based centralized monitoring system of adherence to immunosuppressive medication in kidney transplant recipients: study protocol for a randomized controlled trial. Trials 2017 Oct 16;18(1):480 [FREE Full text] [doi: 10.1186/s13063-017-2221-z] [Medline: 29037222]
46. Bachman DM, Gifford AL, Keyi X, Li Z, Feng C, Brooks M, et al. Feasibility and acceptability of a real-time adherence device among HIV-positive IDU patients in China. AIDS Res Treat 2013;2013:957862 [FREE Full text] [doi: 10.1155/2013/957862] [Medline: 23956851]
47. Wang N, Shewade HD, Thekkur P, Huang F, Yuan Y, Wang X, et al. Electronic medication monitor for people with tuberculosis: implementation experience from thirty counties in China. PLoS One 2020;15(4):e0232337 [FREE Full text] [doi: 10.1371/journal.pone.0232337] [Medline: 32348351]
48. Drabarek D, Anh NT, Nhung NV, Hoa NB, Fox GJ, Bernays S. Implementation of medication event reminder monitors among patients diagnosed with drug susceptible tuberculosis in rural Viet Nam: a qualitative study. PLoS One 2019;14(7):e0219891 [FREE Full text] [doi: 10.1371/journal.pone.0219891] [Medline: 31329610]
49. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
50. Campbell JI, Musiimenta A, Burns B, Natukunda S, Musinguzi N, Haberer JE, et al. The importance of how research participants think they are perceived: results from an electronic monitoring study of antiretroviral therapy in Uganda. AIDS Care 2019 Jun;31(6):761-766 [FREE Full text] [doi: 10.1080/09540121.2018.1556381] [Medline: 30606037]
51. Siu HY, Mangin D, Howard M, Price D, Chan D. Developing and testing an electronic medication administration monitoring device for community dwelling seniors: a feasibility study. Pilot Feasibility Stud 2017;3:5 [FREE Full text] [doi: 10.1186/s40814-016-0118-3] [Medline: 28168041]
52. Hoshi K, Kawakami J, Aoki S, Hamada K, Sato K. Compliance monitoring system using screen printing technology based on conductive ink. Technol Health Care 2012;20(6):511-520. [doi: 10.3233/THC-2012-00701] [Medline: 23187016]
53. Thakkar D, Piparva KG, Lakkad SG. A pilot project: 99DOTS information communication technology-based approach for tuberculosis treatment in Rajkot district. Lung India 2019;36(2):108-111 [FREE Full text] [doi: 10.4103/lungindia.lungindia_86_18] [Medline: 30829243]
54. Schukat M, Rudroju B. DigiSpenser--a GSM-based drug management and compliance monitoring system. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:5311-5314. [doi: 10.1109/IEMBS.2011.6091314] [Medline: 22255537]
55. Morak J, Schwarz M, Hayn D, Schreier G. Feasibility of mHealth and Near Field Communication technology based medication adherence monitoring. Conf Proc IEEE Eng Med Biol Soc 2012;2012:272-275. [doi: 10.1109/EMBC.2012.6345922] [Medline: 23365883]
56. van Onzenoort HA, Neef C, Verberk WW, van Iperen HP, de Leeuw PW, van der Kuy PM. Determining the feasibility of objective adherence measurement with blister packaging smart technology. Am J Health Syst Pharm 2012 May 15;69(10):872-879. [doi: 10.2146/ajhp100592] [Medline: 22555083]
57. Bruxvoort K, Festo C, Cairns M, Kalolella A, Mayaya F, Kachur SP, et al. Measuring patient adherence to malaria treatment: a comparison of results from self-report and a customised electronic monitoring device. PLoS One 2015;10(7):e0134275 [FREE Full text] [doi: 10.1371/journal.pone.0134275] [Medline: 26214848]
58. Hafezi H, Robertson TL, Moon GD, Au-Yeung K, Zdeblick MJ, Savage GM. An ingestible sensor for measuring medication adherence. IEEE Trans Biomed Eng 2015 Jan;62(1):99-109. [doi: 10.1109/tbme.2014.2341272]
59. Martani A, Geneviève LD, Poppe C, Casonato C, Wangmo T. Digital pills: a scoping review of the empirical literature and analysis of the ethical aspects. BMC Med Ethics 2020 Jan 08;21(1):3 [FREE Full text] [doi: 10.1186/s12910-019-0443-1] [Medline: 31914995]
60. Chai PR, Rosen RK, Boyer EW. Ingestible biosensors for real-time medical adherence monitoring: MyTMed. Proc Annu Hawaii Int Conf Syst Sci 2016 Jan;2016:3416-3423 [FREE Full text] [doi: 10.1109/HICSS.2016.426] [Medline: 27182206]
61. Eisenberger U, Wüthrich RP, Bock A, Ambühl P, Steiger J, Intondi A, et al. Medication adherence assessment: high accuracy of the new Ingestible Sensor System in kidney transplants. Transplantation 2013 Aug 15;96(3):245-250 [FREE Full text] [doi: 10.1097/TP.0b013e31829b7571] [Medline: 23823651]

62. Kamal S, Rosen MI, Lazar C, Siqueiros L, Wang Y, Daar ES, et al. Perceptions of people living with HIV and HIV healthcare providers on real-time measuring and monitoring of antiretroviral adherence using ingestible sensors: a qualitative study. AIDS Res Treat 2020;2020:1098109 [FREE Full text] [doi: 10.1155/2020/1098109] [Medline: 32566285]
63. Browne SH, Umlauf A, Tucker AJ, Low J, Moser K, Garcia JG, et al. Wirelessly observed therapy compared to directly observed therapy to confirm and support tuberculosis treatment adherence: a randomized controlled trial. PLoS Med 2019 Oct;16(10):e1002891 [FREE Full text] [doi: 10.1371/journal.pmed.1002891] [Medline: 31584944]
64. Flores GP, Peace B, Carnes TC, Baumgartner SL, Buffkin DE, Euliano NR, et al. Performance, reliability, usability, and safety of the ID-Cap system for ingestion event monitoring in healthy volunteers: a pilot study. Innov Clin Neurosci 2016;13(9-10):12-19 [FREE Full text] [Medline: 27974996]
65. Chai PR, Carreiro S, Innes BJ, Rosen RK, O'Cleirigh C, Mayer KH, et al. Digital pills to measure opioid ingestion patterns in emergency department patients with acute fracture pain: a pilot study. J Med Internet Res 2017 Dec 13;19(1):e19 [FREE Full text] [doi: 10.2196/jmir.7050] [Medline: 28087496]
66. Rahman T. Should trackable pill technologies be used to facilitate adherence among patients without insight? AMA J Ethics 2019 Apr 01;21(4):332-336 [FREE Full text] [doi: 10.1001/amajethics.2019.332] [Medline: 31012420]
67. McCall C, Maynes B, Zou C, Zhang N. RMAIS: RFID-based medication adherence intelligence system. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; Aug. 31 - Sept. 4, 2010; Buenos Aires, Argentina. [doi: 10.1109/iembs.2010.5627529]
68. Shtrichman R, Conrad S, Schimo K, Shachar R, Machluf E, Mindal E, et al. Use of a digital medication management system for effective assessment and enhancement of patient adherence to therapy (ReX): feasibility study. JMIR Hum Factors 2018 Nov 26;5(4):e10128 [FREE Full text] [doi: 10.2196/10128] [Medline: 30478020]
69. Roh H, Shin S, Han J, Lim S. A deep learning-based medication behavior monitoring system. Math Biosci Eng 2021 Jan 28;18(2):1513-1528 [FREE Full text] [doi: 10.3934/mbe.2021078] [Medline: 33757196]
70. Bilodeau G, Ammouri S. Monitoring of medication intake using a camera system. J Med Syst 2011 Jun;35(3):377-389. [doi: 10.1007/s10916-009-9374-6] [Medline: 20703552]
71. Steinkamp JM, Goldblatt N, Borodovsky JT, LaVertu A, Kronish IM, Marsch LA, et al. Technological interventions for medication adherence in adult mental health and substance use disorders: a systematic review. JMIR Ment Health 2019 Mar 12;6(3):e12493 [FREE Full text] [doi: 10.2196/12493] [Medline: 30860493]
72. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e18 [FREE Full text] [doi: 10.2196/mhealth.7030] [Medline: 28223265]
73. Shields MD, ALQahtani F, Rivey MP, McElnay JC. Mobile direct observation of therapy (MDOT) - A rapid systematic review and pilot study in children with asthma. PLoS One 2018;13(2):e0190031 [FREE Full text] [doi: 10.1371/journal.pone.0190031] [Medline: 29401500]
74. Sekandi JN, Buregyeya E, Zalwango S, Dobbin KK, Atuyambe L, Nakkonde D, et al. Video directly observed therapy for supporting and monitoring adherence to tuberculosis treatment in Uganda: a pilot cohort study. ERJ Open Res 2020 Jan;6(1) [FREE Full text] [doi: 10.1183/23120541.00175-2019] [Medline: 32280670]
75. Fozoonmayeh D, Le HV, Wittfoth E, Geng C, Ha N, Wang J, et al. A scalable smartwatch-based medication intake detection system using distributed machine learning. J Med Syst 2020 Feb 28;44(4):76. [doi: 10.1007/s10916-019-1518-8] [Medline: 32112271]
76. Lim TH, Abdullah AH. Medication adherence using non-intrusive wearable sensors. ICST Transactions on Ambient Systems 2017 Dec 19;4(16):153484. [doi: 10.4108/eai.19-12-2017.153484]
77. Hettema JE, Hosseinbor S, Ingersoll KS. Feasibility and reliability of interactive voice response assessment of HIV medication adherence: research and clinical implications. HIV Clin Trials 2012;13(5):271-277 [FREE Full text] [doi: 10.1310/hct1305-271] [Medline: 23134627]
78. Mulvaney SA, Ho Y, Cala CM, Chen Q, Nian H, Patterson BL, et al. Assessing adolescent asthma symptoms and adherence using mobile phones. J Med Internet Res 2013;15(7):e141 [FREE Full text] [doi: 10.2196/jmir.2413] [Medline: 23864345]
79. Thomas BE, Kumar JV, Chiranjeevi M, Shah D, Khandewale A, Thiruvengadam K, et al. Evaluation of the accuracy of 99DOTS, a novel cellphone-based strategy for monitoring adherence to tuberculosis medications: comparison of digital adherence data with urine isoniazid testing. Clin Infect Dis 2020 Mar 28. [doi: 10.1093/cid/ciaa333] [Medline: 32221550]
80. Leu C, Mabragaña M, Giguere R, Dolezal C, Carballo-Diéguez A, McGowan I. Use of a novel technology to track adherence to product use in a microbicide trial of short duration (MTN-007). AIDS Behav 2013 Nov;17(9):3101-3107 [FREE Full text] [doi: 10.1007/s10461-013-0549-2] [Medline: 23842718]
81. Khurana L, Durand EM, Gary ST, Otero AV, Gerzon MC, Beck J, et al. Subjects with osteoarthritis can easily use a handheld touch screen electronic device to report medication use: qualitative results from a usability study. Patient Prefer Adher 2016;10:2171-2179 [FREE Full text] [doi: 10.2147/PPA.S94247] [Medline: 27822018]
82. Fosser SM, Mahmoud N, Habib B, Weir DL, Chan F, El Halabieh R, et al. Smart about medications (SAM): a digital solution to enhance medication management following hospital discharge. JAMIA Open 2021 Apr;4(2):ooab037 [FREE Full text] [doi: 10.1093/jamiaopen/ooab037] [Medline: 34159299]

83. Melilli E, Cestone G, Revuelta I, Meneghini M, Lladó L, Montero N, et al. Adoption of a novel smart mobile-health application technology to track chronic immunosuppression adherence in solid organ transplantation: results of a prospective, observational, multicentre, pilot study. Clin Transplant 2021 May;35(5):e14278. [doi: 10.1111/ctr.14278] [Medline: 33682207]
84. Ellis RJ, Hill JH, Kerley KD, Sinha A, Ganci A, Russell CL. The feasibility of a using a smart button mobile health system to self-track medication adherence and deliver tailored short message service text message feedback. JMIR Form Res 2019 Jun 25;3(2):e13558 [FREE Full text] [doi: 10.2196/13558] [Medline: 31237568]
85. Huang R, Ren G, Hu J. Bracelet- and self-directed observational therapy for control of tuberculosis: study protocol for a cluster randomized controlled trial. Trials 2017 Jul 04;18(1):286 [FREE Full text] [doi: 10.1186/s13063-017-1996-2] [Medline: 28673323]
86. Roy PC, Abidi SR, Abidi SSR. Monitoring Activities Related to Medication Adherence in Ambient Assisted Living Environments. Stud Health Technol Inform 2017;235:28-32. [Medline: 28423749]
87. Nouvet E, Knoblauch AM, Passe I, Andriamiadanarivo A, Ravelona M, Ainanomena Ramtariharisoa F, et al. Perceptions of drones, digital adherence monitoring technologies and educational videos for tuberculosis control in remote Madagascar: a mixed-method study protocol. BMJ Open 2019 May 09;9(5):e028073 [FREE Full text] [doi: 10.1136/bmjopen-2018-028073] [Medline: 31076475]
88. Moshnyaga V, Koyanagi M, Hirayama F, Takahama A, Hashimoto K. A medication adherence monitoring system for people with dementia. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2016 Presented at: IEEE International Conference on Systems, Man, and Cybernetics (SMC); 9-12 Oct. 2016; Budapest, Hungary. [doi: 10.1109/smc.2016.7844240]
89. Hayakawa M, Uchimura Y, Omae K, Waki K, Fujita H, Ohe K. A smartphone-based medication self-management system with realtime medication monitoring. Appl Clin Inform 2013;4(1):37-52 [FREE Full text] [doi: 10.4338/ACI-2012-10-RA-0045] [Medline: 23650486]
90. Hoshi K, Kawakami J, Aoki S, Hamada K, Sato K. Real-time wireless compliance monitoring system using calendar-type pill organizer. Technol Health Care 2013;21(5):455-467. [doi: 10.3233/THC-130747] [Medline: 24029048]
91. Rantanen P, Parkkari T, Leikola S, Airaksinen M, Lyles A. An in-home advanced robotic system to manage elderly home-care patients' medications: a pilot safety and usability study. Clin Ther 2017 May;39(5):1054-1061 [FREE Full text] [doi: 10.1016/j.clinthera.2017.03.020] [Medline: 28433400]
92. Gengiah TN, Upfold M, Naidoo A, Mansoor LE, Feldblum PJ, Karim QA, et al. Monitoring microbicide gel use with real-time notification of the container's opening events: results of the CAPRISA Wisebag study. AIDS Behav 2014 May;18(5):833-840 [FREE Full text] [doi: 10.1007/s10461-014-0750-y] [Medline: 24643314]
93. Schuman-Olivier Z, Borodovsky JT, Steinkamp J, Munir Q, Butler K, Greene MA, et al. MySafeRx: a mobile technology platform integrating motivational coaching, adherence monitoring, and electronic pill dispensing for enhancing buprenorphine/naloxone adherence during opioid use disorder treatment: a pilot study. Addict Sci Clin Pract 2018 Sep 24;13(1):21 [FREE Full text] [doi: 10.1186/s13722-018-0122-4] [Medline: 30249279]
94. Spratt ES, Papa CE, Mueller M, Patel S, Killeen T, Maher E, et al. Using technology to improve adherence to hiv medications in transitional age youth: research reviewed, methods tried, lessons learned. J Gen Med (Dover) 2017;1(1) [FREE Full text] [Medline: 30345429]
95. Jung H, Jeon Y, Seong SJ, Seo JJ, Choi J, Cho J, et al. ICT-based adherence monitoring in kidney transplant recipients: a randomized controlled trial. BMC Med Inform Decis Mak 2020 Jun 10;20(1):105 [FREE Full text] [doi: 10.1186/s12911-020-01146-6] [Medline: 32522263]
96. Velligan DI, Kamil SH. Enhancing patient adherence: introducing smart pill devices. Ther Deliv 2014 Jun;5(6):611-613 [FREE Full text] [doi: 10.4155/tde.14.33] [Medline: 25090273]
97. Corden ME, Koucky EM, Brenner C, Palac HL, Soren A, Begale M, et al. MedLink: a mobile intervention to improve medication adherence and processes of care for treatment of depression in general medicine. Digit Health 2016;2:2055207616663069 [FREE Full text] [doi: 10.1177/2055207616663069] [Medline: 29942564]
98. Ermisoglu E, Bayrak C, Mendi E. Simulation of mobile treatment monitoring system. In: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2013 Presented at: ASONAM '13: Advances in Social Networks Analysis and Mining; August 25 - 28, 2013; Niagara Ontario Canada p. 978-981. [doi: 10.1145/2492517.2500318]
99. Haberer JE, Kiwanuka J, Nansera D, Wilson IB, Bangsberg DR. Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. AIDS Behav 2010 Dec;14(6):1294-1301 [FREE Full text] [doi: 10.1007/s10461-010-9720-1] [Medline: 20532605]
100. Bahrainwala L, Knoblauch AM, Andriamiadanarivo A, Diab MM, McKinney J, Small PM, et al. Drones and digital adherence monitoring for community-based tuberculosis control in remote Madagascar: a cost-effectiveness analysis. PLoS One 2020;15(7):e0235572 [FREE Full text] [doi: 10.1371/journal.pone.0235572] [Medline: 32634140]
101. Musiimenta A, Tumuhimbise W, Mugaba AT, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: mixed methods study. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119] [Medline: 31788561]

102. Haberer JE, Kiwanuka J, Nansera D, Muzoora C, Hunt PW, So J, et al. Realtime adherence monitoring of antiretroviral therapy among HIV-infected adults and children in rural Uganda. AIDS 2013 Aug 24;27(13):2166-2168 [FREE Full text] [doi: 10.1097/QAD.0b013e328363b53f] [Medline: 23751260]
103. Belknap R, Weis S, Brookens A, Au-Yeung KY, Moon G, DiCarlo L, et al. Feasibility of an ingestible sensor-based system for monitoring adherence to tuberculosis therapy. PLoS One 2013;8(1):e53373 [FREE Full text] [doi: 10.1371/journal.pone.0053373] [Medline: 23308203]
104. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):52-54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
105. The app evaluation model. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2022-02-27]

Abbreviations

DOT: directly observed therapy
EMMS: Electronic Medication Management System
HCP: health care provider
MBMS: Medication Behavior Monitoring System
MEMS: Medication Event Monitoring System
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RFID: radio frequency identification
RMAIS: radio frequency identification–based medication adherence intelligence system
VDOT: video-directly observed therapy

Edited by L Buis; submitted 23.11.21; peer-reviewed by H Ayatollahi, Z Wang, H Mehdizadeh; comments to author 28.12.21; revised version received 23.01.22; accepted 28.01.22; published 10.03.22.

Please cite as:
Mason M, Cho Y, Rayo J, Gong Y, Harris M, Jiang Y
Technologies for Medication Adherence Monitoring and Technology Assessment Criteria: Narrative Review
JMIR Mhealth Uhealth 2022;10(3):e35157
URL: https://mhealth.jmir.org/2022/3/e35157
doi:10.2196/35157
PMID:35266873

Review

Integrating Behavioral Science and Design Thinking to Develop Mobile Health Interventions: Systematic Scoping Review

Paula Voorheis[1,2], MSc; Albert Zhao[2,3], MSc; Kerry Kuluski[1,4], PhD; Quynh Pham[1,5], PhD; Ted Scott[2,6], PhD; Peter Sztur[2], BTech; Nityan Khanna[2], BTech; Mohamed Ibrahim[2], BTech; Jeremy Petch[1,2,7,8], PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[2]Centre for Data Science and Digital Health, Hamilton Health Sciences, Hamilton, ON, Canada
[3]Master of eHealth Program, McMaster University, Hamilton, ON, Canada
[4]Institute for Better Health, Trillium Health Partners, Toronto, ON, Canada
[5]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[6]Faculty of Health Sciences, School of Nursing, McMaster University, Hamilton, ON, Canada
[7]Division of Cardiology, Department of Medicine, McMaster University, Hamilton, ON, Canada
[8]Population Health Research Institute, Hamilton, ON, Canada

Corresponding Author:
Paula Voorheis, MSc
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College St 4th Floor
Toronto, ON, M5T 3M6
Canada
Phone: 1 7058163180
Email: paula.voorheis@mail.utoronto.ca

Abstract

Background: Mobile health (mHealth) interventions are increasingly being designed to facilitate health-related behavior change. Integrating insights from behavioral science and design science can help support the development of more effective mHealth interventions. Behavioral Design (BD) and Design Thinking (DT) have emerged as best practice approaches in their respective fields. Until now, little work has been done to examine how BD and DT can be integrated throughout the mHealth design process.

Objective: The aim of this scoping review was to map the evidence on how insights from BD and DT can be integrated to guide the design of mHealth interventions. The following questions were addressed: (1) what are the main characteristics of studies that integrate BD and DT during the mHealth design process? (2) what theories, models, and frameworks do design teams use during the mHealth design process? (3) what methods do design teams use to integrate BD and DT during the mHealth design process? and (4) what are key design challenges, implementation considerations, and future directions for integrating BD and DT during mHealth design?

Methods: This review followed the Joanna Briggs Institute reviewer manual and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist. Studies were identified from MEDLINE, PsycINFO, Embase, CINAHL, and JMIR by using search terms related to mHealth, BD, and DT. Included studies had to clearly describe their mHealth design process and how behavior change theories, models, frameworks, or techniques were incorporated. Two independent reviewers screened the studies for inclusion and completed the data extraction. A descriptive analysis was conducted.

Results: A total of 75 papers met the inclusion criteria. All studies were published between 2012 and 2021. Studies integrated BD and DT in notable ways, which can be referred to as “Behavioral Design Thinking.” Five steps were followed in Behavioral Design Thinking: (1) empathize with users and their behavior change needs, (2) define user and behavior change requirements, (3) ideate user-centered features and behavior change content, (4) prototype a user-centered solution that supports behavior change, and (5) test the solution against users’ needs and for its behavior change potential. The key challenges experienced during mHealth design included meaningfully engaging patient and public partners in the design process, translating evidence-based

behavior change techniques into actual mHealth features, and planning for how to integrate the mHealth intervention into existing clinical systems.

Conclusions: Best practices from BD and DT can be integrated throughout the mHealth design process to ensure that mHealth interventions are purposefully developed to effectively engage users. Although this scoping review clarified how insights from BD and DT can be integrated during mHealth design, future research is needed to identify the most effective design approaches.

(JMIR Mhealth Uhealth 2022;10(3):e35799) doi:10.2196/35799

KEYWORDS

behavior change; design thinking; digital health; health behavior; mobile application; mobile health; mobile phone; product design; scoping review; systems design; telemedicine; user-centered design

Introduction

Background

Digital health interventions are increasingly being designed to help people manage their health [1]. Many of these digital health interventions seek to facilitate behavior change and are often referred to as "digital behavior change interventions" (DBCIs) [2]. Among the wide range of DBCIs available, mobile health (mHealth) interventions have the potential to improve the reach and efficiency of health support owing to the widespread use of mobile phones [3,4]. Despite the potential impact of mHealth DBCIs, there is mixed evidence on whether they are effective at changing health behavior and improving health outcomes [3]. One concern is that patients and the public struggle to "effectively engage" with mHealth DBCIs [5,6]. Effective engagement with mHealth DBCIs has been defined as necessitating both microengagement with the mHealth interface itself (eg, logging into the app, entering data) and macroengagement with the behavior changes the mHealth intervention aims to support (eg, performing the exercises prescribed by the app) [6]. Therefore, for an mHealth DBCI to "effectively engage" users, it must be designed with engaging user-centered features to support microengagement and evidence-based behavior change techniques (BCTs) to support macroengagement.

To design effective behavior change content, there is evidence that mHealth interventions developed using the behavior change theory and BCTs are more likely to be effective than those without [7]. Behavioral scientists have been developing methods to systematically transition from diagnosing a behavioral problem to designing a behavior change intervention [8-11], which can be referred to as Behavioral Design (BD) [12]. Bondaronek and colleagues [13] provide several examples of how BD can be operationalized in publicly available physical activity apps. For instance, the app "Movesum" uses BCT 1.1 Goal Setting (Behavior) in the form of an easily adjustable step count goal [13,14]. BCT 1.1 Goal Setting (Behavior) is particularly useful when users struggle to plan for what they want to achieve or how they want to act [15]. Overall, BD can be operationalized in diverse ways but generally involves the following steps: (1) understanding the behavioral problem, (2) making a behavioral diagnosis for the target behavior using behavioral theories, models, and frameworks, (3) identifying relevant BCTs using taxonomies and classifications, (4) translating BCTs into intervention features, and (5) evaluating behavior change outcomes.

To design engaging user-centered mHealth features, there is evidence that mHealth interventions developed using person- and user-centered design processes are more likely to facilitate user engagement and improve intervention effectiveness [16,17]. Person- and user-centered design processes vary in terms of their operationalization but always put user needs at the forefront of design. Design Thinking (DT) is a common framework used in design science to guide creative user-centric designs for mHealth [18]. Design thinkers can use a range of approaches during the design process to ensure user-centeredness, such as directly involving users in app development or referring to Nielsen's 10 usability heuristics [19]. The Nielsen Norman Group provides several detailed examples of how different usability principles can be implemented in application design [20,21]. Overall, DT can be executed in different ways but generally involves the following steps: (1) empathizing with the user, (2) defining the user requirements, (3) ideating functional concepts, (4) prototyping the user-centered solution, and (5) testing the solution to see if users' needs are met. Despite knowledge about what constitutes "effective engagement" with mHealth DBCIs, a recent scoping review of DBCIs developed over the past 2 decades found that most design teams make limited use of BCTs and do not adequately describe the methods they employ to meet users' needs [22]. This scoping review concluded that DBCI practitioners have little guidance on how to integrate best practices from behavioral science and design science, and a need exists to develop guidance to support them through this process [22]. This conclusion has been reiterated across the literature, with experts agreeing that methodological guidance is required to design effectively engaging DBCIs [23].

Experts have also begun to discuss similarities and differences between approaches used by design scientists and behavioral scientists [24-26]. For example, design scientists often rely on end users to ideate content based on their stated preferences, needs, and recommendations. They iteratively build interventions by using ongoing feedback, with a focus on producing creative solutions that users will enjoy (ie, ensuring the users are microengaged). In contrast, behavioral scientists focus on producing solutions that will nudge behavior change (ie, ensuring the users are macroengaged). Behavioral scientists often rely on theory- and evidence-based linkages to understand behavioral problems and select intervention content. They rigorously test solutions against their ability to effect behavior change and not necessarily how and whether the user is engaging with them. Nonetheless, both approaches emphasize the importance of understanding and diagnosing the problem at

hand before proposing, designing, and implementing a solution. Both approaches also aim to ensure that the resulting solutions are designed purposefully to achieve user engagement. Amalgamating best practices from DT and BD may be mutually beneficial and help design teams develop more "effectively engaging" DBCIs [24-26]. There is currently a knowledge gap with respect to how best practices from DT and BD can be integrated to develop "effectively engaging" mHealth DBCIs. Specifically, little is known about how DT and BD can be blended throughout the mHealth DBCI design process to ensure that microengagement and macroengagement needs will be met.

Aims and Objectives

The aim of this scoping review was to identify and map how design teams have integrated best practices from BD and DT throughout the mHealth DBCI design process. By clarifying how BD and DT can be integrated, this review aimed to provide guidance on how mHealth DBCIs can be designed to more "effectively engage" patients and the public. This scoping review addressed the following questions: (1) what are the main characteristics of the studies that integrated BD and DT during the mHealth DBCI design process? (2) what theories, models, and frameworks did design teams use during the mHealth DBCI design process? (3) what methods did design teams use to integrate BD and DT during the mHealth DBCI design process? and (4) what are the key design challenges, implementation considerations, and future directions for the integration of BD and DT during mHealth DBCI design?

Methods

Study Design

The Joanna Briggs Institute reviewer's manual was used to guide the conduct of this scoping review [27]. The scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (see Multimedia Appendix 1) [28]. A scoping review protocol was drafted internally among key stakeholders, including mHealth software architects, mHealth design team managers, an information specialist, and several researchers with experience in the topic.

Search Strategy

The search strategy was developed by the lead author in collaboration with an information specialist. Initial searches were conducted in journals highly relevant to our research topic (eg, Cochrane, JMIR journals) to identify suitable search terms. A series of initial searches in MEDLINE were completed to analyze the text words contained in the title and abstracts of retrieved papers and index terms used to describe the papers. A final list of search terms was compiled, and a search of the databases MEDLINE, PsycINFO, EMBASE, and CINAHL was completed on May 15, 2021. A handsearch of *Journal of Medical Internet Research* was also completed on May 15, 2021, as this was recognized as a journal highly relevant to the research topic. A gray literature search and a review of paper reference lists were not conducted. This decision was made to tighten the scope of the review around the research objectives, given time and resource constraints. Search terms combined the following topics: mHealth, behavior change, and design thinking. The full search strategy can be seen in Multimedia Appendix 2.

Eligibility Criteria

Included papers had to be primary studies, where a full-text paper described the design process of an mHealth DBCI. More specifically, included papers had to describe how BD and DT practices were integrated throughout the design of an mHealth DBCI that aimed to support behavior change in patients or the public. To meet these criteria, the papers had to clearly describe their mHealth DBCI design process, addressing at least 3 of the 5 design process steps suggested in DT (empathizing, defining, ideating, prototyping, and testing) [18]. In addition, papers had to clearly describe how behavior change theories, models, frameworks, or techniques were incorporated into the mHealth design process. Studies that only used behavior change insights to evaluate the mHealth DBCI after development were excluded. Furthermore, studies that described an mHealth intervention designed to provide a psychological treatment (eg, cognitive behavioral therapy) without describing how their design process utilized behavior change theories, models, and frameworks to support "effective engagement" were also excluded. No limitations were put on the year of publication; however, only papers published in English were included. To be eligible, the intervention must have been an mHealth intervention, defined by the World Health Organization as "health care and public health practice supported by mobile devices such as mobile phones, tablets, patient monitoring devices, and other wireless devices" [4]. The intervention must also have been designed for use by patients or the public. If the intervention was designed only for use by health care professionals, it was excluded.

Evidence Selection

Studies from the database searches were handled using Covidence (Veritas Health Innovation Ltd) reference management software. Papers were deduplicated and imported for screening using Covidence. A 2-level screening was performed after duplicate removal. During level 1 screening, titles and abstracts were screened using the eligibility criteria. Publications with title or abstract not meeting the eligibility criteria were excluded. During level 2 screening, full-text papers that passed level 1 were screened. Studies that met the eligibility criteria were included for full data extraction. Consistent with PRISMA-ScR, reasons for exclusion were recorded at the full-text level [28]. Prior to the selection of sources, 2 reviewers completed a pilot screening of 50 titles and abstracts to assess the reliability of the eligibility criteria. Interrater agreement for study inclusion was calculated using percentage agreement. If agreement was lower than 80%, the eligibility criteria would be clarified and another pilot test would occur. All interrater discrepancies during level 1 and 2 evidence selection were resolved between the 2 reviewers upon discussion. The 2 reviewers screened all titles, abstracts, and full-text papers for inclusion.

Data Extraction

The reviewers extracted data from the eligible papers by using 2 data extraction forms. The first data extraction form elicited the main study characteristics, including lead author, year of publication, journal of publication, country of origin, study design, study purpose, target user population, the health issue, the target health behavior, mHealth DBCI summary, mHealth DBCI design process duration, and the members of the mHealth DBCI design team. The second data extraction form elicited details about how BD and DT were integrated throughout the design process, which included extracting types of theories, models, and frameworks used; the approaches design teams used to integrate best practices from BD with DT over the course of the mHealth DBCI design process, key challenges in the mHealth DBCI design process, key implementation considerations for mHealth DBCIs, and future considerations for the mHealth DBCI design. The data extraction forms were drafted, revised, and agreed upon by the 2 reviewers after an iterative process of implementing the extraction forms on a sample of papers.

Analysis and Presentation of Results

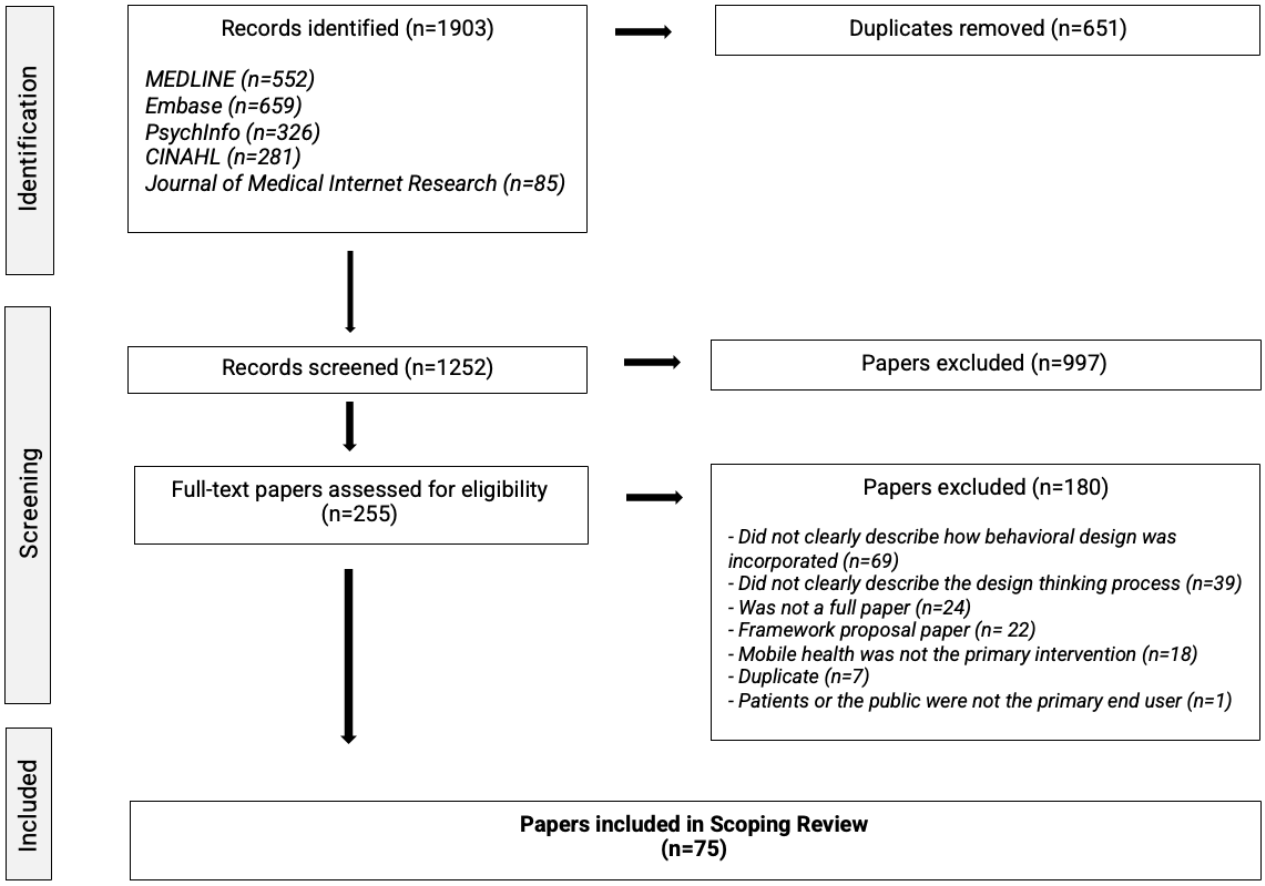
A descriptive analysis of the included papers was conducted to meet the objectives of the scoping review. Narrative descriptions, frequency calculations, and visual diagrams were utilized to communicate the results.

Results

Evidence Selection

A total of 1912 papers were identified from the searches. After 651 duplicate studies were removed, 1252 papers were screened based on their titles and abstracts, with 255 full-text papers meeting the eligibility criteria. After full-text review, 75 papers fulfilled the eligibility criteria and were included in the final review. Figure 1 shows the PRISMA-ScR flow diagram illustrating the paper selection process [28].

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram of the study selection process.



Main Characteristics of the Included Papers

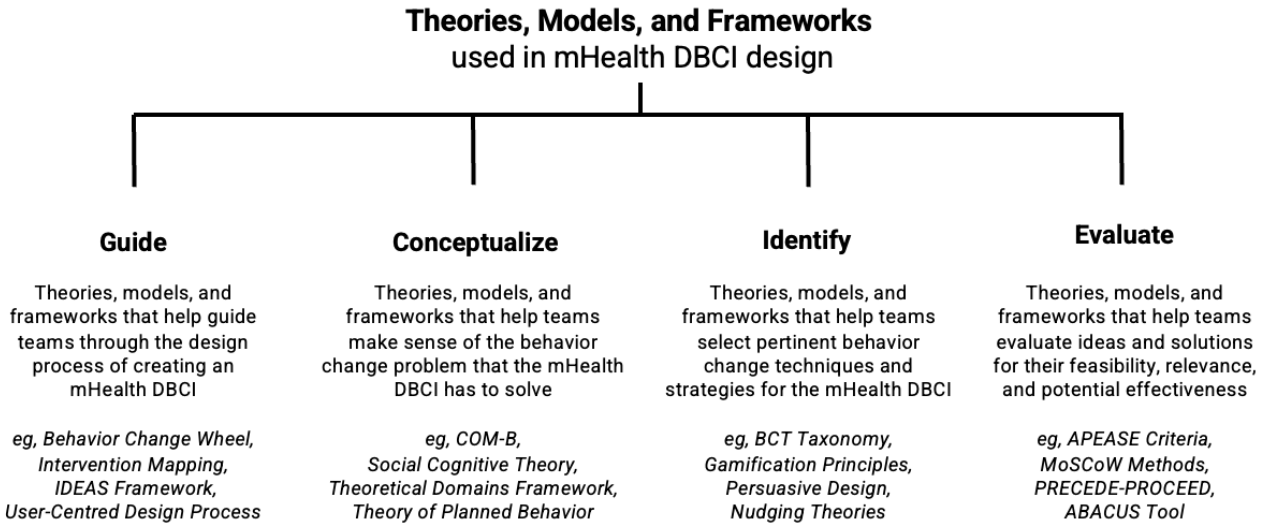
All 75 primary studies were published between 2012 and 2021, with a surge from 2018 onward (51/75, 68%). The Journal of Medical Internet Research and its sister journals accounted for almost 55% (41/75) of the papers included. The United States (14/75, 19%), the United Kingdom (13/75, 17%), Australia (10/75, 13%), and Holland (9/75, 12%) were the most common study locations. The target population of the mHealth DBCIs varied, with the most common being patients with cardiovascular issues (7/75, 9%), patients with diabetes (5/75, 7%), adults with overweight and obesity (5/75, 7%), adults who smoke (5/75, 7%), adults with poor physical activity levels (5/75, 7%), and cancer survivors (5/75, 7%). The target health behaviors addressed also varied, with the most prominent being improved physical activity (18/75, 24%), improved diet (17/75, 23%), disease self-management (12/75, 16%), preventative health

behaviors (6/75, 8%), adherence to medication (5/75, 7%), adherence to rehabilitation programming (5/75, 7%), and smoking cessation (5/75, 7%). mHealth DBCI design teams were multidisciplinary in their membership. Members included researchers, patients, caregivers, community partners, clinicians, technology developers, and experts in behavior change, health psychology, health communications, health promotion, and health informatics. Topic experts were also relied on depending on the type of intervention (eg, diabetes educators). Software engineers, computer scientists, videographers, product designers, and graphic designers were brought to assist with the development of the mHealth intervention, although it was usually unclear when they were included. The design process duration was usually not reported. Out of the 14 interventions that clearly reported design process duration, there was a large variation in timespan, ranging from less than 3 months to upwards of 4 years. Multimedia Appendix 3 provides an overview of the main study characteristics, and Multimedia Appendix 4 provides a full list of the 75 studies included.

Theories, Models, and Frameworks Used During Design

Studies used a variety of theories, models, and frameworks in their mHealth DBCI design process. Theories, models, and frameworks were most often used to (1) guide the design process itself, (2) conceptualize the behavior change problem, (3) identify relevant BCTs, and (4) evaluate ideas for their applicability, feasibility, or potential effectiveness. Figure 2 summarizes the types of theories, models, and frameworks used in the mHealth DBCI design process, and Multimedia Appendix 5 provides a detailed breakdown.

Figure 2. Theories, models, and frameworks used in mobile health digital behavior change intervention design. ABACUS: App Behavior Change Scale; APEASE: Acceptability, Practicability, Effectiveness, Affordability, Side-effects, and Equity; BCT: behavior change technique; COM-B: capability, opportunity, motivation-behavior; DBCI: digital behavior change intervention; IDEAS: Integrate, Design, Assess, and Share; mHealth: mobile health; MoSCoW: must-have, should-have, could-have, and won't-have, or will not have right now; PRECEDE: Predisposing, Reinforcing, and Enabling Constructs in Educational Diagnosis and Evaluation; PROCEED: Policy, Regulatory, and Organizational Constructs in Educational and Environmental Development.

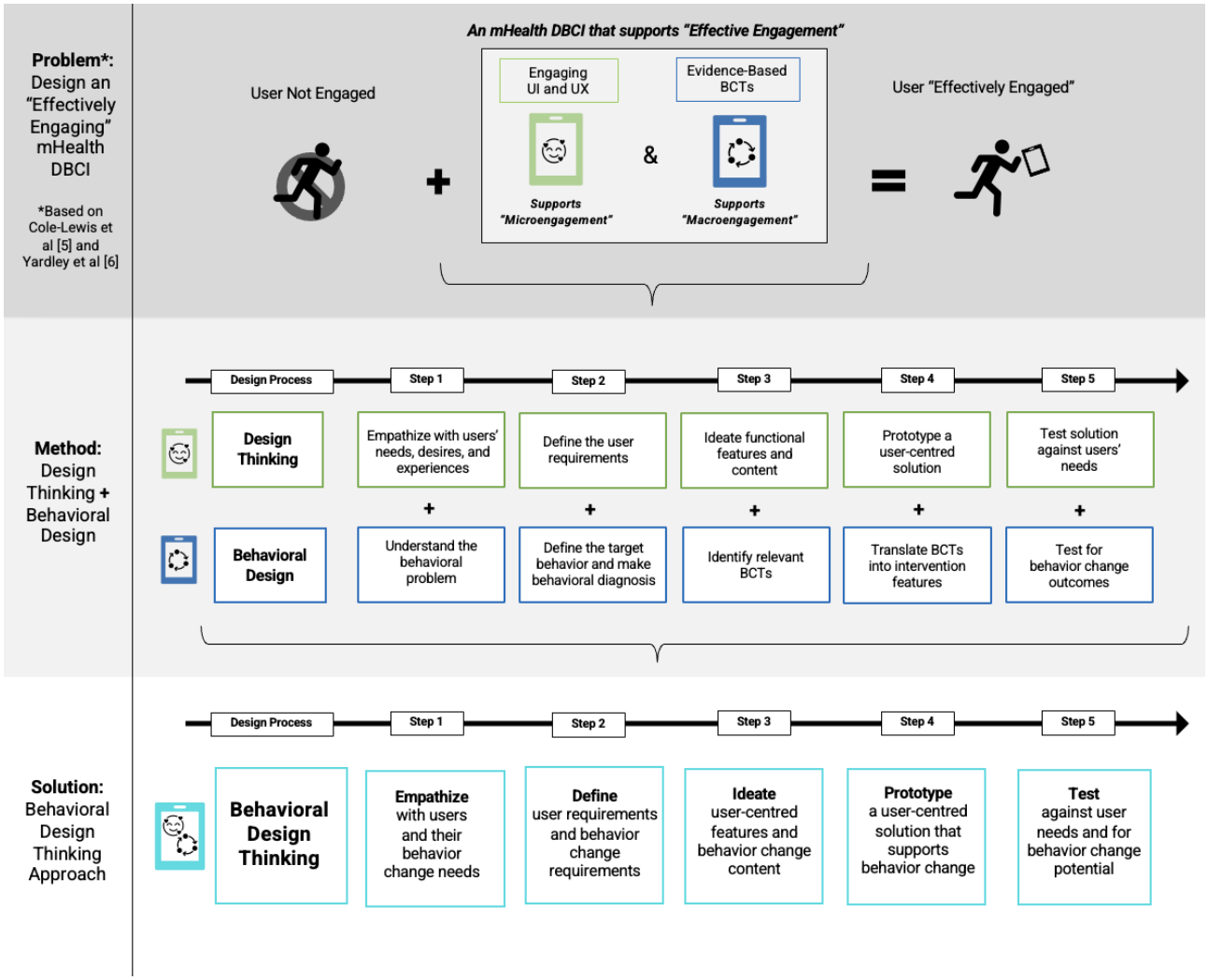


Methods Used to Integrate BD and DT During the mHealth DBCI Design Process

Regardless of the theories, models, and frameworks teams used to design their mHealth DBCIs, they integrated best practices from BD and DT in notable ways. We refer to the mixing of BD and DT throughout the mHealth DBCI design process as the "Behavioral Design Thinking Approach" (see Figure 3). The Behavioral Design Thinking Approach presents a new method of designing mHealth DBCIs, which is the result of merging together best practices from BD and DT. Multimedia Appendix 6 provides further detail on the Behavioral Design Thinking Approach along with several specific examples on how BD and DT can be integrated.

Generally, 5 steps are followed in the Behavioral Design Thinking Approach: (1) empathize with users and their behavior change needs, (2) define user and behavior change requirements, (3) ideate user-centered features and behavior change content, (4) prototype a user-centered solution that supports behavior change, and (5) test the solution against users' needs and for its behavior change potential. Across these steps, studies integrated DT and BD in different ways, often "driving" their mixing in a certain direction. The 3 ways DT and BD were mixed included (1) DT and BD approaches were equally weighted (*notation* = *DT+BD*), (2) DT drove the approach, with concepts from BD supplementing the DT approaches (*notation* = *DT→BD*), or (3) BD drove the approach, with concepts from DT supplementing the BD approaches (*notation* = *BD→DT*). Overall, studies tended to "empathize" by blending DT and BD equally (*DT+BD*), "define" by blending DT and BD equally (*DT+BD*), "ideate" by driving the process by BD (*BD→DT),* "prototype" by driving the process by DT *(DT→ BD*), and "test" by driving the process by DT *(DT→BD).* These approaches are summarized in the text below.

Figure 3. The behavioral design thinking approach [5,6]. BCT: behavior change technique; DBCI: digital behavior change intervention; mHealth: mobile health; UI: user interface; UX: user experience.



Behavioral Design Thinking Approach

Behavioral Design Thinking Step 1: Empathize With Users and Their Behavior Change Needs

Most studies described how they empathized with users while simultaneously conceptualizing the behavior change problem (66/75, 88%), blending concepts from DT and BD harmoniously *(notation = DT+BD)*. To empathize with the users who would perform a target behavior, studies analyzed users' experiences, perceptions, beliefs, needs, and preferences with their health issues, health behaviors, health interventions, and mobile app usage. To understand the behaviors that the users will perform, studies examined applicable target health behaviors, behavioral determinants, BCTs, and behavioral theories, models, and frameworks. Finally, to understand the context that the users perform the behavior in, studies assessed current practices and programs, relevant personal, social, environmental, and structural factors, and pertinent clinical, usage, and behavioral aims. Studies used primary research (eg, interviews, focus groups, surveys, creative workshops) and secondary research (eg, secondary data analyses, literature reviews, reviews of other interventions, guidelines, practices) to empathize. Studies not only involved patient and public end users during this step (48/75, 64%), but also involved health care practitioners, community partners, behavioral experts, design scientists, and technology developers. To directly involve patients and the public, studies used a variety of tools such as interview guides informed by behavior change models and visual presentations of apps to stimulate collaborative discussion.

Behavioral Design Thinking Step 2: Define User and Behavior Change Requirements

Studies tended to analyze the empathy results in order to define system requirements that would meet both users' needs and behavior change needs *(notation = DT+BD)*. Studies tended to define user-centered requirements by sorting users' stated preferences into key themes to be addressed. Studies tended to define the behavior change requirements by using the empathy results to formulate a "behavioral diagnosis," which outlines the behavioral determinants that need to be addressed. When studies organized the requirements into an amalgamated format, DT or BD usually drove the organization. For example, some studies used tables organized by the relevant behavioral determinants, whereby they would then list corresponding user quotes and resultant requirements alongside the behavioral determinants. Other studies opted to use the requirements to create holistic user personas and scenarios. Regardless of the

approach to amalgamate and make sense of the requirements, the requirement lists were often lengthy and needed to be refined. Refining the requirements involved (1) identifying the requirements that met the project scope and objectives, (2) using feasibility criteria to refine and select certain requirements, (3) consulting experts and patient and public users to prioritize the requirements, and (4) ranking the most important requirements according to their likelihood to elicit behavior change, alignment with current practices, adaptability to a digital interface, acceptability to users, and compatibility with data collection needs.

Behavioral Design Thinking Step 3: Ideate User-Centered Features and Behavior Change Content

Most studies described how they translated requirements into design content and features by using ideation methods (67/75, 89%). Generally, studies drove their ideation methods using BD *(notation = BD→DT)*. For instance, several studies relied only on BD to ideate the relevant behavior change content (19/67, 28%). These studies used researcher-led "behavioral mapping" to match relevant behavior change determinants with BCTs by using evidence-based taxonomies and tested linkages. It was often unclear, however, exactly how these studies translated the identified BCTs into content for the mobile app. Other studies attempted to completement BD approaches with DT methods during their ideation (48/67, 72%). In these cases, studies would creatively ideate design content and features that would support behavior change while also meeting users' needs and preferences. These studies relied on different techniques such as design team brainstorming, expert stakeholder panels, technology partner consultation, and end user co-design efforts. Design teams tended to brainstorm how the behavior change insights could be integrated with users' preferences, existing programming, norms in mHealth design, and clinical management needs. Stakeholder panels would discuss all the requirements and consider the most appropriate strategies (dosage, delivery, organization, and personalization). Technology partners would specify how requirements could be operationalized within an mHealth app. Patient and public partners would contribute to co-design sessions to create and reflect on ideas, content, and designs. Approximately 34% (23/75) of the studies appeared to directly involve patient and public users during this step to ensure that the ideated content met their unique needs.

Behavioral Design Thinking Step 4: Prototype a Solution That Is User-Centered and Supports Behavior Change

Many studies did not clearly describe the prototyping methods they used to translate ideated content and features into functional solutions (28/75, 36%). Studies that described their prototyping approach tended to drive their prototyping by using DT (*notation = DT→BD*). For instance, a large proportion of studies only used prototyping methods, tools, and aids traditional to DT (37/47, 79%). DT prototyping methods usually involved iterative prototyping, feedback, and refinement, often utilizing sprint, scrum, or agile methods. DT prototyping tools included wireframing, paper prototyping, application flowcharts, and use case scenarios. DT prototyping aids included tools such as Nielsen's usability heuristics [19] and the Eight Golden Rules [29]. Despite reliance on DT for prototyping, several studies supplemented DT methods with BD considerations to ensure that behavior change content would be operationalized within the prototyped solution (10/47, 21%). For example, to supplement DT prototyping methods, some studies utilized BCT codevelopment to ensure that BCTs were not lost in translation during technological development. To supplement DT prototyping tools, some studies elected to make BCT flowcharts to clarify how the BCTs would be interacted with. To supplement DT prototyping aids, some studies also used the Behavioral Intervention Technology Model [30] or the Persuasive Systems Design Framework [31], which were behaviorally informed models used to guide feature selection. Among the studies that clearly described their prototyping methods, just over half appeared to directly involve patient and public end users in their prototyping process (26/47, 55%).

Behavioral Design Thinking Step 5: Test Solution Against User Needs and for Its Behavior Change Potential

A large proportion of studies did not clearly describe how they tested their mHealth DBCI solution within the design process (18/75, 24%). Studies that described their testing methods tended to drive their testing by methods traditional to DT (*notation =DT→BD)*. For instance, a considerable number of studies relied exclusively on DT evaluation approaches such as heuristic evaluation, usability testing, expert evaluation, and pilot testing to evaluate the solutions against users' needs (33/57, 58%). Despite reliance on DT for testing, some studies supplemented these testing methods with BD considerations to evaluate the solution for its behavior change potential (24/57, 42%). For instance, in addition to traditional DT heuristic evaluation, some studies conducted a BCT evaluation to assess the final solution for the presence of known BCTs. In addition to traditional DT usability testing, some studies "tagged" BCT components within the app to follow how users engaged, accepted, and perceived the intended BCTs. Some studies also conducted posttest user interviews directed by interview guides based on constructs of a behavioral model. In addition to traditional DT expert evaluations, some studies brought on behavioral science experts to assess the extent to which the intervention content had fidelity to the intended BCTs. Experts also could assess the quality of the mHealth DBCI by using the App Behavior Change Scale [32]. If pilot tests were conducted, specific behavior change outcomes could be evaluated in addition to traditional usage metrics. Metrics included change in knowledge, change in intentions, state of behavior change, user experience with BCTs, perceived potential of BCTs, and user engagement with BCTs.

Design Challenges, Implementation Considerations, and Future Directions

In addition to describing their design process, studies also identified key implementation considerations, design challenges, and future directions for mHealth DBCI design. These have been summarized in Textbox 1. Multimedia Appendix 7 expands on these results in further detail.

Textbox 1. Key design challenges, implementation considerations, and future directions.

Design challenges

- Design process can be time and resource consuming, especially when the design approach is unclear.
- Recruiting and involving representative end users and key stakeholders can be difficult.
- Conflicting ideas can result from integrating behavioral theory, user needs, and stakeholder views.
- The translation of behavior change techniques into actual mobile health (mHealth) digital behavior change intervention (DBCI) features and content can be confusing.
- The design process is time limited and usually may not allow for comprehensive evaluation.
- Integrating the mHealth DBCI into clinical practice can be complex.
- mHealth platforms come with their own technical challenges and limiting factors.

Implementation considerations

- Evaluate potential implementation barriers and facilitators during the "testing" step.
- Facilitate early stakeholder buy-in.
- Plan for the integration of the mHealth DBCI into clinical systems.
- Use feasibility criteria throughout the entire design process.
- Use an implementation plan (marketing, dissemination, onboarding, adoption, usage, sustainability).

Future directions

- Guidance on the design process for mHealth DBCIs.
- Guidance on how to operationalize behavioral change techniques within mHealth DBCIs in a user-friendly way.
- Guidance on how to meaningfully involve users and stakeholders in the design process.
- Guidance on how to tailor mHealth DBCIs to meet behavioral, personal, and clinical needs.

Discussion

Primary Findings

This paper presents a systematic scoping review of 75 papers that described their design process for developing an mHealth DBCI. This review addressed a gap in the literature about how mHealth interventions can be designed to integrate practices from DT and BD. Although the number of mHealth DBCIs seems to be growing, the results highlight substantial heterogeneity in the methods studies used to design them. This scoping review aimed to clarify, map, and synthesize the different methods that can be used to design mHealth DBCIs. A new consolidated approach to mHealth DBCI design is presented—the Behavioral Design Thinking Approach. A large proportion of mHealth DBCIs were designed to facilitate diet- and exercise-related behavior change. Similar findings were noted in a scoping review on DBCI design over the past 2 decades [22]. Nonetheless, the results point to a lack of clarity about how mHealth DBCI content and features were ideated and operationalized. Several studies in this review struggled to translate BCTs into user-friendly mHealth features. Studies also ended up trying to fit several BCTs within the mHealth app, increasing app complexity. To simplify the user experience, several studies noted that personalization may be an important future direction for the field. Tailoring mHealth content to users' unique behavioral, clinical, and personal needs can help facilitate the delivery of features to support effective engagement [33]. Ensuring that design teams use appropriate questions and metrics to inform personalization will be essential. This further exemplifies the need for the meaningful involvement of patients and the public in mHealth DBCI design. Regarding the meaningful involvement of patients and the public in the mHealth DBCI design process, this review found that patients and the public were most often involved as participants (eg, in user interviews, surveys, testing) rather than as partners on the design team itself. Patients and the public appeared to have minimal direct input on design decision-making throughout the mHealth DBCI design process. Growing calls are being made to prioritize the unique perspectives of patients and the public during design [16,17]. Involving end users in the design process has been suggested to increase the effectiveness, relevance, and appropriateness of mHealth DBCIs [16,17]. Nonetheless, design teams appear to have little guidance on how to meaningfully engage patient and public end users in the design process itself [34-37].

Theories, Models, and Frameworks Used During Design

Regarding the theories, models, and frameworks that studies used during mHealth DBCI design, design teams appeared to face a vast array of these "tools" from different fields, created for different purposes. Although studies reported benefiting from the structure that these tools offered, a best practice approach for integrating DT and BD insights to develop "effectively engaging" mHealth DBCIs did not appear to exist. This scoping review offers a classification of the types of theories, models, and frameworks that teams can use during

mHealth DBCI design and offers a consolidated approach for guiding the design process in general.

Methods Used to Integrate BD and DT During the mHealth DBCI Design Process

During the integration of DT and BD throughout the mHealth DBCI design, key similarities and differences between the 2 approaches were observed. During the empathizing step, users' preferences often differed from what was needed to change their behavior. Balancing user-stated preferences with evidence-based behavior change strategies and mHealth platform requirements was a challenge for teams. During the defining step, it was clear that blending DT and BD perspectives resulted in lengthy and complex requirements. Involving users and topic experts to refine these requirements appeared to be a helpful strategy. During ideation, BD relied on evidence-based linkages to ideate content, whereas DT relied on iterative brainstorming and collaborative creation. Studies reconciled these different approaches by starting with BD to develop a list of relevant BCTs and then using DT to creatively integrate BCTs and other requirements into the mHealth DBCI. The prototyping step was the least well described among studies but usually relied on approaches traditional to DT. Studies that brought on their technology partner only at the prototyping stage often had to add additional methods to ensure that BCTs were operationalized as imagined. The testing step was also dominated by DT, as it was challenging to meaningfully test for behavior change outcomes within the time constraints imposed by the design process. Nonetheless, several measures were used by studies to assess the solution's potential to support behavior change.

Design Challenges, Implementation Considerations, and Future Directions

Finally, regarding key implementation considerations, design challenges, and future directions, several cross-cutting themes were identified. In addition to ensuring that BD and DT insights were incorporated into the mHealth DBCI design, many teams noted that planning for implementation of the mHealth DBCI was just as important for ensuring success. Using insights from BD and DT may also be relevant in designing an implementation plan for the mHealth DBCI. Implementation planning appears to be particularly complex in this context, as many mHealth DBCIs need to be integrated with existing clinical systems and norms. Regardless of how evidence-based and user-friendly the mHealth DBCI may be, if stakeholder buy-in, system adaptation, and clinical sustainability issues are not taken into consideration, the mHealth DBCIs are likely to fail. Future research must address the lack of guidance design teams have in developing and implementing effective mHealth DBCIs. Particularly, design teams have little guidance on the "tools" (ie, theories, models, and frameworks) they can use, how to meaningfully involve patient and public end users, and how to tailor mHealth DBCI content to meet behavioral, personal, and clinical needs.

Limitations

The most significant limitation of this scoping review was that the inclusion criteria necessitated that each included paper had to be a full-text primary study that clearly described the mHealth design process. This decision was made to tighten the scope of this review and owing to limited time and resources. Although the systematic database search was not supplemented by a gray literature search or a review of reference lists, the 75 included papers offer a diverse range of insights that meet the research objectives while addressing a prominent gap in the literature. It should be noted that the original scope of this review included a database, reference list, and gray literature search to identify available design frameworks that could be used by design teams to integrate BD and DT insights during mHealth development. Presenting these results alongside the data extracted from 75 primary studies was beyond the scope of this paper. The results of this scoping review are only inclusive of studies written in English; therefore, findings may not be generalizable internationally.

Implications

The Behavioral Design Thinking Approach offers a way forward in the field of mHealth DBCI design. mHealth design teams may consider using the insights presented in the Behavioral Design Thinking Approach to inform their future work. mHealth design teams may also find it helpful to reflect on the different types of theories, models, and frameworks they can use during the design process, as well as the key challenges they may face along the way. The findings presented in this review may also be relevant to researchers in the fields of behavioral science and design science who are interested in interdisciplinary collaboration. It is reasonable to assume that breaking down silos between these 2 fields may improve the success of mHealth DBCIs. Overall, the main benefits of this research include (1) clarifying what approaches can be used to design mHealth DBCIs, (2) promoting transparency in the choices that studies must make during the mHealth DBCI design process, and (3) enabling future researchers to test what design approaches are the most effective to develop "effectively engaging" mHealth DBCIs.

Conclusion

The number of mHealth interventions designed to support behavior change is increasing. Integrating best practices from BD and DT may allow for the development of mHealth DBCIs that more effectively engage patients and the public. This paper has helped identify and conceptualize the methods that can be used to integrate BD and DT throughout the mHealth DBCI design process. Raising the standard of mHealth design methods will be essential to ensure confidence in the impact mHealth interventions can have on improving health outcomes. If more mHealth DBCIs are purposefully designed to address effective engagement, it is likely that patients, the public, and health care practitioners will be more confident in adopting mHealth. Given the predicted increase in the demand of mHealth interventions, the time is now to ensure mHealth design methods are appropriately suited to increase effective engagement.

Acknowledgments

This review was funded from the core operating budget of the Centre for Data Science and Digital Health (CREATE) at Hamilton Health Sciences. The most responsible author is the Director of CREATE. The larger organization of Hamilton Health Sciences had no role in the design, conduct, or content of the review.

Authors' Contributions

PV was responsible for leading all stages of the scoping review. KK, QP, and JP reviewed the scoping review protocol to ensure its research rigor. TS, PS, NK, and MI ensured that the scoping review objectives and protocol were appropriate and applicable to the field. PV and AZ completed the study screening and data extraction. PV was responsible for writing the scoping review draft. All authors were involved in reviewing and preparing the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 367 KB - mhealth_v10i3e35799_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOC File , 44 KB - mhealth_v10i3e35799_app2.doc]

Multimedia Appendix 3
Main characteristics of the studies.
[DOC File , 107 KB - mhealth_v10i3e35799_app3.doc]

Multimedia Appendix 4
Summary of all the studies included.
[DOC File , 205 KB - mhealth_v10i3e35799_app4.doc]

Multimedia Appendix 5
Theories, models, and frameworks used in design.
[DOC File , 676 KB - mhealth_v10i3e35799_app5.doc]

Multimedia Appendix 6
The behavioral design thinking approach.
[PDF File (Adobe PDF File), 1248 KB - mhealth_v10i3e35799_app6.pdf]

Multimedia Appendix 7
Design challenges, implementation considerations, and future directions.
[DOC File , 42 KB - mhealth_v10i3e35799_app7.doc]

References

1. 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]
2. 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]
3. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
4. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies based on the findings of the second global survey on eHealth. Global Observatory for eHealth Series, vol 3. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-05-15]

5. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding Health Behavior Technology Engagement: Pathway to Measuring Digital Behavior Change Interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
6. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
7. 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]
8. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008 Oct;13 Suppl 3:64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
9. 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]
10. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Applied Psychology 2008;57(4):660-680 [FREE Full text] [doi: 10.1111/j.1464-0597.2008.00341.x]
11. Cash PJ, Hartlev CG, Durazo CB. Behavioural design: A process for integrating behaviour change and design. Design Studies 2017 Jan;48:96-128. [doi: 10.1016/j.destud.2016.10.001]
12. Salzer S. Behavioral design 2020 and beyond: opinions, predictions, and thoughts from leaders in the field. Medium. 2020 Feb 27. URL: https://medium.com/behavior-design-hub/behavioral-design-2020-and-beyond-dc88a87f3b97 [accessed 2021-05-15]
13. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of Publicly Available Physical Activity Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
14. 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]
15. Carey RN, Connell LE, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. Behavior Change Techniques and Their Mechanisms of Action: A Synthesis of Links Described in Published Intervention Literature. Ann Behav Med 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]
16. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
17. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
18. Design thinking bootleg. Stanford d.school. 2011. URL: https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/t/5b19b2f2aa4a99e99b26b6bb/1528410876119/dschool_bootleg_deck_2018_final_sm+%282%29.pdf [accessed 2021-05-15]
19. Nielsen J. 10 usability heuristics for user interface design. Nielsen Norman Group. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2021-05-15]
20. Moran K. Design thinking: study guide. Nielsen Norman Group. 2021. URL: https://www.nngroup.com/articles/design-thinking-study-guide/ [accessed 2021-09-01]
21. Kaplan K. 10 usability heuristics applied to complex applications. Nielsen Norman Group. 2021. URL: https://www.nngroup.com/articles/usability-heuristics-complex-applications/ [accessed 2021-09-01]
22. Taj F, Klein MCA, van Halteren A. Digital Health Behavior Change Technology: Bibliometric and Scoping Review of Two Decades of Research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
23. 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]
24. Tantia P. What is the future of design and behavioral science? A conversation with Cliff Kuang. Behavioral Scientist. 2021. URL: https://behavioralscientist.org/what-is-the-future-of-design-and-behavioral-science-a-conversation-with-cliff-kuang/ [accessed 2021-05-15]
25. Choudhary S. Human centered design, behavioral design and methods of behavioral design. Medium. 2019. URL: https://medium.com/@i.shubhangich/human-centred-design-behavioural-design-and-methods-of-behavioural-design-aa2147515267 [accessed 2021-05-15]
26. Hallsworth M, Kirkman E. The future of behavioral insights demands human-centered design. Behavioral Scientist. URL: https://behavioralscientist.org/the-future-of-behavioral-insights-demands-human-centered-design/ [accessed 2021-05-15]
27. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: JBI; 2020.

28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/m18-0850]
29. Shneiderman B. Designing for fun. Interactions 2004 Sep;11(5):48-50. [doi: 10.1145/1015530.1015552]
30. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
31. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. CAIS 2009;24:pp-pp. [doi: 10.17705/1cais.02428]
32. 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]
33. Lustria MLA, 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;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
34. Shen N, Jankowicz D, Strudwick G. Patient and Family Engagement Approaches for Digital Health Initiatives: Protocol for a Case Study. JMIR Res Protoc 2021 Jul 21;10(7):e24274 [FREE Full text] [doi: 10.2196/24274] [Medline: 34287212]
35. Bjerkan J, Kane B, Uhrenfeldt L, Veie M, Fossum M. Citizen-Patient Involvement in the Development of mHealth Technology: Protocol for a Systematic Scoping Review. JMIR Res Protoc 2020 Aug 28;9(8):e16781 [FREE Full text] [doi: 10.2196/16781] [Medline: 32857061]
36. Morton E, Barnes SJ, Michalak EE. Participatory digital health research: A new paradigm for mHealth tool development. Gen Hosp Psychiatry 2020;66:67-69 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.07.005] [Medline: 32702489]
37. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]

Abbreviations

BCT: behavior change technique
BD: Behavioral Design
DBCI: digital behavior change intervention
DT: Design Thinking
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by L Buis; submitted 03.01.22; peer-reviewed by H Johnsen, L Hickman, C Sahin; comments to author 24.01.22; revised version received 03.02.22; accepted 10.02.22; published 16.03.22.

Please cite as:
Voorheis P, Zhao A, Kuluski K, Pham Q, Scott T, Sztur P, Khanna N, Ibrahim M, Petch J
Integrating Behavioral Science and Design Thinking to Develop Mobile Health Interventions: Systematic Scoping Review
JMIR Mhealth Uhealth 2022;10(3):e35799
URL: https://mhealth.jmir.org/2022/3/e35799
doi:10.2196/35799
PMID:35293871

Original Paper

Predicting Changes in Depression Severity Using the PSYCHE-D (Prediction of Severity Change-Depression) Model Involving Person-Generated Health Data: Longitudinal Case-Control Observational Study

Mariko Makhmutova[1], MSc; Raghu Kainkaryam[2], PhD; Marta Ferreira[2], MSc; Jae Min[2], PhD; Martin Jaggi[1], PhD; Ieuan Clay[2,3], PhD

[1]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2]Evidation Health Inc, San Mateo, CA, United States
[3]Digital Medicine Society, Boston, MA, United States

Corresponding Author:
Ieuan Clay, PhD
Digital Medicine Society
90 Canal Street, 4th Floor
Boston, MA, 02114
United States
Phone: 1 1733095953
Email: ieuan@dimesociety.org

Abstract

Background: In 2017, an estimated 17.3 million adults in the United States experienced at least one major depressive episode, with 35% of them not receiving any treatment. Underdiagnosis of depression has been attributed to many reasons, including stigma surrounding mental health, limited access to medical care, and barriers due to cost.

Objective: This study aimed to determine if low-burden personal health solutions, leveraging person-generated health data (PGHD), could represent a possible way to increase engagement and improve outcomes.

Methods: Here, we present the development of PSYCHE-D (Prediction of Severity Change-Depression), a predictive model developed using PGHD from more than 4000 individuals, which forecasts the long-term increase in depression severity. PSYCHE-D uses a 2-phase approach. The first phase supplements self-reports with intermediate generated labels, and the second phase predicts changing status over a 3-month period, up to 2 months in advance. The 2 phases are implemented as a single pipeline in order to eliminate data leakage and ensure results are generalizable.

Results: PSYCHE-D is composed of 2 Light Gradient Boosting Machine (LightGBM) algorithm–based classifiers that use a range of PGHD input features, including objective activity and sleep, self-reported changes in lifestyle and medication, and generated intermediate observations of depression status. The approach generalizes to previously unseen participants to detect an increase in depression severity over a 3-month interval, with a sensitivity of 55.4% and a specificity of 65.3%, nearly tripling sensitivity while maintaining specificity when compared with a random model.

Conclusions: These results demonstrate that low-burden PGHD can be the basis of accurate and timely warnings that an individual's mental health may be deteriorating. We hope this work will serve as a basis for improved engagement and treatment of individuals experiencing depression.

(JMIR Mhealth Uhealth 2022;10(3):e34148) doi:10.2196/34148

KEYWORDS

depression; machine learning; person-generated health data

Introduction

Major depressive disorder is a leading cause of disability worldwide, impacting the lives of more than 264 million people globally, according to the World Health Organization [1]. The COVID-19 pandemic has further increased the number of people experiencing depressive symptoms [2]. Despite its prevalence, depression often remains undiagnosed and untreated. In 2017, an estimated 17.3 million adults in the United States experienced at least one major depressive episode, with 35% of them not receiving any treatment [3].

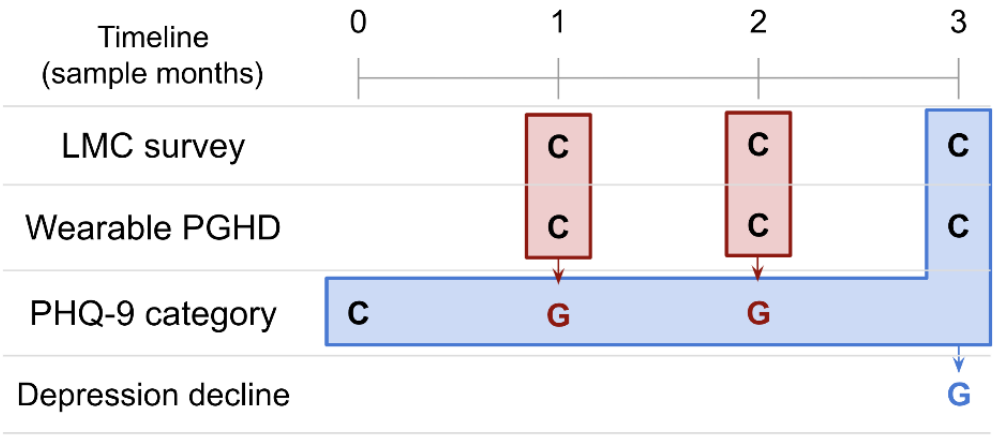
Underdiagnosis of depression has been attributed to many reasons, including stigma surrounding mental health, limited access to medical care, and barriers due to cost [4]. Undiagnosed and untreated depression has significant economic consequences, adding an economic burden of over US $200 billion annually in the United States alone [5]. Thus, it is essential to make the detection and monitoring of depression symptoms easier and more affordable.

An increasingly explored and promising way to accomplish this is through person-generated health data (PGHD) in the form of self-reports and data from consumer-grade wearable devices [6]. Multiple studies have shown that early indicators of changes in depression status can be detected from PGHD in the form of social media use [7] or physical activity patterns [8]. For example, a recent study, using consumer wearable devices to track the sleep of 368 participants, found several strong associations (Z-scores up to 6.19) between sleep features and self-reported depression [9]. Another study showed that activity features collected for 23 participants could accurately (κ=0.773) classify individuals with depression from controls, and predict changes in depression status over a 2-week period [10]. Although these studies are limited in sample size and time duration to generalize across larger populations, they demonstrate the potential of this approach versus more burdensome active assessments [11].

In this work, we present PSYCHE-D (Prediction of Severity Change-Depression), a 2-phase prediction model that uses PGHD to predict longitudinal changes in an individual's depression severity level (Figure 1). Input data include self-reported sociodemographic data and medical history, and objective behavioral data derived from consumer-grade wearables. The presented model has been developed using the largest longitudinal cohort study ever considered for depression at the time of publication [12], collecting PGHD over a 1-year period from more than 10,000 participants.

In previous work, we presented initial results [13] for the first phase of the model, and exploratory results for the second phase are also available [14]. These initial results demonstrate the feasibility of the PSYCHE-D approach, yet the stepwise development approach creates the possibility of data leakage between the phases and therefore misleading results. This work presents results from an improved pipeline that eliminates the leakage, thus ensuring generalizable results and laying the foundation for a very low–burden, consumer-facing, personalized system that could improve engagement and outcomes in people with depression.

Figure 1. A schematic overview of the PSYCHE-D (Prediction of Severity Change-Depression) model. Phase 1c uses screener survey responses (regarding sociodemographics and chronic comorbidities at baseline), self-reported lifestyle and medication changes (LMC) survey data from the month in which the Patient Health Questionnaire-9 (PHQ-9) label is generated, and data from consumer-grade wearables to categorize each individual's likely PHQ-9 category. In the second phase, this generated information is combined with the initial PHQ-9 category, screener survey responses, additional LMC self-reports, and consumer-grade wearable device person-generated health data (PGHD) to make the final prediction of whether the individual is likely to have experienced increased depression severity over the 3-month period. Red blocks represent Phase 1, and blue blocks represent Phase 2. C: collected. G: generated.



Methods

Data Collection

The data used in this work are part of the DiSCover (Digital Signals in Chronic Pain) Project (ClinicalTrials.gov identifier: NCT03421223). The DiSCover Project is a 1-year longitudinal study consisting of 10,036 individuals in the United States, who, between January 2018 and January 2020, provided data from consumer-grade wearable devices and completed surveys about their mental health and lifestyle changes quarterly and monthly,

respectively. Detailed design and baseline participant characteristics are described in the report by Lee et al [12].

The data subset used in this work comprises the following:

1. Wearable PGHD: Step and sleep data from the participants' consumer-grade wearable devices (Fitbit) worn throughout the study were collected.
2. Screener survey: Prior to the study, participant self-reported sociodemographic information, as well as comorbidities were collected.
3. Lifestyle and medication changes (LMC) survey: Every month, participants were requested to complete a brief survey reporting changes in their lifestyle and medication over the past month.
4. Patient Health Questionnaire-9 (PHQ-9) score: Every 3 months, participants were requested to complete the PHQ-9, a 9-item questionnaire that has proven to be a reliable and valid measure of depression severity [15].

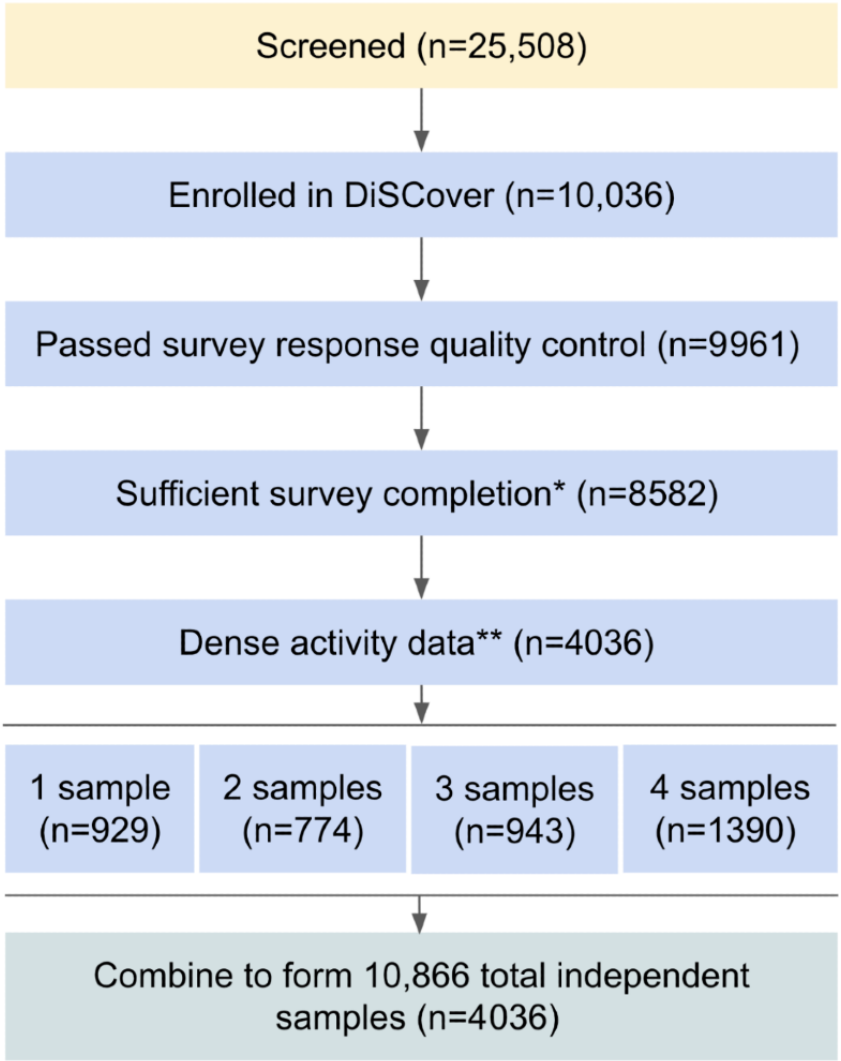
From these input sources, we defined a range of input features, both static (defined once, remain constant for all samples from a given participant throughout the study, eg, demographic features) and dynamic (varying with time for a given participant, eg, behavioral features derived from consumer-grade wearables). Feature extraction and engineering are described in Multimedia Appendix 1.

Data Processing

Data Filtering Process

Figure 2 outlines the processing of the initial data set into the samples used for developing the phase 1c model. Of the 10,036 enrolled participants, 9961 passed the survey response quality control, defined as completion of the PHQ-9 for at least two contiguous quarters, as well as the LMC survey for the same month as the second PHQ-9. Additional filtering, based on the density of available activity data in the 2 weeks matching the PHQ-9 recall period, was performed according to standards proposed in the literature [16,17]. We ultimately obtained a total of 10,866 samples from 4036 unique participants.

Figure 2. Illustration of the participant filtering process. *Completion of the Patient Health Questionnaire-9 (PHQ-9) for the current quarter, the PHQ-9 for the previous quarter, and the lifestyle and medication changes (LMC) survey for the current month. **≥10 hours daily wear time for ≥4 days per week in the 2-week interval. DiSCover: Digital Signals in Chronic Pain.



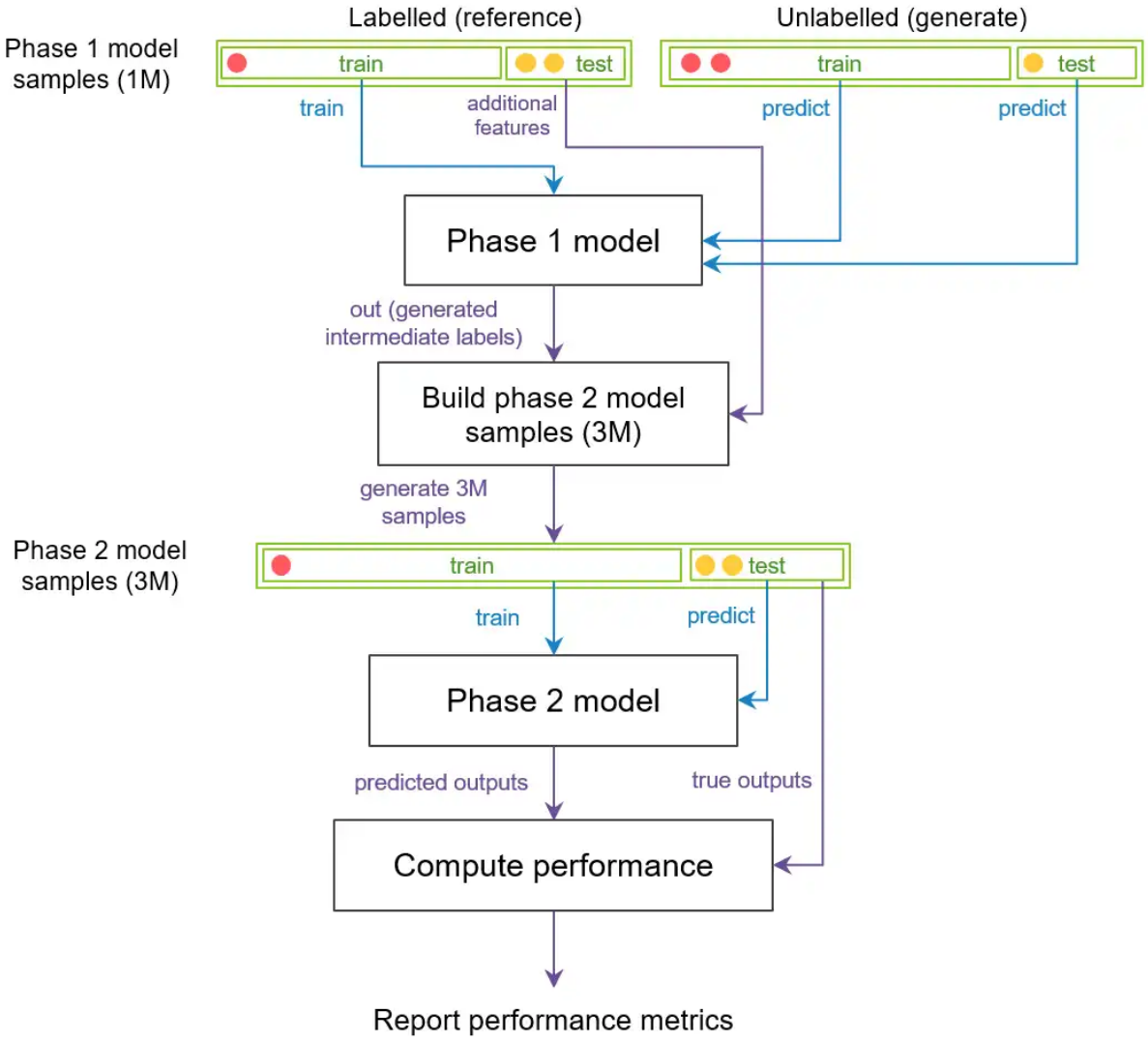
Data Set Construction

Initial data exploration showed that the evolution of PHQ-9 scores over 3-month intervals was constant throughout the study period, when grouping by demographic variables, such as sex, age, race, and geographic location. Based on this observation, we divided the data into 3-month long, nonoverlapping, independent samples. We used the notations "SM0" (sample month 0), "SM1," "SM2," and "SM3" to refer to relative time points within each sample. Each 3-month sample consisted of 1 set of screener survey responses, PHQ-9 survey responses at SM0 and SM3, LMC survey responses at SM3 (as well as SM1 and SM2, if available), and wearable PGHD for SM3 (as well as SM1 and SM2, if available). The wearable PGHD included data collected from 8 to 14 days prior to the PHQ-9 label generation date (SM1 and SM2 in phase 1, SM3 in phase 2).

Modeling

Figure 1 illustrates the overall approach, and the inputs and outputs for phase 1c and phase 2c. Figure 3 illustrates the modeling approach, which is explained in more detail below. The key design feature is that the models are combined into a single combined pipeline, and participant-based train-test partitioning is performed once at the start, in order to eliminate the possibility of data leakage. The combined pipeline is thus fitted on 1 set of participants and tested on another set of previously unseen participants.

Figure 3. Schematic representation of the PSYCHE-D (Prediction of Severity Change-Depression) combined pipeline architecture, used to estimate performance on previously unseen participants. The phase 1c model is trained on a subset of participants in the training set, and predictions for the training and test set participants are made. The phase 2c model has the same participant split for the training and test sets. Red and yellow circles represent samples from 2 different participants. All samples from the red participant are in the training set, and all samples from the yellow participant are in the test set for both phases 1c and 2c. Green blocks represent data, and black blocks represent models and data processing stages. Blue arrows represent input to classification models for training or predicting, and purple arrows represent data passage for other purposes (eg, providing true output values for testing). Note: multiple circles represent multiple samples from the same participant. This procedure is repeated over 5 random participant-based splits of the training and test data, to obtain confidence intervals for the combined pipeline performance.



Phase 1c: Categorization of the Intermediate PHQ-9

The goal of the phase 1c model was to predict participants' PHQ-9 score categories from sociodemographic, medical, and wearable PGHD. The initial version of phase 1c is described in the report by Makhmutova et al [13]. Here, we describe an improved variant that has been adapted to reduce overfitting. The Light Gradient Boosting Machine (LightGBM) algorithm with Dropouts Meet Multiple Additive Regression Trees (DART) boosting [18], an ensemble model of boosted regression trees with dropout, was chosen due to its ability to handle sparse data and the ability to tune an additional dropout parameter to reduce overfitting. Feature selection removed highly correlated features, and used recursive feature elimination [19] in order to eliminate features that had lower contributions to model performance. Model performance was primarily measured using quadratic weighted Cohen κ [20], with adjacent accuracy (ie, fraction of samples predicted at most one off from the target value), balanced accuracy, and weighted F1-scores as secondary performance metrics. We performed randomized search 5-fold cross-validation to tune the hyperparameters of our LightGBM model. We chose to perform a 5-fold cross-validation to reduce the impact of overfitting. We reported the performance metrics of the best tuned models with 95% CIs across 5 training runs (5 outer shuffle splits). Further details on hyperparameters are reported in Multimedia Appendix 1 and elsewhere [14]. Due to the very large feature space that covers a range of static and dynamic input features, we constructed the model in 3 steps. We first performed an extensive exploration accessing the best feature subsets of each type of input. We then carried out an initial optimization on input sets, which combined different types of input and considered an initial estimation of model error. Subsequently, we conducted a final tuning to obtain the best performing model. The output of phase 1c generated intermediate monthly PHQ-9 score categories for SM1, denoting sample month 1, and SM2.

Phase 2c: Prediction of Longitudinal PHQ-9 Change

In phase 2c, we predicted an increase in the PHQ-9 category using the participants' PHQ-9 scores from SM0, intermediate generated PHQ-9 categories at SM1 and SM2 as well as the generated probabilities of each PHQ-9 category for SM1 and SM2, and LMC survey responses and wearable PGHD collected over the 2 weeks prior to final PHQ-9 completion at SM3. We also used the screener survey responses as input features to control for sociodemographic factors. To compute the target variable in each sample in the phase 2c model, we observed whether there was an increase in the PHQ-9 category between SM0 and SM3.

A similar model construction procedure was used for phase 2c as for phase 1c. The feature selection procedure consisted of reducing the initial number of input features through the removal of highly correlated features and selecting the most important features using recursive feature elimination with cross-validation for the largest sets of input features, grouped by source. Then, we performed forward sequential feature selection [21], a greedy method that has been successfully used to develop digital measures in mental health studies [22], to identify the optimal features. We then again used LightGBM DART, as it has been shown to deliver high accuracy in comparable classification tasks [18], is able to handle sparse data, and generates interpretable models.

Specificity and area under the precision-recall curve (AUPRC) [23] were prioritized as performance metrics. Feature importance was assessed using a combination of the following 2 key metrics: "gain" importance and "split" importance [24]. Gain importance measures the improvement in accuracy that a feature provides, while split importance considers the number of times the feature is used in a model. Taken together, these metrics help us understand which features contribute the most to the "decisions" that the model makes.

The construction of the PSYCHE-D combined pipeline consisting of phase 1c, followed by phase 2c, is summarized in Figure 3. The diagram also illustrates the participant-based splitting approach used to ensure that we generate predictions on previously unseen participants, to evaluate the approach's generalization capabilities. Further details are presented in Multimedia Appendix 1.

Code Availability

The codes of the models in this study, along with their trained weights, are available on GitHub [25].

Data Availability

Data are made available to academic researchers on Zenodo [26].

Ethics Approval

This study received expedited review and Institutional Review Board (IRB) approval from the Western Institutional Review Board-Copernicus Group (IRB study number: 1181760; protocol number: 20172916; initial approval date: December 21, 2017).

Results

Overview

In the following section, we present the performance and informative features for the combined pipeline. Importantly, we wanted to build the model in a manner representative of how such a model might be deployed "in the real world." In such a situation, a trained model (eg, as part of an app) would need to make predictions for participants that the model is naive to, that is, people who have just downloaded the app and perhaps only filled out the baseline assessments, and did not contribute data used in the model construction. This pipeline is therefore designed to test the generalizability of the models by eliminating any data leakage, and using a participant-based validation strategy, that is, the model is tested on participants that it is completely naive to. Results for the 2 phases are presented separately.

Intermediate Classification of Depression Severity

Acquiring PGHD on a large scale requires a low-burden data collection approach; thus, participants were only asked to complete the PHQ-9 at sparse intervals, once every 3 months. Consequently, we were limited to a relatively small set of reference labels, with 2.07 labels on average per enrolled

participant over the course of 1 year. The first phase of our approach thus generated more frequent intermediate depression severity labels, which were used in combination with self-reported reference labels to reduce the sparsity of the data set by up to 3 times.

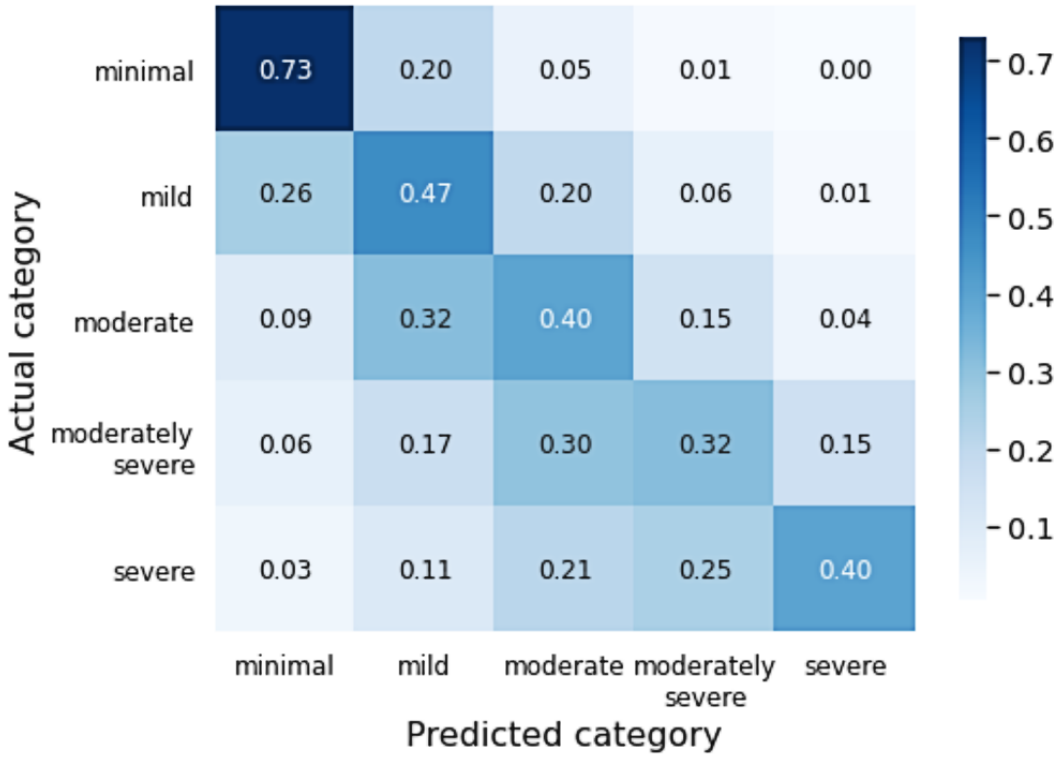
We were able to construct a multi-class classification model that determines a participant's depression severity for a given month, by assigning an individual to 1 of 5 ordinal PHQ-9 classes describing severity from minimal to severe [15]. The details and distributions of the observed classes are presented in Table S1 in Multimedia Appendix 1.

The best performing model, based on the LightGBM DART algorithm, after hyperparameter tuning, had a κ value of 0.476 (95% CI ±0.017) and an adjacent accuracy of 77.6%.

The performance of the model was not equal across all PHQ-9 severity categories. Comparing actual to predicted categories in a confusion matrix (Figure 4), we observed that performance was high for samples from individuals with either relatively low (minimal or mild) or high (moderately severe or severe) depression.

The model included features selected across all input sources, including demographics (gender, birth year, education, and BMI), life events and conditions at baseline (whether they had received financial assistance, experienced trauma or given birth, or were diagnosed with a range of chronic conditions), LMC (changes to medications or lifestyle), and sleep-related wearable PGHD (the number of hypersomnia nights, range of bedtime, and average ratio of the time spent asleep to the time spent in bed). A full list of the final features and their relative importance is included in Multimedia Appendix 2. This model was then used to generate intermediate PHQ-9 category labels for each individual for SM1 and SM2.

Figure 4. Confusion matrix showing the phase 1c model's Patient Health Questionnaire-9 (PHQ-9) score category accuracy distribution across PHQ-9 severity groups. Darker blue represents higher accuracy. Performance overall is weak, but adjacent accuracy is high, and classification performance in samples from individuals with lower (minimal to mild) and higher (moderately severe to severe) severity is relatively high, compared to the performance seen for intermediate severity samples.



Prediction of Longitudinal Change

The intermediate generation of depression severity labels means that each sample consisted of the PHQ-9 depression severity at SM0, the LMC surveys, wearable PGHD, and up to two generated labels that provide a weak estimate of depression severity (PHQ-9 category) at SM1 and SM2.

We posed our original aim as a binary problem as follows: can we predict increased depression severity? We defined increased depression severity as that when a participant changed the PHQ-9 category between SM0 and SM3. From our 10,866 samples, 2252 (20.7%) were thus labeled as positive cases.

The construction of the second phase model was optimized across possible input feature sets and LightGBM model hyperparameters. As summarized in Figure 3, we noted that with this approach, the optimization process also depended on the outputs generated by the first phase.

We used a range of metrics to assess performance, but prioritized sensitivity as the key metric, as our primary goal in this work was to correctly identify the highest proportion of individuals reporting increased depression severity. As the data set was highly imbalanced, with 21% of individuals in the data set reporting increased depression severity, we optimized for performance for both the majority and minority classes. We

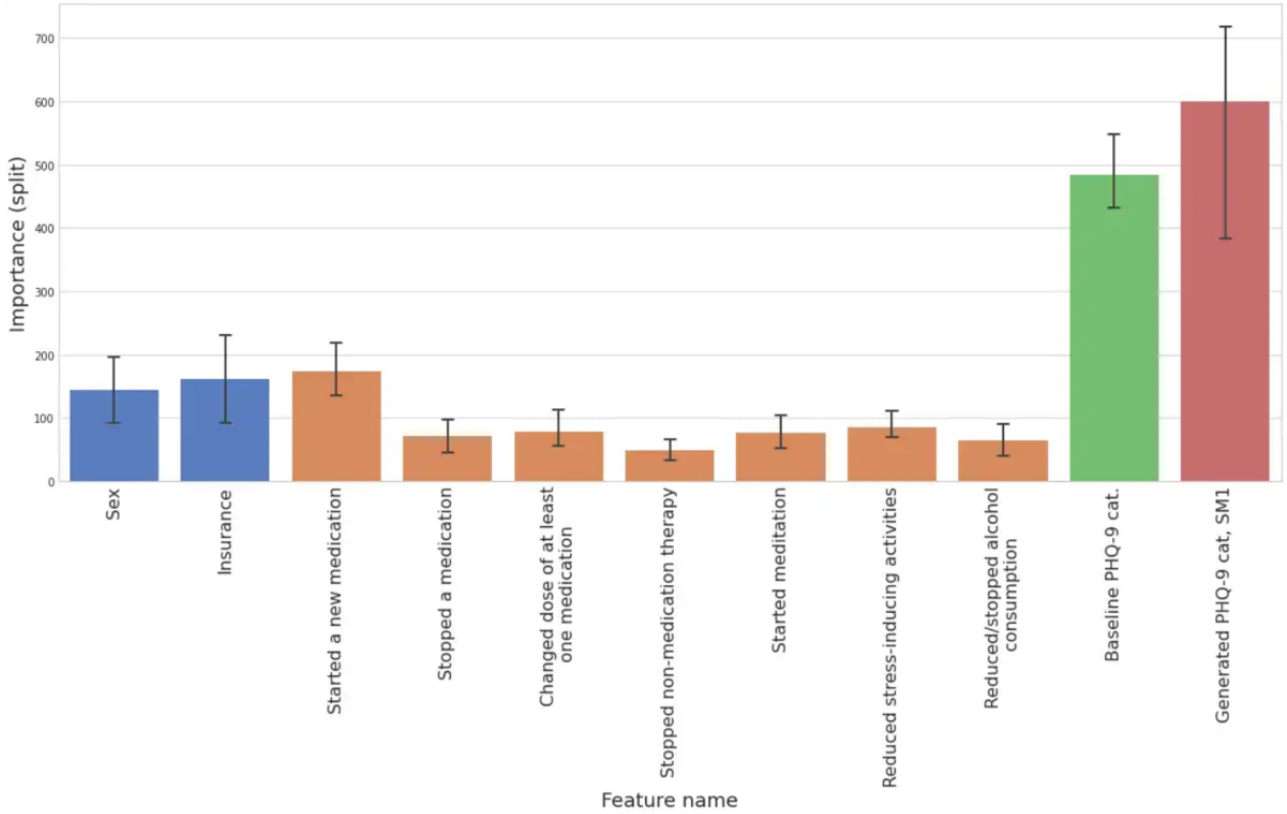
thus took into account specificity and AUPRC as secondary performance metrics, to observe the tradeoff in performance for each class.

Based on this, the best performing model selected 13 input features, with a sensitivity of 55.4% (95% CI ±0.8%), specificity of 65.3% (95% CI ±4.2%), and AUPRC of 0.31 (95% CI ±0.024). In comparison, a baseline model, which randomly assigned 20.7% positive labels, reported an AUPRC of 0.21, a sensitivity of 19.8%, and a specificity of 80.0% (averaged across 10 runs of 1000 samples).

We examined the most important features in the second phase of the combined pipeline and observed that the selected features to predict relative changes in depression were similar to the features selected to predict absolute depression in the first phase.

The most important features are presented in Figure 5, with further details in Multimedia Appendix 2. Features that were most frequently selected as strong predictors of an increase in depression severity, regardless of the cohort, were PHQ-9–related features. Specifically, the self-reported starting PHQ-9 category and the generated intermediate PHQ-9 category for SM1 were the most important features, as we can see in Figure 5. Among the static demographic and socioeconomic features, we noted that sex and having health insurance were the most important. Various self-reported LMC features were frequently selected, including medication changes (starting, stopping, and changing doses) and stress-related lifestyle changes (starting meditation and reducing stress-inducing activities), as well as reducing or stopping alcohol consumption. We observed that objective sleep features were again selected, but no specific individual wearable PGHD feature (sleep or otherwise) was sufficiently consistently selected to be included in the final model.

Figure 5. Split feature importance of the features selected in at least three of five train-test splits in the best performing phase 2c model. Colors represent the types of features (static screener features are blue, lifestyle and medication changes [LMC] features are orange, baseline Patient Health Questionnaire-9 [PHQ-9] features are green, and generated PHQ-9 features are red). The 95% CIs for the split feature importance are also visualized. SM: sample month.



Discussion

Principal Findings

PGHD represent a low burden direct connection to the patient journey, and such data have already been demonstrated to be a valuable component of models that predict health-relevant outcomes [27,28]. We present a 2-phase approach for predicting longitudinal deterioration in depression status. In phase 1c, we increased the label density by generating intermediate PHQ-9 category labels using wearable PGHD and LMC information. In the second phase, we combined self-reported and generated PHQ-9 category labels with additional recent wearable PGHD and LMC information to predict the deterioration of depression status 3 months after the initial self-report. This 2-phase approach has a very low burden and requires very little participant interaction. The information we used as input consists of simple self-reports and data from consumer-grade wearables.

Even though overall performance in phase 1c was not particularly strong (κ=0.476, 95% CI ±0.017), we were encouraged by 2 factors: the adjacent accuracy was high (77.6%), and an examination of features in the final tuned models showed good correspondence to factors known to be important risk factors for depression, for example, gender,

experience of trauma, and chronic comorbidities. Large-scale studies have shown that these have an influence on depression [29]. We also observed that objective sleep features were selected. Sleep features and depressive disorders have been previously associated using low-cost wearable devices [9], PGHD [30], and smartphones [31]. Additionally, we observed that performance was not even across severity groups and was high for individuals with either relatively mild or relatively severe depression.

In phase 2c, our best performing model achieved a sensitivity of 55.4%, specificity of 65.3% (95% CI ±4.2%), and AUPRC of 0.31 (95% CI ±0.024). In comparison, simulating random assignment of 20.7% positive labels across 10 iterations of 1000 samples, we noted an AUPRC of 0.21, a sensitivity of 19.8%, and a specificity of 80.0%. This means that sensitivity nearly tripled, while specificity only slightly reduced. We prioritized sensitivity because the potential consequences of false negatives (ie, not identifying a person with deteriorating depression) is much higher than the cost of false positives (ie, incorrectly suspecting someone of deteriorating depression).

We observed that features from all input sources were selected in the best performing models, but with different relative importance. We saw that static features (ie, those defined at enrollment, which do not change afterwards) were selected, but were of relatively low importance. This included features that are known to be relevant to the risk of developing depression, including the presence of chronic comorbidities [32], ethnicity [33], financial difficulties [34], and pregnancy [35]. We also saw features derived from wearable devices, including trends in sleep onset time, percentage of sleep time spent awake, and overall number of hypersomnia days. The most important features were those generated in phase 1c, that is, the probability of an individual being in a given PHQ-9 class, summarizing features from across all input sources. The intermediate labels generated in phase 1c are inspired by the concept of "weak labeling," which can help reduce large-scale noisy data to a signal useful for supervised learning (eg, the report by Zhan et al [36]). We noted that due to data sparsity, intermediate labeling was not always available, and thus, some samples did not have 2 intermediate PHQ-9 category labels, but sometimes had 1 or none. Nonetheless, as LightGBM was able to deal with missing values, the lack of intermediate labeling or missing PGHD values did not pose problems in the phase 2c model predictions, highlighting that the approach described in our work is indeed low-burden and robust.

From this, we were able to deduce that the average sleep onset time is a good determinant of increasing depression severity, which is consistent with previous research [9], but that variability in sleep is participant specific and not necessarily a good predictor for generalizing to other participants.

Limitations

The work presented here demonstrates the potential of a PGHD-based model for predicting long-term changes in depression status in new individuals. This initial approach nevertheless has several limitations in practice, which will be addressed in future work.

The model relies on the completion of several self-reported surveys over time. Participants were highly engaged with the year-long research study, but to lower the barrier to participation, the number of surveys could be reduced or replaced with alternative sources of data. For example, instead of LMC surveys, medication change data could be obtained through electronic health records [37] or through other consumer-grade wearables that incorporate engagement, such as the Oura ring, which allows participants to annotate days with a number of tags like medication [38].

The performance of PSYCHE-D was below our initial expectations, despite more than triple sensitivity versus a random model, and was weaker than the initial nongeneralized performance [13,14]. However, further validation and prospective data collection could seek to build off this "out of the box" performance using an active learning approach to improve individualized performance [39,40]. We also plan to perform further validation with independently generated data [41]. The study design also limits us to making predictions of depression status change over a 3-month time window. Thus, future work will focus on testing predictions beyond that time horizon.

We will also explore the application of PSYCHE to other aspects of mental health like anxiety [31], fatigue [42], and stress [22].

Outlook and Conclusion

Effective treatments for depression exist, but they must be delivered in a timely manner, as the benefits of early intervention are established for both older [43] and younger [44] patients. Moreover, the objectivity of our system provides a nonstigmatizing environment to engage people about depression [4]. We hope that this demonstration of the ability to predict long-term changes in depression using a low-burden PGHD-based approach will have great potential to deliver value to patients.

Acknowledgments

We would like to thank Nikki Marinsek, Allison Shapiro, and Piyusha Gade for their advice, as well as Raul Eulogio and Filip Jankovic for technical input. We would also like to thank the study participants and the DiSCover project team (especially Shefali Kumar, Claire Meunier, Katie Ulhorn, Lauren Vucovich, Jermyn See, Christine Sang, Jen Lee, Kelly Scherer, and Jessie Juusola) for allowing us to work with the participant data sets. This work was entirely self-funded by Evidation Health Inc.

Authors' Contributions

MM and IC contributed to the conception of this work. All authors contributed to the analysis and interpretation of data. MM, RK, MF, and JM contributed to the code writing. All authors contributed to drafting and revising the manuscript. All authors approved the submitted and final version of the manuscript.

Conflicts of Interest

MM was a paid intern at Evidation Health and was completing her master's degree at École Polytechnique Fédérale de Lausanne at the time the work was completed. RK, MF, and JM are employees of and hold stock options in Evidation Health. IC has no competing interests to declare. MJ has no competing interests to declare.

Multimedia Appendix 1
Supplementary methods.
[DOCX File , 150 KB - mhealth_v10i3e34148_app1.docx]

Multimedia Appendix 2
Selected features.
[DOCX File , 95 KB - mhealth_v10i3e34148_app2.docx]

References

1. Depression. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2022-03-06]
2. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
3. Major Depression. National Institute of Mental Health. URL: https://www.nimh.nih.gov/health/statistics/major-depression.shtml [accessed 2022-03-06]
4. Chekroud AM, Foster D, Zheutlin AB, Gerhard DM, Roy B, Koutsouleris N, et al. Predicting Barriers to Treatment for Depression in a U.S. National Sample: A Cross-Sectional, Proof-of-Concept Study. Psychiatr Serv 2018 Aug 01;69(8):927-934 [FREE Full text] [doi: 10.1176/appi.ps.201800094] [Medline: 29962307]
5. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
6. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, et al. Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors. Front Psychiatry 2020;11:584711 [FREE Full text] [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
7. Vahedi Z, Zannella L. The association between self-reported depressive symptoms and the use of social networking sites (SNS): A meta-analysis. Curr Psychol 2019 Jan 26;40(5):2174-2189 [FREE Full text] [doi: 10.1007/s12144-019-0150-6]
8. Renn BN, Pratap A, Atkins DC, Mooney SD, Areán PA. Smartphone-Based Passive Assessment of Mobility in Depression: Challenges and Opportunities. Ment Health Phys Act 2018 Mar;14:136-139 [FREE Full text] [doi: 10.1016/j.mhpa.2018.04.003] [Medline: 30123324]
9. 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]
10. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2:3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
11. Sverdlov O, Curcic J, Hannesdottir K, Gou L, De Luca V, Ambrosetti F, et al. A Study of Novel Exploratory Tools, Digital Technologies, and Central Nervous System Biomarkers to Characterize Unipolar Depression. Front Psychiatry 2021 May 6;12:640741 [FREE Full text] [doi: 10.3389/fpsyt.2021.640741] [Medline: 34025472]
12. Lee J, Cerrada C, Vang M, Scherer K, Tai C, Tran J, et al. The DiSCover Project: Protocol and Baseline Characteristics of a Decentralized Digital Study Assessing Chronic Pain Outcomes and Behavioral Data. medRxiv. 2021. URL: https://www.medrxiv.org/content/10.1101/2021.07.14.21260523v1 [accessed 2022-03-06]
13. Makhmutova M, Kainkaryam R, Ferreira M, Min J, Jaggi M, Clay I. Prediction of self-reported depression scores using person-generated health data from a virtual 1-year mental health observational study. In: DigiBiom '21: Proceedings of the 2021 Workshop on Future of Digital Biomarkers. 2021 Presented at: 2021 Workshop on Future of Digital Biomarkers; June 25, 2021; Virtual Event Wisconsin. [doi: 10.1145/3469266.3469878]

14. Makhmutova M. Predicting changes in depression using person-generated health data. École Polytechnique Fédérale de Lausanne. 2021. URL: https://infoscience.epfl.ch/record/288435/files/mariko_makhmutova_master_thesis.pdf [accessed 2022-03-06]
15. 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]
16. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, 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]
17. Tudor-Locke C, Camhi SM, Troiano RP. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012;9:E113 [FREE Full text] [doi: 10.5888/pcd9.110332] [Medline: 22698174]
18. Korlakai VR, Gilad-Bachrach R. DART: Dropouts meet Multiple Additive Regression Trees. In: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics. 2015 Presented at: Eighteenth International Conference on Artificial Intelligence and Statistics; May 9-12, 2015; San Diego, CA, USA p. 489-497.
19. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning 2002;46:389-422. [doi: 10.1023/A:1012487302797]
20. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement 2016 Jul 02;33(3):613-619. [doi: 10.1177/001316447303300309]
21. Aha D, Bankert R. A Comparative Evaluation of Sequential Feature Selection Algorithms. In: Fisher D, Lenz H, editors. Learning from Data. Lecture Notes in Statistics, vol 112. New York, NY: Springer; 1996:199-206.
22. Sano A, Taylor S, McHill A, Phillips A, Barger L, 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]
23. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
24. Breiman L. Random Forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
25. PSYCHE-D GitHub open-source repository. GitHub. URL: https://github.com/evidation-opensource/PSYCHE-D [accessed 2022-03-06]
26. Makhmutova M, Kainkaryam R, Ferreira M, Min J, Jaggi M, Clay I. PSYCHE-D: predicting change in depression severity using person-generated health data. JMIR mHealth and uHealth (forthcoming) 2022. [doi: 10.2196/34148]
27. Karas M, Marinsek N, Goldhahn J, Foschini L, Ramirez E, Clay I. Predicting Subjective Recovery from Lower Limb Surgery Using Consumer Wearables. Digit Biomark 2020;4(Suppl 1):73-86 [FREE Full text] [doi: 10.1159/000511531] [Medline: 33442582]
28. Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Fiorenza Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. Nat Med 2021 Jun;27(6):1105-1112 [FREE Full text] [doi: 10.1038/s41591-021-01339-0] [Medline: 34031607]
29. Brailean A, Curtis J, Davis K, Dregan A, Hotopf M. Characteristics, comorbidities, and correlates of atypical depression: evidence from the UK Biobank Mental Health Survey. Psychol Med 2020 May;50(7):1129-1138. [doi: 10.1017/S0033291719001004] [Medline: 31044683]
30. Kumar S, Tran J, Ramirez E, Lee W, Foschini L, Juusola J. Design, Recruitment, and Baseline Characteristics of a Virtual 1-Year Mental Health Study on Behavioral Data and Health Outcomes: Observational Study. JMIR Ment Health 2020 Jul 23;7(7):e17075 [FREE Full text] [doi: 10.2196/17075] [Medline: 32706712]
31. 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;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
32. Li H, Ge S, Greene B, Dunbar-Jacob J. Depression in the context of chronic diseases in the United States and China. Int J Nurs Sci 2019 Jan 10;6(1):117-122 [FREE Full text] [doi: 10.1016/j.ijnss.2018.11.007] [Medline: 31406877]
33. Bailey RK, Mokonogho J, Kumar A. Racial and ethnic differences in depression: current perspectives. Neuropsychiatr Dis Treat 2019;15:603-609 [FREE Full text] [doi: 10.2147/NDT.S128584] [Medline: 30863081]
34. Richardson T, Elliott P, Roberts R, Jansen M. A Longitudinal Study of Financial Difficulties and Mental Health in a National Sample of British Undergraduate Students. Community Ment Health J 2017 Apr;53(3):344-352 [FREE Full text] [doi: 10.1007/s10597-016-0052-0] [Medline: 27473685]
35. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2019 Feb 12;321(6):588-601. [doi: 10.1001/jama.2018.20865] [Medline: 30747970]

36. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. JAMA Neurol 2018 Jul 01;75(7):876-880 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0809] [Medline: 29582075]
37. Kukafka R, Ancker JS, Chan C, Chelico J, Khan S, Mortoti S, et al. Redesigning electronic health record systems to support public health. J Biomed Inform 2007 Aug;40(4):398-409 [FREE Full text] [doi: 10.1016/j.jbi.2007.07.001] [Medline: 17632039]
38. Ōura Ring company homepage. Ōura Ring. URL: https://ouraring.com/ [accessed 2022-03-06]
39. Huang S, Jin R, Zhou Z. Active Learning by Querying Informative and Representative Examples. IEEE Trans Pattern Anal Mach Intell 2014 Oct;36(10):1936-1949. [doi: 10.1109/TPAMI.2014.2307881] [Medline: 26352626]
40. Bachman P, Sordoni A, Trischler A. Learning Algorithms for Active Learning. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, NSW, Australia p. 301-310.
41. Goldsack JC, Dowling AV, Samuelson D, Patrick-Lake B, Clay I. Evaluation, Acceptance, and Qualification of Digital Measures: From Proof of Concept to Endpoint. Digit Biomark 2021;5(1):53-64 [FREE Full text] [doi: 10.1159/000514730] [Medline: 33977218]
42. Luo H, Lee P, Clay I, Jaggi M, De Luca V. Assessment of Fatigue Using Wearable Sensors: A Pilot Study. Digit Biomark 2020;4(Suppl 1):59-72 [FREE Full text] [doi: 10.1159/000512166] [Medline: 33442581]
43. Reynolds CF, Cuijpers P, Patel V, Cohen A, Dias A, Chowdhary N, et al. Early intervention to reduce the global health and economic burden of major depression in older adults. Annu Rev Public Health 2012 Apr;33:123-135 [FREE Full text] [doi: 10.1146/annurev-publhealth-031811-124544] [Medline: 22429161]
44. Schley C, Pace N, Mann R, McKenzie C, McRoberts A, Parker A. The headspace Brief Interventions Clinic: Increasing timely access to effective treatments for young people with early signs of mental health problems. Early Interv Psychiatry 2019 Oct;13(5):1073-1082. [doi: 10.1111/eip.12729] [Medline: 30160372]

Abbreviations

AUPRC: area under the precision-recall curve
DART: Dropouts Meet Multiple Additive Regression Trees
DiSCover: Digital Signals in Chronic Pain
IRB: Institutional Review Board
LightGBM: Light Gradient Boosting Machine
LMC: lifestyle and medication changes
PGHD: person-generated health data
PHQ-9: Patient Health Questionnaire-9
PSYCHE-D: Prediction of Severity Change-Depression
SM: sample month

Edited by L Buis; submitted 08.10.21; peer-reviewed by C Stewart, M Kraus; comments to author 08.12.21; revised version received 21.12.21; accepted 11.02.22; published 25.03.22.

Please cite as:
Makhmutova M, Kainkaryam R, Ferreira M, Min J, Jaggi M, Clay I
Predicting Changes in Depression Severity Using the PSYCHE-D (Prediction of Severity Change-Depression) Model Involving Person-Generated Health Data: Longitudinal Case-Control Observational Study
JMIR Mhealth Uhealth 2022;10(3):e34148
URL: https://mhealth.jmir.org/2022/3/e34148
doi:10.2196/34148
PMID:35333186

Original Paper

Reaching Patients With Noncommunicable Diseases in Rural Tanzania Using Mobile Devices and Community Trust: Qualitative Study

Ayano Miyashita[1], MSc; Keiko Nakamura[1], MD, PhD; Mayumi Ohnishi[2], RN, MPH, PhD; Deogratius Bintabara[1,3], MD, PhD; Festo K Shayo[1], MD, MSc, PhD; Isaac I Maro[1,4], MD, MPH, PhD; Hideko Sato[1], RN, MPH; Kaoruko Seino[1], MSc, PhD; Stephen Kibusi[3], MPH, PhD

[1]Department of Global Health Entrepreneurship, Tokyo Medical and Dental University, Tokyo, Japan

[2]Department of Community-Based Rehabilitation Sciences, Nagasaki University, Nagasaki, Japan

[3]College of Health Sciences, The University of Dodoma, Dodoma, United Republic of Tanzania

[4]The Ifakara Health Institute, Dar es Salaam, United Republic of Tanzania

Corresponding Author:
Keiko Nakamura, MD, PhD
Department of Global Health Entrepreneurship
Tokyo Medical and Dental University
Yushima 1-5-45, Bunkyo-ku
Tokyo, 113-8519
Japan
Phone: 81 3 5803 4048
Fax: 81 3 5803 4034
Email: nakamura.ith@tmd.ac.jp

Abstract

Background: A health service using mobile devices, mobile health (mHealth), has been widely applied to programs focusing on maternal and child health and communicable diseases in sub-Saharan African countries. However, mHealth apps for noncommunicable disease (NCD) services remain limited.

Objective: This study aimed to explore the acceptability and potential usability of SMS text messaging for patients and health care providers for the management of NCDs as part of an implementation research in rural Tanzania.

Methods: Nine focus group discussions were conducted with 56 participants (21 community health workers [CHWs], 17 patients, and 18 health care professionals [HPs]) in 3 districts in the Dodoma region, Tanzania. The interview guides were prepared in Swahili, and each session was recorded, transcribed, and translated into English. The focus group discussions consisted of the following topics: (1) perceptions of the participants about the possible use of mobile devices and SMS text messages as an mHealth platform in community health services; and (2) experiences of mobile device use in health activities or receiving health services via a mobile phone in the past.

Results: CHWs and HPs reported having familiarity using mobile devices to provide health services, especially for reaching or tracing patients in remote settings; however, patients with NCDs were less familiar with the use of mobile devices compared with the other groups. Hesitation to receive health services via SMS text messaging was seen in the patient group, as they wondered who would send health advice to them. Some patients expected services beyond what mHealth could do, such as aiding in recovery from a disease or sending notifications about the availability of prescription medications. CHWs showed interest in using text messaging to provide health services in the community; however, the concerns raised by CHWs included the cost of using their own mobile devices. Moreover, they demanded training about NCD management before engaging in such an activity.

Conclusions: This study explored views and experiences regarding the possible installation of an mHealth intervention for managing NCDs in rural Tanzania. Although HPs and CHWs had experience using mobile devices to provide health services in non-NCD projects, only a few patients (3/17, 17%) had heard about the use of mobile devices to receive health services. To improve the suitability and acceptability of the intervention design for patients with NCDs, their trust must be earned. Involving CHWs in the intervention is recommended because they have already been appointed in the community and already know how to communicate effectively with patients in the area.

(JMIR Mhealth Uhealth 2022;10(3):e29407) doi:10.2196/29407

KEYWORDS

noncommunicable disease; community health workers; Tanzania; communication; rural; community; trust; disease; acceptability; usability; text message; SMS; mobile phone; implementation

Introduction

In 2011, mobile health (mHealth) was defined as "the use of mobile and wireless technologies to support the achievement of health objectives" by the World Health Organization [1] and is expected to be an increasingly important tool for patients in resource-scarce settings due to the fast-growing penetration rates of mobile phones in developing countries [2-4]. mHealth has already been implemented for a wide range of health promotion and management activities: to act as reminders of clinic visits, to encourage medication adherence, and to promote behavioral changes [5,6]. The increase in the numbers of patients with noncommunicable disease (NCDs) is likely to dominate health care needs and expenditures in most low-income and middle-income countries [7]. Tanzania is no exception and is facing an increasing prevalence of NCDs at a time when its health system is already overloaded with a high prevalence of communicable diseases [8,9]. A national survey in 2012 showed that the prevalence of Diabetes Mellitus (DM) and hypertension (HT) was found to be 9.1% and 25.9%, respectively [10]. In 2016, there were approximately 0.14 physicians per 10,000 individuals nationwide, with a lower ratio in the rural areas [11]. The Tanzania Ministry of Health's strategic and action plan for the prevention and control of NCDs is targeting to reduce mortality by 25% from baseline by 2025 by improving patients' disease management through community and facility care interventions [12]. Although the country has a well-thought-out NCD strategic plan, the already overburdened health system is limiting the effective control of NCDs. mHealth, which uses text messages in Tanzania, has been applied only in programs to combat communicable diseases and reproductive health issues. This study is a part of the Community mHealth Integrated Care (ComHIC) to manage hypertension (HT) or diabetes (DM) in Tanzania's overburdened health system [13]. Our project aims to apply mHealth with support by community health workers (CHWs) under an NCD program to overcome the lack of resources and infrastructure for health service access in the rural Dodoma districts. Our mHealth system is aiming to support the existing local health systems to help control and manage patients with HT and DM. The objectives of this study were to obtain insights into perceptions of the concept of mHealth among CHWs, patients with NCDs, and health care professionals (HPs) in rural Tanzania, to assess the possible use of mobile devices and SMS text messaging as an mHealth platform among CHWs and patients, and to develop an appropriate mHealth intervention for the study site.

Methods

Setting and Research Team

Focus group discussions (FGDs) were conducted in the Kondoa, Kongwa, and Mpwapwa districts in the Dodoma region of Tanzania from January to February 2020. The research team consists of health system researchers, health literacy and communication experts with experience of qualitative data collection, physicians, and nurses with experiences in Tanzania, who developed a protocol of the study.

Eligibility Criteria of the Participants

Quota sampling method was applied to recruit 3 participant groups of CHWs, patients with diabetes or hypertension, and HPs at each district hospital, a total of 9 groups with 6-8 people in each group. A Dodoma regional medical officer gave a recruiting instruction to NCD coordinators who work for the NCD clinics at each district hospital, and they called in the potential candidates. For the 3 groups, the candidates were listed based on the following eligibility criteria: for all—no gender quota; aged above 18 years; and willing and capable to attend a 60-minute interview session held at the district hospital; for CHWs—those who have registered at the district hospitals and therefore have been engaging in a health program or have worked as a CHW in the past; for patients—attending NCD clinics at the district hospital more than 1 year for HT, DM, or both; and for HPs—those who are currently working at NCD clinics at the district hospitals.

Interview Guides Development

The 3 interview guides for CHWs, for patients with DM or HT, and for HPs were prepared in Swahili by the study team through a series of discussions among the study team. A semistructured interview design was adopted with consultations with researchers who had experience in designing interview guides. The main topics of the discussions were (1) perceptions of the participants about the possible use of mobile devices and SMS text messaging as an mHealth platform in community health services; and (2) experiences of mobile device use in health activities or receiving health services via a mobile phone in the past.

Qualitative Interview and Analysis

A facilitator and a notetaker, both native Swahili speakers, moderated the FGD sessions. Prior to the initiation of the FGDs, the facilitator gave a brief explanation of the study purpose and obtained written informed consent from all participants who had agreed to participate. All FGDs were conducted in a room at the district hospital. The FGDs, which were approximately 60 minutes per session, were audio recorded, transcribed, and translated into English. The transcripts were then uploaded to NVivo, version 12 (QSR International) [14], and the analysis was performed inductively. The initial sets of code were extracted by 2 investigators individually, and they were compared to each other to increase transparency in the process of code and category generation and to ensure consistency in the codes. Any inconsistencies or questions that arose during the individual analysis were recorded and discussed between the 2 investigators. When such issues could not be resolved by comparing the generated sets of code, the investigators referred

to the original texts and interviewed the researcher members who conducted the interviews. Those processes of reexamination were repeated until the final analysis results were obtained. Further, the results were reviewed by the study team including members who facilitated the focus group interviews and the physicians and nurses who were familiar with the regional context for the analyses to capture the participants' perspectives.

Ethics

Ethical clearance certificate for conducting medical research in Tanzania was obtained from the National Institute for Medical Research (HQ/R.8a/Vol.IX/3220). The study was also approved by the Research Review Committee of Tokyo Medical and Dental University, Japan (M2019-191).

Results

Participants

Nine FGDs were conducted with a total of 21 (38%) CHWs, 17 (30%) patients, and 18 (32%) HPs (n=56; 22 men [39%], 34 women [61%]) in the 3 districts in the Dodoma region. The age ranges of the participants were 24-61 years for CHWs, 27-54 years for HPs, and 55-68 years for patients; however, 11 (65%) of the 17 patients did not agree to provide their date of birth, as some believed the information could be used for witchcraft. Multimedia Appendix 1 shows the participants' perspectives and experiences regarding the use of mHealth. The results of the discussion are described as follows.

Perceptions of mHealth

The participants' perceptions about the possible use of mobile devices and SMS text messaging as an mHealth platform in community health services are as follows:

Positive Perceptions

All groups provided essentially positive feedback about the use of mHealth for supporting NCD control and management. CHWs supported the idea of using SMS text messaging to contact patients for several reasons. The main reason was saving time and resources, as mHealth enables them to contact many patients in a limited time, to overcome transportation issues, and to save time reporting. Possible mobile uses for NCD control included providing health information, offering encouragement, and sending reminders for clinical appointments and taking medications.

> *I see this is a simple way of communicating with them because you cannot visit them all; you may find that you have eighty or ninety patients; you cannot visit all of them; so mobile phones are the easiest means.* [CHW, Kondoa]

The majority of patients (4/5, 80% in Kondoa; 5/6, 83% in Kongwa; and 5/6, 83% in Mpwapwa) agreed about the use of mobile devices for reminders for clinical appointments and medications. All FGD participants were confident about reading SMS text messages and using mobile devices for communication. Some patients were willing to follow treatment guidelines if they were provided with one via mobile, while others said that regular SMS text messages would make them feel encouraged and supported.

> *It's a good idea, everybody knows how to read a message here. It is readable, and I think more beautiful.* [Patient, Mpwapwa]

HPs agreed about the use of SMS text messages to support patients with NCD, as they believe the majority of people currently have mobile devices, and they have seen some successful health projects using mobile devices in the past. They think that messages can motivate patients to follow their recommendations and reduce the number of missed appointments; therefore, they hope the intervention could lead to a reduced burden of complications, blood pressure, and blood sugar. Some believe that emotional support for patients with long-term health issues could be provided by text because some patients have been described as feeling lonely.

> *I think it's a good idea. We have a habit that everyone likes to be reminded of things, and for most patients, especially those with loneliness diseases, a little reminder makes him/her know that you do care.* [HP, Kongwa]

The use of mHealth as an educational tool was also expected, and reaching patients who do not know how to read or have a mobile device was suggested via family members.

> *If I try to look at the kind of patients we see, most are old, and these old people are over sixty years of age; most of them don't use mobile phones on a large scale, and even if they do, they maybe just receive a call from their children. But many have difficulty reading; so, the use of mobile phones may not necessarily help to reach them directly. We can reach them through their children and grandchildren and explain things to them as they receive the text message.* [HP, Kondoa]

Negative Perceptions

As CHWs had already been using their mobile devices to reach out to patients for the current project, they were also aware of the possible difficulties that may arise. They mentioned that not all patients have a mobile device; however, they said it was possible to reach patients via a mobile device belonging to a relative or someone the patients could trust. Their biggest concern was how the project provides funds to purchase airtime vouchers, as many of them cover the cost out-of-pocket.

> *To add to that, it is our routine/arrangement that every patient must provide a mobile phone number; if he/she doesn't have a number, then they provide the number of aclose relative to whom he/she trusts to share his/her health problem. We have established this system so that we can reach our clients; otherwise, we would lose many*. [CHW, Kondoa]

Some patients had never heard the term NCDs, so they were not very sure about the content of the texts and from whom they would receive them. Especially in Mpwapwa, the concept of mHealth was not easily grasped during the interview.

> *I see the question… but maybe the question was, I did not really understand, if I ever received any messages on the phone about these diseases we are talking about, diabetes and hypertension? I said I never did,*

> *except if it is health care, this the system that the whole of Tanzania should use to be treated.* [Patient, Mpwapwa]

> *Really good advice, now how do they get my phone number? And who will send me such a message? You or someone else? He has to, he doesn't have my phone number, I don't know him, how will he serve me?* [Patient, Mpwapwa]

Although the majority of the HPs provided positive feedback about mHealth, some shared concerns that the use of mHealth only may not be sufficient as an intervention. They suggested combining other types of health promotion such as TV, radio, billboards, and social networking services to not only cover patients who are illiterate or do not have a mobile device, but also to promote health in the general population.

> *Yes, it can help a lot for those who have access to cell phones because not all of them have phones or they have them but they cannot use them; he/she can answer a call but cannot read a text message. Let's not just use the cell phones; also, posters like that big one at the bus terminal with a message "TB is preventable." It can also be done through TV because many watch TV.* [HP, Kondoa]

Other concerns were related to difficulties eliciting behavioral change, general negligence toward interventions, and a lack of awareness leading to misperceptions.

Recommendations or Expectations for mHealth for Each Participant Group

CHWs expect the project to provide airtime vouchers for communicating with patients. In addition, some hope to have a device such as tablet or smartphone for reporting their activities. Regarding the mode of communication, in general, CHWs prefer two-way communication.

> *My opinions are: we suggest making it possible to be funded for airtime vouchers so that we can communicate with our clients. Most of them have phones, only a few of them do not. If we would be given airtime vouchers, we could contact our clients anytime we needed to.* [CHW, Kongwa]

Once patients understood the mHealth concept, reminders for clinical appointments and taking medications were popular functions they hoped to have. Some described that they were expecting to receive medical advice via their mobile device, while others expected to see clinical outcomes such as decreased blood pressure or alleviated complications after following the obtained advice. Most preferred receiving short messages, except some, who preferred calls instead of texts for medication reminders. Expectations of being able to order prescribed medications and receive notifications about medication supplies were also reported.

> *I really thought it would be better because sometimes you can forget the time to take a medicine; but if we get a call to remind us, it will be better for me.* [Patient, Kongwa]

HPs suggested SMS text message contents such as giving reminders of clinical appointments and medication times, promoting dietary salt reduction and exercise, and advising about lifestyle modifications. Regarding strategies for the effective delivery of messages to patients, concrete advice such as sending a series of short, well-written messages at least once a week (the time of the day depending on the disease) and risk communication, including the impact of complications and the economic burden for families, were also raised during the interviews.

> *There are reminders that could be sent on lifestyle modifications and adherence to medications, so these could be just short messages, but specific, such as only food intake or exercise, not just one message that has everything in it.* [HP, Mpwapwa]

Experiences With mHealth

Here, we turn to the experiences of mobile device use in health activities or receiving health services via a mobile phone in the past. FGDs included participant critiques of using mobiles, which are as follows:

1. CHWs reported that at least 5 organizations were using some sort of communication tools, including those for maternal and child health, family planning, human immunodeficiency virus infection and acquired immunodeficiency syndrome (HIV/AIDS), and tuberculosis. Some nongovernmental organizations provide a tablet to every CHW so that they can write a monthly report.
2. Patients in the Mpwapwa district were not aware of any projects using mobile devices. One of the patients in the Kongwa district mentioned the use of mobile devices for reminders to patients about clinical visits at a referral hospital. Two patients in Kondoa reported hearing about HIV/AIDS projects that use mobile devices to contact patients for checking and tracing (3/17, 17% patients).
3. HPs reported knowing about various projects that were using mHealth for adherence support and follow-up for patients, such as those for maternal and child health, family planning, tuberculosis, and leprosy. Apparently, mobile devices were not the only mode of communication; for example, radios were used for a tuberculosis project and billboards for an HIV project. HPs also shared the lessons learned from past projects; for instance, messages that were too long to read were often ignored by the receivers, and inconsistent content demoralized participants in terms of motivation.

Discussion

Regarding the extent of the brief explanation of mHealth during the interviews, the idea of text-based interventions for NCD management was positively accepted by the CHWs, patients, and HPs. However, some patients had difficulty understanding the concept. Hesitation (eg, who sends the messages?) and unrealistically high expectations (eg, it may fix all problems) about mHealth's effectiveness were reported. Positive reasoning was based on the commonality of mobile device use for daily activities and geographical hardships experienced during

health-seeking and health-providing behaviors. CHWs and HPs supported the use of mobile devices as they had already been using them to reach out to their patients. Even if those targeted among the older population cannot read or do not have a mobile device, CHWs and HPs believed that their children or grandchildren could read out the texts for them.

According to the 2012 STEP (World Health Organization Stepwise Approach to NCD Risk Factor Surveillance) survey in Tanzania, the prevalence of diabetes for men increased with age and reached at peak at the age group of 45-54 while the prevalence for women was high at the age group of 55-64. In the same survey, the prevalence of both diastolic and systolic blood pressure increased with age, and it reached peak level at the age group of 55-64 years [10]. This study covered the major age range of the DM and HT patients in Tanzania. mHealth interventions must be accessible to patients with NCD in rural Dodoma, and if SMS text messages are used, they must be effective and deemed culturally appropriate in order to manage the disease. For example, based on observations made by the facilitator during the focus groups, the research team interpreted different levels of understanding for the mHealth concept among patients across 3 districts. Dialects and cultural diversity were reported within the region. A CHW suggested a combination of SMS text messaging and supports by CHWs; this would facilitate patients' management of disease and lead to culturally appropriate understanding by local patients. The mHealth intervention program would ideally be implemented locally in terms of consistency and sustainability.

Although the utility of text messages was positively accepted among patients, preferably, the sender of the text messages should acquire community trust in advance to disseminate the NCD management information. HPs and CHWs shared some concerns that a text-only approach may not be sufficient to elicit behavioral change among patients. CHWs stated that they would also like to continue making phone calls, as calling may be easier than typing text messages (although our intervention design did not ask CHWs to prepare text messages), and patients' understanding can be checked immediately. HPs preferred to combine text messages with other means of health promotion, such as billboards, TV, or radio, since they thought those means cover more people and offer more preventive effects.

The involvement of CHWs in the program could be helpful to reinforce the outcomes of text-based learning or reminders for the patients. CHWs were appointed by a chairperson in village meetings, and they know how to authorize their activity in each community. As a result, they could connect the community (patients) with health care facilities (HPs). CHWs' knowledge about how to operate health projects in socially tight communities could be valuable to acquire community trust and achieve effective communication with patients. Involving experienced CHWs in the mHealth intervention may help to improve the implementation fidelity of the project.

To achieve the above, training for CHWs about managing patients with NCDs is necessary. Many CHWs have expressed uncertainty regarding knowledge about NCDs, and if there were such a program, they would like to undergo the necessary training before joining. Some concerns were also raised by CHWs such as the cost of using their own mobile device when working for patients, which they must do in current programs. To improve the acceptability and sustainability of the intervention, it is important to minimize the cost of using the intervention for CHWs.

Acknowledgments

The study has been supported by Japan Agency for Medical Research and Development and Grant in Aid for Scientific Studies, Japan Society for Promotion of Science grant JP19H04511.

Authors' Contributions

KN and SK conceptualized the research framework and designed the study. KN, MO, SK, IIM, DB, FKS, and AM developed focus group discussion guidelines. DB and IIM conducted focus group discussions and collected the data. DB and FKS compiled and translated the data. AM, HS, KN, MO, SK, IIM, DB, FKS, and SK analyzed and interpreted the qualitative data. AM drafted the manuscript. KN and MO reviewed and provided conceptual advice for the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Perspectives of the participants in the focus group discussion.
[PDF File (Adobe PDF File), 78 KB - mhealth_v10i3e29407_app1.pdf]

References

1. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2022-01-25]

2. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
3. Goel S, Bhatnagar N, Sharma D, Singh A. Bridging the Human Resource Gap in Primary Health Care Delivery Systems of Developing Countries With mHealth: Narrative Literature Review. JMIR Mhealth Uhealth 2013 Dec 03;1(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2688] [Medline: 25099436]
4. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote Patient Monitoring via Non-Invasive Digital Technologies: A Systematic Review. Telemedicine and e-Health 2017 Jan;23(1):3-17. [doi: 10.1089/tmj.2016.0051]
5. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013 Jun;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
7. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
8. Mayige M, Kagaruki G, Ramaiya K, Swai A. Non communicable diseases in Tanzania: a call for urgent action. Tanzania J Hlth Res 2011 Dec;13(5):378-386. [doi: 10.4314/thrb.v13i5.7] [Medline: 26591992]
9. Edwards R, Unwin N, Mugusi F, Whiting D, Rashid S, Kissima J, et al. Hypertension prevalence and care in an urban and rural area of Tanzania. Journal of Hypertension 2000;18(2):145-152. [doi: 10.1097/00004872-200018020-00003]
10. Mayige M, Kagaruki G. Tanzania STEPS survey report. National Institute for Medical Research. 2013. URL: https://www.who.int/ncds/surveillance/steps/UR_Tanzania_2012_STEPS_Report.pdf [accessed 2022-01-27]
11. Physicians (per 1,000 people). The World Bank. URL: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS [accessed 2022-01-25]
12. Strategic and action plan for the Prevention and control of non communicable diseases in Tanzania 2016 - 2020. World Diabetes Foundation. 2016. URL: https://www.worlddiabetesfoundation.org/sites/default/files/NCD [accessed 2022-01-25]
13. Community mHealth Integrated Care (ComHIC) to manage hypertension (HT)/diabetes (DM) in Tanzania's overburdened health system. Japan Agency for Medical Research and Development. URL: https://www.amed.go.jp/en/news/program/0301C_00025.html [accessed 2022-01-25]
14. NVivo Qualitative Data Analysis Software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-01-25]

Abbreviations

CHW: community health worker
ComHIC: Community mHealth Integrated Care
DM: Diabetes Mellitus
FGD: focus group discussion
HP: health care professional
HT: hypertension
mHealth: mobile health
NCD: noncommunicable disease
STEP: World Health Organization Stepwise Approach to NCD Risk Factor Surveillance

Edited by G Eysenbach; submitted 08.04.21; peer-reviewed by T Abraham; comments to author 29.04.21; revised version received 25.05.21; accepted 31.12.21; published 17.03.22.

Please cite as:
Miyashita A, Nakamura K, Ohnishi M, Bintabara D, Shayo FK, Maro II, Sato H, Seino K, Kibusi S
Reaching Patients With Noncommunicable Diseases in Rural Tanzania Using Mobile Devices and Community Trust: Qualitative Study
JMIR Mhealth Uhealth 2022;10(3):e29407
URL: https://mhealth.jmir.org/2022/3/e29407
doi:10.2196/29407
PMID:35297772

Original Paper

Web-Based TangPlan and WeChat Combination to Support Self-management for Patients With Type 2 Diabetes: Randomized Controlled Trial

Shu-Fang Xia[1*], PhD; Gusonghan Maitiniyazi[1*], BSc; Yue Chen[1*], BSc; Xiao-Ya Wu[2*], BSc; Yu Zhang[3], BSc; Xiao-Yan Zhang[4], BSc; Zi-Yuan Li[1], BSc; Yuan Liu[1], BSc; Yu-Yu Qiu[1], PhD; Jun Wang[2], MD, PhD

[1]Wuxi School of Medicine, Jiangnan University, Wuxi, China

[2]Department of Rehabilitation, Wuxi 9th Affiliated Hospital of Soochow University, Wuxi, China

[3]Department of Endocrinology, Affiliated Hospital of Jiangnan University, Wuxi, China

[4]Department of Hepatobiliary Surgery, Affiliated Hospital of Jiangnan University, Wuxi, China

[*]these authors contributed equally

Corresponding Author:
Jun Wang, MD, PhD
Department of Rehabilitation, Wuxi 9th Affiliated Hospital of Soochow University
NO. 999, Liangxi Road
Wuxi, 214063
China
Phone: 86 51085867999
Email: wangjun2115@suda.edu.cn

Abstract

Background: China has the largest number of patients with type 2 diabetes mellitus (T2DM) in the world. However, owing to insufficient knowledge of self-management in patients with diabetes, blood glucose (BG) control is poor. Most diabetes-related self-management applications fail to bring significant benefits to patients with T2DM because of the low use rate and difficult operation.

Objective: This study aims to examine the effectiveness of the combination of the self-designed web-based T2DM management software TangPlan and WeChat on fasting BG (FBG), glycated hemoglobin (HbA_{1c}), body weight, blood pressure (BP), and lipid profiles in patients with T2DM over a 6-month period.

Methods: Participants were recruited and randomized into the TangPlan and WeChat or control groups. Participants in the control group received usual care, whereas the TangPlan and WeChat participants received self-management guidance with the help of TangPlan and WeChat from health care professionals, including BG self-monitoring; healthy eating; active physical exercise; increasing medication compliance; and health education during follow-ups, lectures, or web-based communication. They were also asked to record and send self-management data to the health care professionals via WeChat to obtain timely and effective guidance on diabetes self-management.

Results: In this study, 76.9% (120/156) of participants completed the 6-month follow-up visit. After the intervention, FBG (mean 6.51, SD 1.66 mmol/L; *P*=.048), HbA_{1c} (mean 6.87%, SD 1.11%; *P*<.001), body weight (mean 66.50, SD 9.51 kg; *P*=.006), systolic BP (mean 127.03, SD 8.00 mm Hg; *P*=.005), diastolic BP (mean 75.25, SD 5.88 mm Hg; *P*=.03), serum low-density lipoprotein cholesterol (mean 2.50, SD 0.61 mmol/L; *P*=.006), and total cholesterol (mean 4.01, SD 0.83 mmol/L; *P*=.02) in the TangPlan and WeChat group were all significantly lower, whereas serum high-density lipoprotein cholesterol (mean 1.20, SD 0.25 mmol/L; *P*=.01) was remarkably higher than in those in the control group. Compared with the baseline data, significance was found in the mean change in FBG (95% CI −0.83 to −0.20; *P*=.002), HbA_{1c} (95% CI −1.92 to −1.28; *P*<.001), body weight (95% CI −3.13 to −1.68; *P*<.001), BMI (95% CI −1.10 to −0.60; *P*<.001), systolic BP (95% CI −7.37 to −3.94; *P*<.001), diastolic BP (95% CI −4.52 to −2.33; *P*<.001), triglycerides (95% CI −0.16 to −0.03; *P*=.004), serum low-density lipoprotein cholesterol (95% CI −0.54 to −0.30; *P*<.001), and total cholesterol (95% CI −0.60 to −0.34; *P*<.001) in the TangPlan and WeChat group but not in the control group (*P*=.08-.88).

Conclusions: Compared with usual care for patients with T2DM, the combination of TangPlan and WeChat was effective in improving glycemic control (decrease in HbA_{1c} and BG levels) and serum lipid profiles as well as reducing body weight in patients with T2DM after 6 months.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000028843; https://tinyurl.com/559kuve6

(JMIR Mhealth Uhealth 2022;10(3):e30571) doi:10.2196/30571

KEYWORDS

type 2 diabetes; glucose control; TangPlan; WeChat; self-management

Introduction

Background

In recent years, diabetes mellitus (DM) has gradually become an epidemic worldwide. China is the epicenter of the diabetes epidemic worldwide with an estimated 114.4 million people with diabetes in 2017 [1], mainly type 2 DM (T2DM). In a nationally representative cross-sectional survey conducted in 2013 in mainland China, the estimated overall prevalence of diabetes and prediabetes was 10.9% and 35.7%, respectively [2]. It has been suggested that medical management of diabetes alone before complications already accounts for 8.5% of national health expenditure in China, which puts heavy financial burdens on the country and patients [3]. Despite the development of therapeutic drugs and advanced techniques, the blood glucose (BG) levels of patients with DM are still poorly controlled [4], leading to various complications [5]. Self-management is considered the most critical factor in ensuring well-controlled BG levels, thereby preventing DM complications [6,7]. Self-management includes tracking BG trends, adhering to medication, monitoring nutrition, and increasing physical activity based on good diabetes health education [8]. Continuous DM care needs effective self-management education and support for both patients and family members.

Previous studies have found that increasing communication with health care professionals and enhancing diabetes management are beneficial for glycemic control [9]. However, Chinese physicians are usually very busy at work; patients only have a few minutes to consult physicians and fail to gain detailed DM self-management knowledge in a limited time. Furthermore, acute diabetic conditions and related chronic conditions force physicians to solve at least two symptomatic problems during the visit rather than the more time-consuming management of diabetes [10]. Only 20% of physicians consider they have the necessary resources to effectively manage patients with diabetes [11] and, during a visit, diabetes-specific assessments such as BG monitoring trends, history of hypoglycemia, foot examination, and blood pressure (BP) are not always performed. In addition, few patients have diabetes diaries, which prevents physicians from providing effective treatment guidance [12].

Growing evidence suggests that emerging telemedicine may further improve diabetes self-management and clinical outcomes via the establishment of an active interaction between patients and health care professionals. Apps are feasible tools to improve self-management of T2DM [13,14], resulting in positive self-management behaviors such as appropriate diet, enhanced physical activity, and BG monitoring [15]. However, despite these positive outcomes, the reality is that only a small proportion of patients use apps for diabetes self-management in China. In particular, older adults with diabetes, who account for 80.8% of the total number of patients with diabetes [16], are unable to use all kinds of professional apps well, resulting in poor use of apps for diabetes self-management. Moreover, most of the apps are completely new, which is not easy for middle-aged and older Chinese adults to use. Even in high-income countries, diabetes app use rate is still low. In Australia, only 8% of patients with diabetes use apps to support diabetes self-management [17]. Furthermore, owing to the lack of interconnected internetworking systems in different hospitals across China, it is very difficult for health care professionals to obtain the patient's diagnosis and treatment records from other hospitals and continuous follow-up information on patients.

With the development of mobile health, the ways of acquiring medical consultations have changed. WeChat is an extremely popular social app in China, and it is also simple for older adults to operate. Many researchers have reported the effectiveness of WeChat in chronic disease management, including diabetes, hypertension, cancer, obesity, stable coronary artery disease, and chronic obstructive pulmonary disease [18-21]. However, it is also challenging for health care professionals to give patients with diabetes detailed self-management advice based only on WeChat, including diet and exercise advice. Therefore, we designed a diabetes management software for health care professionals, TangPlan, which is based on Chinese culture and can be used in conjunction with WeChat to provide detailed diabetes self-management advice.

Objective

The primary objective of this study is to estimate the impact of the combination of TangPlan and WeChat on BG, glycated hemoglobin (HbA_{1c}), body weight, BP, and lipid profiles in patients with T2DM for a 6-month period.

Methods

Study Design

This study was designed as a 6-month, nonblinded randomized controlled trial (ChiCTR2000028843) between April 1, 2020, and October 31, 2020, to examine the efficacy and feasibility of the combination of the web-based TangPlan and WeChat on BG control in patients with T2DM. Potential participants registered in the community who received the notification calls about the trial and came for clinic visits were identified by trained health workers in a community health care center in Wuxi, China. These participants had established a health record

in the community health care center before January 2020. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist is presented in Multimedia Appendix 1.

Ethics Approval

This study was approved by the Jiangnan University Medical Ethics Committee (JNU20200312IRB04). The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) ethical guidelines and the CONSORT-EHEALTH checklist [22]. All researchers involved had been trained uniformly before the trial started to ensure that they were familiar with the trial procedures and methods.

Participants

All participants confirmed their willingness to participate in face-to-face screening interviews to assess their eligibility. Our health care professionals recorded the participants' relevant information, such as the type of disease, cognitive function, literacy capacity, surgery history in the past 6 months, planned residence time in the city, and mobile phone operation ability. After that, based on the inclusion and exclusion criteria, our health care professionals selected eligible participants for enrollment.

The criteria for inclusion in the study were as follows: (1) patients diagnosed with T2DM who met the 1999 World Health Organization diagnostic criteria; (2) participants with normal cognitive function who could read and write and voluntarily participated in the study; (3) no history of major surgery during the past 6 months, no major surgery plan in the next 6 months, and absence of any medication condition that could prevent the patients from walking for 15 to 30 minutes a day; (4) participants who had lived in Wuxi for more than half a year and were willing to participate in regular follow-ups; and (5) participants or family members living with them who could use WeChat proficiently, including sending and receiving messages, voice calls, and video calls.

The exclusion criteria for enrollment were as follows: (1) diagnosis of type 1 DM, gestational DM, maturity-onset diabetes of the young, or any other type of diabetes; (2) patients undergoing hemodialysis for chronic kidney disease; (3) history of any serious heart-related events (such as heart attack or stroke) in the past year; (4) pregnant patients or patients planning for a pregnancy in the next 6 months; (5) patients with disturbance of consciousness and mental disorders; and (6) patients participating in other intervention studies.

The participants received no compensation but were enrolled in the program for free. Before participating in the program, informed consent was obtained from each participant to use their data for clinical research.

Sample Size

The sample size was calculated based on a completely random design using the sample size formula for the comparison of the mean of 2 independent samples. The trial was designed for analysis using 2-tailed tests, with type I and II error rates set at 0.05 and 0.1, respectively. We used the difference in the mean HbA_{1c} (0.91%) between the intervention and control groups along with the SD (1.14 for the intervention group and 1.61 for the control group) from a study on diabetes education and SMS text messaging reminders on metabolic control and disease management in patients with T2DM [23]. The latter study was similar to our trial as both were randomized controlled trials with primary outcomes of HbA_{1c}. The calculations indicated that the total sample size required for each group was 50. Considering a dropout rate of up to 20%, the final sample size was determined to be ≥60 cases in each group.

Program

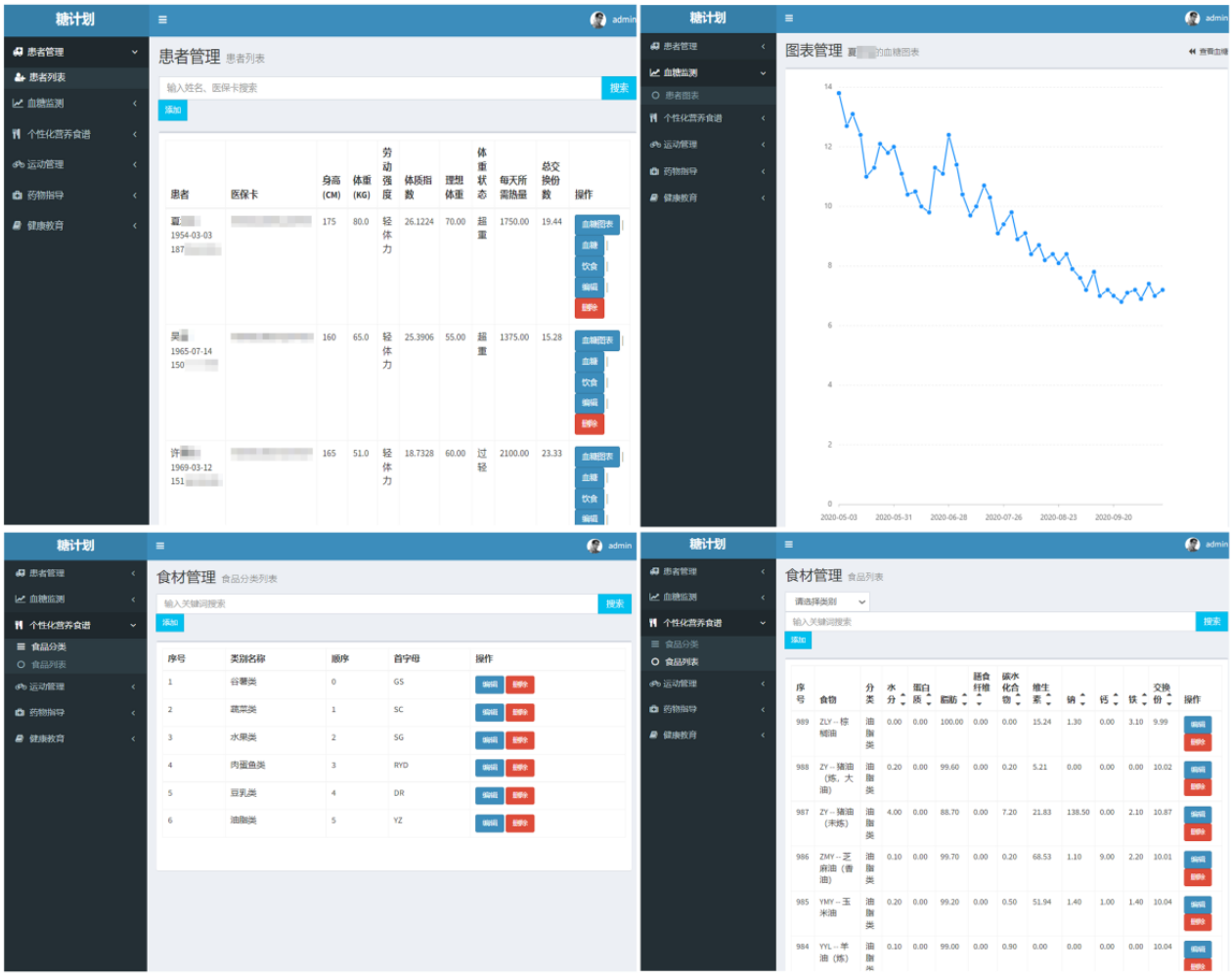
Diabetes Self-management Education Team

Our multidisciplinary team comprised health care professionals, including the general physician from the community health care center, a diabetes specialist nurse, physicians from the department of endocrinology, physicians from the department of rehabilitation, a dietitian, and trained diabetes health educators.

TangPlan Software

The diabetes management software TangPlan (Figure 1) was designed by our multidisciplinary team using the focus group method, and technical support was provided by Wuxi Wutong Leaf Technology Co, Ltd. TangPlan includes 6 functional modules (Textbox 1).

Figure 1. TangPlan website.



Textbox 1. The 6 functional modules included in TangPlan.

1. Patient list module: Each patient's information includes name, sex, date of birth, health insurance card number, mobile phone number, education level, food allergy, religious beliefs, normal exercise type and duration, body height, body weight, BMI, physical activity (intensity grades include very light, light, middle, and heavy), type 2 diabetes mellitus (T2DM) duration in years, glycated hemoglobin (HbA_{1c}), systolic blood pressure, diastolic blood pressure, fasting blood glucose (BG), 2-hour postload BG, serum triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, total cholesterol, medication (oral hypoglycemic agents, insulin, or both), and comorbidity (including hypertension, hyperlipidemia, coronary heart disease, cerebral infarction, and myocardial infarction).
2. BG monitoring module: The patient's BG level is recorded in this module, and a dynamic BG chart is generated.
3. Personalized nutrition customization module: In the food management submodule, all ingredients are divided into six categories: cereals, vegetables, fruits, beans and bean products, meat and fish, and oil. Each ingredient has an exact nutritive value, including water; protein; fat; dietary fiber; carbohydrates; and vitamins, calories, sodium, calcium, iron, glycemic index, and glycemic load. In the recipe design submodule, based on the patient's personalized information from the patient list module and BG monitoring module, the calorie requirement and the dietary recipe are automatically designed for each patient by a general physician, a physician from the department of endocrinology, or a dietitian. The software can automatically generate a weekly recipe and does not use ingredients with a higher glycemic index and load, and some ingredients can be adjusted based on the patient's willingness.
4. Physical activity module: Physical activity includes aerobic and resistance exercises. Aerobic exercises include walking, jogging, long-distance swimming, cycling, tai chi, and fitness dance. Resistance exercises include push-ups, sit-ups, squats, barbell curls, upright lifts, bend lifts, bench presses, and overhead presses. According to the participant's previous exercise experience, physicians from the department of endocrinology and rehabilitation select the appropriate physical activity and ascertain the exercise time after careful assessment.
5. Medication guidance module: According to the patient's economic and medical conditions combined with the patient's medication history, the patient's medication situation can be recorded. In addition, a physician from the department of endocrinology provides reasonable medication guidance.
6. Health education module: The T2DM-related knowledge misunderstandings of each patient are recorded so that our team members, especially the diabetes specialist nurse and trained diabetes health educators, can correct them during diabetes health education.

Allocation

After the participants were finally determined to be eligible to participate in this research, they were notified to go to the community health care center at a designated time and digitally randomized in a 1:1 allocation ratio either to the combination of TangPlan and WeChat group or usual care alone (control) group. After the results of the grouping were released, the participants were not allowed to switch groups.

Intervention and Control Groups

Diabetes Usual Care (Control Group)

Participants in the control group went to the diabetes clinic of the community health care center and received usual care, including medication adjustment, guidance on a healthy and reasonable diet, suggestions on BG self-monitoring, and physical activity.

Combination of TangPlan and WeChat Care (TangPlan and WeChat Group)

Our team members collected all the information needed of each TangPlan and WeChat group participant in the TangPlan patient list module before the program started. During the first visit of the TangPlan and WeChat participants, we added them as WeChat friends, created a WeChat group, and ensured that the participants could use it proficiently. The program coached the participants in five areas: improving BG self-monitoring, healthy eating, active physical exercise, increasing medication compliance, and health education (Figure 2).

In terms of BG self-monitoring, we asked the participants about the frequency of BG monitoring at home and evaluated the effects of medication, diet, and physical activity on BG control. We then gave recommendations on the BG monitoring time and frequencies of the participants with T2DM. We also told the participants that, once they tested their BG at home, they should send the data to us via WeChat, and then we could record it in TangPlan and provide personalized recommendations.

Next, we used TangPlan to automatically design a personalized weekly diet plan, adjusted some ingredients based on the participant's willingness, and finally printed the dietary recipe for the participant (Multimedia Appendix 2). We formulated the various ingredients that the participants needed every day, and the combination of ingredients and cooking methods was determined by the participants themselves. The participants generally thought that they could accept the dietary recipes. We encouraged the participants to keep a diet diary and bring it during follow-ups.

Physical activity was carried out under the guidance of physicians from the departments of endocrinology and rehabilitation. They performed medical assessments of cardiopulmonary and exercise function before physical activity. For general patients with T2DM, the goal was to exercise ≥5 days a week with 30 minutes of aerobic exercise each time, including tai chi and walking. If the participant had no contraindications, resistance exercise should be performed 2-3 times a week at intervals of >48 hours. The participants were asked to keep posting in the WeChat group upon completion of the exercises. If the participant's BG fluctuated greatly or had acute metabolic complications, the participant needed to wait until the condition was stabilized before gradually returning to exercise. We regularly evaluated whether the exercise program was suitable for the participant and made corresponding adjustments.

To reduce or avoid missed medications, the participants were also required to keep posting in the WeChat group after taking medications, and team members were responsible for the statistics. For participants who had not kept posting, they were notified by us on WeChat.

Figure 2. The general framework of the program.

Wed-based health education was conducted between team members and participants. First, before the program was launched, our team members evaluated the participants' knowledge of diabetes, medication use, insulin injection, BG monitoring, and diet and exercise. Second, we held T2DM self-management lectures regularly. Physicians, dietitians, diabetes specialist nurses, and trained diabetes health care educators in our team explained the T2DM-related knowledge, including pathogenesis; inducements; and the importance of regular BG monitoring, reasonable diet, medication administration, exercise, and correct BG monitoring methods. We reserved time during each lecture to encourage the participants to ask questions, and the team members answered patiently. After that, we held a T2DM health knowledge competition and encouraged the participants to interact with others. We gave rewards to the participants with outstanding performances. Finally, we conducted web-based health education through the WeChat group. We reminded the participants to take part in the interaction on time by sending messages and making phone calls, guided the participants to express their thoughts and experiences in the WeChat group, praised them for the right approach, and pointed out their mistakes. The focus was to guide the participants to realize the importance of regular medication, reasonable diet, proper exercise, and BG monitoring for glycemic control and to enhance the participants' health awareness and self-management capabilities.

The participants were asked to record self-management data, including BG, meals, physical activity, and medication administration, and they sent this information to the team via WeChat. We reviewed the participants' data daily, provided dietary recipes once a week, responded to participants' and family members' queries, and provided personalized feedback during each interaction. We also provided weekly and monthly summaries to the participants during follow-ups or through WeChat voice calls during the program.

Outcome Measures

Overview

Outcome data were collected at baseline and 6 months after the intervention began. All participants were told to go to the community health care center for follow-up at the specified follow-up time and keep a fasting state. The primary outcome measure of the study was the change in fasting BG (FBG) and HbA_{1c} levels in the control and TangPlan and WeChat groups before and after the intervention. The main secondary outcomes included changes in body weight, BMI, BP, and serum lipid profiles. The above indicators were all determined by the laboratory department of the community health care center and by research assistants who did not know the grouping.

Anthropometric Parameters

When the participants arrived at the community health care center in a fasting state, their body weight and height were measured. Height was determined to the nearest 0.5 cm using a standard height gauge. BMI was calculated as weight (kg)/height $(m)^2$.

Blood Test and BP

The nurses took venous blood samples for FBG, HbA_{1c}, and serum lipid profile measurement. After distributing breakfast and instructing the participants to eat, blood was drawn again 2 hours later to measure 2-hour postload BG. During the waiting interval of the participants, their BP was measured after ensuring that the participants had been resting for at least 30 minutes.

Statistical Analyses

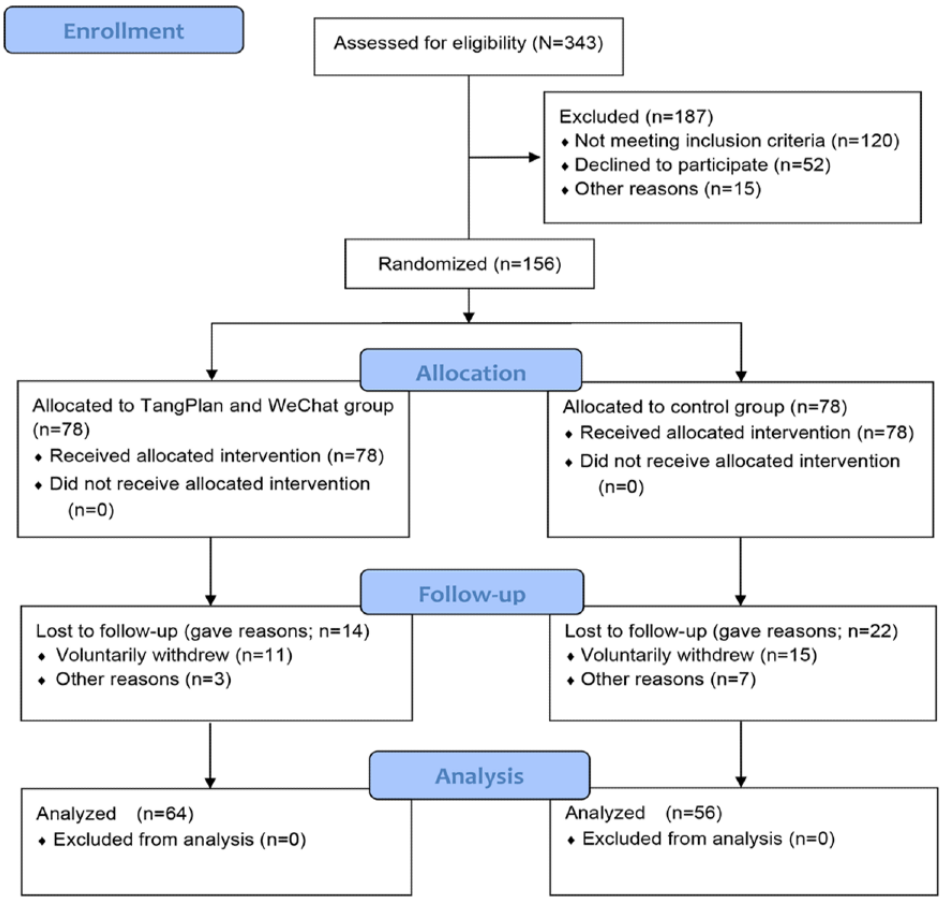
Statistical analyses were performed using SPSS (version 22.0; IBM Corp). The Kolmogorov–Smirnov test or *Q–Q* plot was used to evaluate the normal distribution. Descriptive data are presented as mean and SD for continuous variables and frequency, with proportions for categorical variables. The 2-tailed independent sample *t* test and Mann–Whitney *U* test were used to assess the differences between groups of normally and nonnormally distributed data, respectively. The paired *t* test and Wilcoxon test were used to test the differences before and after the intervention for normally and nonnormally distributed data, respectively. According to the *Guidelines for the Prevention and Treatment of Type 2 Diabetes in China* proposed by the Chinese Diabetes Society in 2017, a reasonable HbA_{1c} control target is <7%. In treatment, HbA_{1c} ≥7% can be used as an important criterion for the initiation of clinical treatment of T2DM or the need to adjust the treatment plan. Accordingly, we divided the HbA_{1c} levels of the patients into two categories: HbA_{1c} <7% (normal) and HbA_{1c} ≥7% (abnormal). The McNemar test was used to determine the impact of the intervention on the HbA_{1c} levels of the control and TangPlan and WeChat groups before and after the intervention. P<.05 was considered statistically significant.

Results

Overview

A total of 343 participants with T2DM were assessed for eligibility, of whom 187 (54.5%) were excluded (Figure 3). In total, 52 participants declined to take part because they were not interested in the program (18/52, 35%), did not want to pay much attention to diabetes (14/52, 27%), thought they did not need help (15/52, 29%), or had no reason (5/52, 10%). A total of 156 participants were randomized into the TangPlan and WeChat group (78/156, 50%) or the control group (78/156, 50%). Of these, 120 participants (TangPlan and WeChat: 64/120, 53.3%; control: 56/120, 46.7%) completed the follow-up assessments, yielding a retention rate of 76.9% (120/156).

Figure 3. Participant flow diagram. HbA_{1c}: glycated hemoglobin.



Baseline Characteristics of the Participants

The baseline characteristics of the participants are presented in Multimedia Appendix 3. Approximately 64% (36/56) male and 36% (20/56) female participants completed the trial in the control group, whereas 63% (40/64) male and 38% (24/64) female participants completed the trial in the TangPlan and WeChat group. No statistically significant differences in baseline characteristics, including age (*P*=.09), education level (*P*=.65), family monthly income (*P*=.49), T2DM duration in years (*P*=.66), body weight (*P*=.13), BMI (*P*=.10), HbA_{1c} (*P*=.84), FBG (*P*=.35), 2-hour postload BG (*P*=.36), serum lipid profiles, medication (*P*=.61), and presence of comorbidities (*P*=.76), were found between the control and TangPlan and WeChat groups.

Changes in FBG and HbA_{1c}

After 6 months, the FBG (mean 6.51, SD 1.66 mmol/L) and HbA_{1c} (mean 6.87%, SD 1.11%) levels of the TangPlan and WeChat group were both significantly lower than those of the control group (FBG: mean 7.71, SD 2.70 mmol/L; HbA_{1c}: mean 8.42%, SD 1.83%; Figure 4A and B). Compared with the baseline data, the mean change in FBG and HbA_{1c} in the control group was −0.21 mmol/L (SD 0.87 mmol/L; 95% CI −0.02 to 0.44 mmol/L; *P*=.08; Figure 4C) and −0.11% (SD 0.55%; 95% CI −0.26% to 0.04%; *P*=.15; Figure 4E), respectively, whereas the mean change in FBG and HbA_{1c} in the TangPlan and WeChat group was 0.51 mmol/L (SD 1.24 mmol/L; 95% CI −0.83 to −0.20 mmol/L; *P*=.002; Figure 4D) and −1.6% (SD 1.28%; 95% CI −1.92% to −1.28%; *P*<.001; Figure 4F), respectively.

In this trial, 23% (13/56) of participants had normal HbA_{1c} levels, and 77% (43/56) of participants had abnormal HbA_{1c} levels in the control group (Table 1). After 6 months of follow-up, there was no change in the number of abnormal and normal HbA_{1c} levels, but 15% (2/13) of the participants with normal HbA_{1c} levels demonstrated abnormal HbA_{1c} levels, and 5% (2/43) of the participants with abnormal HbA_{1c} levels had normal HbA_{1c} levels. The McNemar test revealed that there was no statistical difference in the proportion of participants with normal HbA_{1c} levels before and after the intervention (*P*=.99). In the TangPlan and WeChat group, 27% (17/64) of participants had normal HbA_{1c} levels, and 73% (47/64) of participants had abnormal HbA_{1c} levels before the intervention. After 6 months, 58% (37/64) of participants showed normal HbA_{1c} levels, and 42% (27/64) of participants had abnormal HbA_{1c} levels. Among them, 74% (20/27) of the participants with abnormal HbA_{1c} levels had normal HbA_{1c} levels. There was a significant difference in the proportion of participants with normal HbA_{1c} levels before and after the intervention (*P*<.001).

Figure 4. Comparison of fasting blood glucose and glycated hemoglobin (HbA_{1c}) levels. (A) Fasting blood glucose difference between the control and TangPlan and WeChat groups after 6 months, (B) HbA_{1c} difference between the control and TangPlan and WeChat groups after 6 months, (C) fasting blood glucose changes in the control group at baseline and 6 months, (D) fasting blood glucose changes in the TangPlan and WeChat group at baseline and 6 months, (E) HbA_{1c} changes in the control group at baseline and 6 months, and (F) HbA_{1c} changes in the TangPlan and WeChat group at baseline and 6 months. In (A) and (B), data are shown as mean and SD.

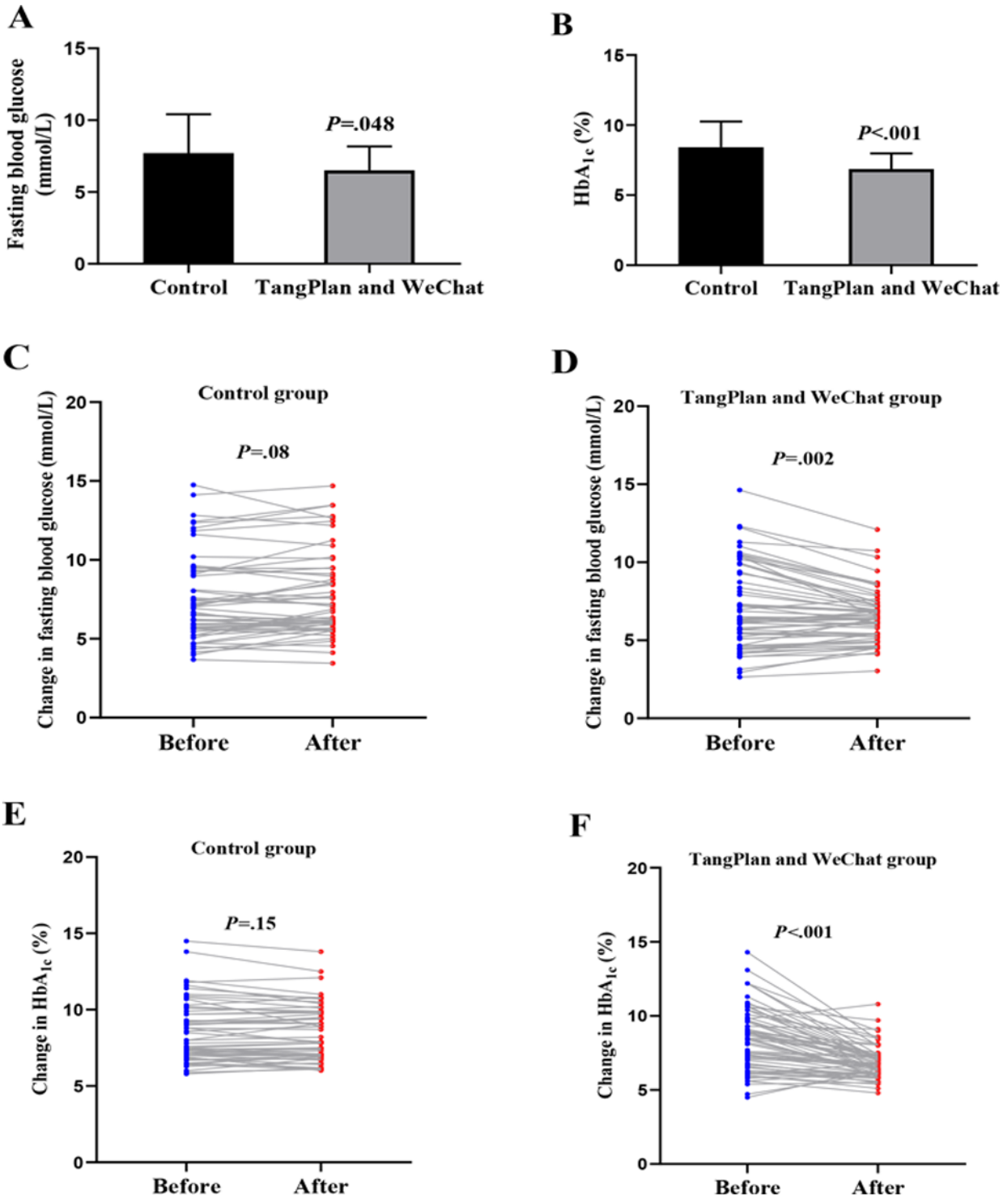


Table 1. Distribution of HbA_{1c} levels before and after the intervention (N=120).

Before the intervention	After the intervention, n (%)		*P* value
	HbA_{1c}[a]<7%	HbA_{1c}≥7%	
Control group (n=56)			.99
HbA_{1c}<7%	11 (19)	2 (4)	
HbA_{1c} ≥7%	2 (4)	41 (73)	
TangPlan and WeChat group (n=64)			<.001
HbA_{1c}<7%	17 (27)	0 (0)	
HbA_{1c} ≥7%	20 (31)	27 (42)	

[a]HbA_{1c}: glycated hemoglobin.

Changes in Body Weight, BMI, BP, and Serum Lipid Profiles

The body weight, BP, and serum lipid profiles of the participants at the end of the study are shown in Table 2. The body weight (mean 66.50, SD 9.51 kg) of the TangPlan and WeChat group was significantly lower than that of the control group (mean 71.91, SD 11.36 kg; *P*=.006), but no significant difference in BMI was observed (*P*=.10). Compared with the control group, the serum high-density lipoprotein cholesterol (HDL-C) level of the TangPlan and WeChat group (mean 1.20, SD 0.25 mmol/L) was significantly higher (*P*=.01), whereas systolic and diastolic BP (systolic BP: mean 127.03, SD 8.00 mm Hg and *P*=.005; diastolic BP: mean 75.25, SD 5.88 mm Hg and *P*=.03), serum low-density lipoprotein cholesterol (LDL-C; mean 2.50, SD 0.61 mmol/L; *P*=.006), and total cholesterol (TC; mean 4.01, SD 0.83 mmol/L; *P*=.02) levels in the TangPlan and WeChat group were significantly lower. In addition, the proportion of participants with HbA_{1c} <7% in the TangPlan and WeChat group was significantly higher than that in the control group (*P*<.001).

Table 2. Differences in body weight, BMI, blood pressure, and serum lipid profiles after 6 months (N=120).

Characteristic	Overall	Control group (n=56)	TangPlan and WeChat group (n=64)	*P* value
Body weight (kg), mean (SD)	69.02 (10.72)	71.91 (11.36)	66.50 (9.51)	.006
BMI (kg/m^2), mean (SD)	23.67 (5.32)	24.53 (5.74)	22.92 (4.86)	.10
SBP[a] (mm Hg), mean (SD)	129.80 (10.22)	132.96 (11.56)	127.03 (8.00)	.005[b]
DBP[c] (mm Hg), mean (SD)	76.52 (6.87)	77.96 (7.64)	75.25 (5.88)	.03[b]
Serum lipid profiles (mmol/L), mean (SD)				
TG[d]	1.59 (0.83)	1.78 (1.10)	1.43 (0.40)	.33[b]
HDL-C[e]	1.15 (0.26)	1.08 (0.25)	1.20 (0.25)	.01
LDL-C[f]	2.71 (0.76)	2.96 (0.84)	2.50 (0.61)	.006[b]
TC[g]	4.22 (0.95)	4.42 (1.00)	4.01 (0.83)	.02
HbA_{1c}[h] distribution, n (%)				<.001
HbA_{1c} <7%	50 (41.7)	13 (23.2)	37 (57.8)	
HbA_{1c} ≥7%	70 (58.3)	43 (76.8)	27 (42.2)	

[a]SBP: systolic blood pressure.

[b]Mann–Whitney *U* test was used.

[c]DBP: diastolic blood pressure.

[d]TG: triglycerides.

[e]HDL-C: high-density lipoprotein cholesterol.

[f]LDL-C: low-density lipoprotein cholesterol.

[g]TC: total cholesterol.

[h]HbA_{1c}: glycated hemoglobin.

The mean changes in body weight, BP, and serum lipid profiles during the 6-month follow-up are illustrated in Table 3. Compared with the baseline data, the mean changes in body weight (*P*<.001), BMI (*P*<.001), systolic BP (*P*<.001), diastolic

BP (*P*<.001), triglycerides (TG; *P*=.004), HDL-C (*P*=.001), LDL-C (*P*<.001), and TC (*P*<.001) in the TangPlan and WeChat group were significantly improved after the intervention, which was not observed in the control group.

Table 3. Changes in body weight, BMI, blood pressure, and serum lipid profiles during the 6-month follow-up (N=120).

Characteristic	Control group (n=56)		TangPlan and WeChat group (n=64)	
	Mean (SD; 95% CI)[a]	*P* value	Mean (SD; 95% CI)	*P* value
Body weight (kg)	−0.17 (0.91; −0.42 to 0.08)	.19[b]	−2.40 (2.84; −3.13 to −1.68)	<.001
BMI (kg/m^2)	−0.06 (0.33; −0.15 to 0.03)	.09[b]	−0.85 (0.99; −1.10 to −0.60)	<.001[b]
SBP[c] (mm Hg)	−0.16 (5.99; −1.76 to 1.44)	.84	−5.66 (6.87; −7.37 to −3.94)	<.001
DBP[d] (mm Hg)	−0.66 (4.36; −1.83 to 0.51)	.26	−3.38 (4.57; −4.52 to −2.33)	<.001
Serum lipid profiles (mmol/L)				
TG[e]	0.01 (0.11; −0.02 to 0.04)	.53	−0.09 (0.24; −0.16 to −0.03)	.004
HDL-C[f]	−0.02 (0.95; −0.05 to 0.01)	.16	0.05 (0.10; 0.02 to 0.07)	.001
LDL-C[g]	−0.01 (0.45; −0.13 to 0.11)	.88	−0.42 (0.46; −0.54 to −0.30)	<.001
TC[h]	−0.03 (0.48; −0.16 to 0.11)	.68	−0.47 (0.51; −0.60 to −0.34)	<.001

[a]Mean was obtained by calculating the average of the values derived from the value after the intervention minus the value before the intervention for the same individual.

[b]Wilcoxon test was used.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]TG: triglycerides.

[f]HDL-C: high-density lipoprotein cholesterol.

[g]LDL-C: low-density lipoprotein cholesterol.

[h]TC: total cholesterol.

Discussion

Principal Findings

This study assessed the effectiveness of the combination of the web-based TangPlan software and WeChat app in improving glycemic control (decrease in HbA_{1c} and BG levels) and serum lipid profiles (decline in TG, LDL-C, and TC and increase in HDL-C) as well as reducing body weight in patients with T2DM after 6 months of the program. Patients who used the TangPlan and WeChat–based intervention for 6 months reduced their HbA_{1c} levels by 1.6%, FBG levels by 0.51 mmol/L, body weight by 2.40 kg, BMI by 0.85 kg/m^2, TG levels by 0.09 mmol/L, LDL-C levels by 0.42 mmol/L, and TC levels by 0.47 mmol/L and increased their HDL-C levels by 0.05 mmol/L.

Although many well-established methods of patient care have improved the clinical profile and complications associated with diabetes, the BG control rate is still at a low level. Owing to poor health education, the proportion of patients with diabetes who have optimal glycemic control (HbA_{1c} <7%) in China is <40%, and the rate is much lower in older adults [24,25]. With the inability to achieve control of BG, diabetes symptoms, and diabetes-related comorbidities through routine follow-up and patient self-management, interventions using mobile technology may improve the treatment effects of diabetes. As an adjuvant to standard self-management, some diabetes apps lead to a clinically remarkable reduction in HbA_{1c} levels [20], whereas some apps fail to improve HbA_{1c} levels [26]. However, although many diabetes apps have been developed in China, their use rate is only 15.44%. Use is higher among patients with type 1 diabetes than among patients with T2DM. The reasons why patients discontinue the use of an app include limited time (29.9%), complicated operations (25.4%), ineffectiveness for glycemic control (24.4%), and cost (19.3%) [27]. It has been reported that older patients have increased difficulty navigating and engaging with diabetes apps and are less likely to benefit from diabetes apps than younger patients [13].

WeChat is an extremely popular social app in China, and it is also easy to operate and can offer multiple functions, including texting and voice messages, free voice and video calls, group chats, and subscribing to public accounts. The value of WeChat in chronic disease management in China lies in that it can effectively overcome current difficulties such as time conflicts, geographic distance, and economic problems. This research in Henan Province, China, showed that offering health education through the WeChat platform for patients with diabetes by sending and explaining diabetes-related knowledge improved glycemic control [28]. Patients with diabetes who intervened through WeChat could receive a better education on BG self-monitoring, reasonable diet, exercise prescription, compliance with prescribed drugs, management of hypoglycemia and hyperglycemia, and weight control through communicating with nurses on WeChat. The frequency of communication between nurses and patients was 3 to 5 times in the first week,

2 to 4 times in the second week, and only once from the third week to the end. In the sixth month, the HbA_{1c} levels of the patients in the WeChat group were lower than those with usual care [29]. Another study with 2 years of follow-up demonstrated that WeChat could contribute to the establishment of a systematic health education model. Similar to what we did in this study, patients were also encouraged to monitor their BG, communicate regularly with the health education team via WeChat, and bring a record book to appointments, leading to lower HbA_{1c} levels and increasing compliance with the control criteria [30]. However, Chinese health care professionals are busy, and it is difficult for them to spend a lot of time giving patients with diabetes timely and complete self-management guidance only through WeChat. In this study, we designed a software named TangPlan for health care professionals, which contains multiple modules for diabetes management, especially the time-saving personalized nutrition customization module. TangPlan can help health care professionals take a short time to manage patients with diabetes, including automatically generating weekly personalized dietary recipes and recording BG monitoring values and exercise. In Chinese patients with poorly controlled diabetes, it was not easy to achieve long-term effective glucose improvement using app self-management alone, but combining it with wed-based management can help achieve rapid and sustained glycemic control [29]. The combination of WeChat and TangPlan in this trial brought benefits to patients with T2DM, which is suggested through improved HbA_{1c} levels, FBG levels, BP, and lipid profiles. However, we also need to enhance the interaction with patients to ensure timely tracking and feedback of the patients' data.

The effectiveness of the intervention was evaluated through a significant reduction in HbA_{1c} levels. Studies have shown that a 0.5% to 1% reduction in HbA_{1c} levels is considered clinically significant and can reduce the risk of complications [31]. Even the Food and Drug Administration in the United States requires a 0.4% decline in HbA_{1c} levels for drug evaluation [32]. The results of the UK Prospective Diabetes Study showed that a 0.9% reduction in HbA_{1c} levels was related to a 25% decrease in microvascular complications, a 10% decrease in diabetes-related mortality, and a 6% reduction in all-cause mortality [33,34]. According to the evidence-based practice guidelines for diabetes in the United States, only 37% of persons with diabetes meet the HbA_{1c} target of <7%, and only 7% meet the combined glycemic, lipid, and BP goals [35]. We found that the combination of TangPlan and WeChat increased the proportion of patients with diabetes who met the HbA_{1c} target in the TangPlan and WeChat group from 27% (17/64) to 58% (37/64) after the intervention. This did not change in the control group. We believe that the average decline of 1.6% in HbA_{1c} levels in the TangPlan and WeChat group was significant in reducing the risk of diabetes-related complications and mortality.

FBG is another indicator of glycemic control and correlates with HbA_{1c} levels. A study showed that FBG levels >5.6 mmol/L but not of 3.9 to 5.6 mmol/L were associated with death [36]. Evidence indicates that a chronic hyperglycemic state is associated with impaired immunity [37], and FBG levels ≥7.0 mmol/L at admission are an independent predictor for 28-day mortality in patients with COVID-19 without a previous diagnosis of diabetes [38]. Hyperglycemia is also a risk factor for cardiovascular disease in T2DM [39]. The combination of TangPlan and WeChat remarkably decreased FBG levels, which indicated an additional benefit of reduction in cardiovascular risk among these patients.

Weight loss is one of the goals in the management of diabetes and is associated with improvements in HbA_{1c} levels [40]. The prospective Swedish Obese Subjects study revealed that weight reduction through gastric surgery performed on patients with obesity had a dramatic effect on the 8-year incidence of diabetes but no effect on the 8-year incidence of hypertension [41]. Each 5 kg/m^2 decrease in BMI will prevent approximately 30% overall mortality in the population with diabetes [42]. The significant decrease in weight (2.40 kg) and BMI (0.85 kg/m^2) among the TangPlan and WeChat group after 6 months of the program highlights the importance of weight reduction in ameliorating HbA_{1c} levels.

A variety of studies suggest that glucose alone is not responsible for diabetic complications, especially in individuals with T2DM [43,44]. Rather, the responsibility lies in a cluster of factors, including dyslipidemia, obesity, and hypertension, which have an impact on the adipose tissue fatty acid metabolism that underlines the onset and progression of diabetic complications [45]. The combination of TangPlan and WeChat exerted a beneficial effect on serum lipid profiles, illustrated by TG, TC, and LDL-C reduction and HDL-C elevation, further confirming the possibility of the TangPlan and WeChat intervention to reduce the incidence of diabetic complications.

To our knowledge, this study is one of the first to report the effectiveness of the combination of a self-designed diabetes management software and WeChat on glycemic control and other metabolic indicators. During the intervention process, whether it was during the follow-up, diabetes education lectures, or web-based communication via WeChat, there were frequent interactions between health care professionals and patients with diabetes, which might help promote a more harmonious relationship. However, one of the limitations of this trial was that the occurrence and progression of diabetes complications were not evaluated after the intervention. In addition, the program was performed for a short duration, and we did not independently quantify the influence of other behaviors and lifestyles on glycemic control. The loss of data during follow-up also limited the scope of this study. Future studies with a larger sample and better control will be able to further determine the effectiveness of the program.

Conclusions

The combination of TangPlan and WeChat demonstrated an incremental decline in HbA_{1c} and FBG levels, body weight, and BP as well as improvements in serum lipid profiles during the 6-month program, indicating the feasibility, acceptance, and value of using a novel method for diabetes management. The results of this study can be further explored to assess the long-term acceptability, cost-effectiveness, and durability of the principal findings as well as the feasibility of the program in a larger population.

Acknowledgments

This study was supported by the Wuxi Science and Technology Development Fund (grant CSE31N1625). The authors would like to thank all the participants in the program and Yu-Qiu Ge for the statistical analysis provided.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1325 KB - mhealth_v10i3e30571_app1.pdf]

Multimedia Appendix 2
Weekly recipe of a participant (providing 1750 kcal).
[PNG File , 3541 KB - mhealth_v10i3e30571_app2.png]

Multimedia Appendix 3
Participant baseline characteristics at the start of the program.
[DOCX File , 21 KB - mhealth_v10i3e30571_app3.docx]

References

1. IDF diabetes atlas. 8th edition. International Diabetes Federation. 2017. URL: https://diabetesatlas.org/atlas/eighth-edition/ [accessed 2021-04-19]
2. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. JAMA 2017;317(24):2515-2523 [FREE Full text] [doi: 10.1001/jama.2017.7596] [Medline: 28655017]
3. National Commission of Health. Chinese national statistical yearbook of health. Beijing, China: Peking Union Medical College Press; 2018.
4. Ryan EA, Holland J, Stroulia E, Bazelli B, Babwik SA, Li H, et al. Improved A1C levels in type 1 diabetes with smartphone app use. Can J Diabetes 2017;41(1):33-40. [doi: 10.1016/j.jcjd.2016.06.001] [Medline: 27570203]
5. Devries JH, Snoek FJ, Heine RJ. Persistent poor glycaemic control in adult type 1 diabetes. A closer look at the problem. Diabet Med 2004;21(12):1263-1268. [doi: 10.1111/j.1464-5491.2004.01386.x] [Medline: 15569126]
6. Blondon K, Klasnja P, Coleman K, Pratt W. An exploration of attitudes toward the use of patient incentives to support diabetes self-management. Psychol Health 2014;29(5):552-563. [doi: 10.1080/08870446.2013.867346] [Medline: 24256342]
7. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
8. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
9. Garg SK, Shah VN, Akturk HK, Beatson C, Snell-Bergeon JK. Role of mobile technology to improve diabetes care in adults with type 1 diabetes: the remote-T1D study iBGStar in type 1 diabetes management. Diabetes Ther 2017;8(4):811-819 [FREE Full text] [doi: 10.1007/s13300-017-0272-5] [Medline: 28555339]
10. Hofer TP, Zemencuk JK, Hayward RA. When there is too much to do: how practicing physicians prioritize among recommended interventions. J Gen Intern Med 2004;19(6):646-653 [FREE Full text] [doi: 10.1007/s11606-004-0058-0] [Medline: 15209603]
11. Harris MI. Health care and health status and outcomes for patients with type 2 diabetes. Diabetes Care 2000;23(6):754-758. [doi: 10.2337/diacare.23.6.754] [Medline: 10840991]
12. Given JE, O'Kane MJ, Bunting BP, Coates VE. Comparing patient-generated blood glucose diary records with meter memory in diabetes: a systematic review. Diabet Med 2013;30(8):901-913. [doi: 10.1111/dme.12130] [Medline: 23324062]
13. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
14. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Correction: mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2018;6(1):e20 [FREE Full text] [doi: 10.2196/mhealth.8789] [Medline: 29334479]

15. Kebede MM, Pischke CR. Corrigendum: popular diabetes apps and the impact of diabetes app use on self-care behaviour: a survey among the digital community of persons with diabetes on social media. Front Endocrinol (Lausanne) 2019;10:220 [FREE Full text] [doi: 10.3389/fendo.2019.00220] [Medline: 31024457]
16. Chan JC, Zhang Y, Ning G. Diabetes in China: a societal solution for a personal challenge. Lancet Diabetes Endocrinol 2014;2(12):969-979. [doi: 10.1016/S2213-8587(14)70144-5] [Medline: 25218728]
17. Trawley S, Baptista S, Browne JL, Pouwer F, Speight J. The use of mobile applications among adults with type 1 and type 2 diabetes: results from the second MILES-Australia (MILES-2) study. Diabetes Technol Ther 2017;19(12):730-738. [doi: 10.1089/dia.2017.0235] [Medline: 29028442]
18. Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. Comput Methods Programs Biomed 2020;196:105710 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105710] [Medline: 32858284]
19. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
20. Shi B, Liu X, Dong Q, Yang Y, Cai Z, Wang H, et al. The effect of a WeChat-based tertiary A-level hospital intervention on medication adherence and risk factor control in patients with stable coronary artery disease: multicenter prospective study. JMIR Mhealth Uhealth 2021;9(10):e32548 [FREE Full text] [doi: 10.2196/32548] [Medline: 34569467]
21. Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J, et al. Evaluating an intervention program using WeChat for patients with chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2020;22(4):e17089 [FREE Full text] [doi: 10.2196/17089] [Medline: 32314971]
22. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
23. Güner TA, Coşansu G. The effect of diabetes education and short message service reminders on metabolic control and disease management in patients with type 2 diabetes mellitus. Prim Care Diabetes 2020;14(5):482-487. [doi: 10.1016/j.pcd.2020.04.007] [Medline: 32471770]
24. Chen R, Ji L, Chen L, Chen L, Cai D, Feng B, et al. Glycemic control rate of T2DM outpatients in China: a multi-center survey. Med Sci Monit 2015;21:1440-1446 [FREE Full text] [doi: 10.12659/MSM.892246] [Medline: 25986070]
25. Ji LN, Lu JM, Guo XH, Yang WY, Weng JP, Jia WP, et al. Glycemic control among patients in China with type 2 diabetes mellitus receiving oral drugs or injectables. BMC Public Health 2013;13:602 [FREE Full text] [doi: 10.1186/1471-2458-13-602] [Medline: 23800082]
26. Castensøe-Seidenfaden P, Husted GR, Jensen AK, Hommel E, Olsen B, Pedersen-Bjergaard U, et al. Testing a smartphone app (Young with Diabetes) to improve self-management of diabetes over 12 months: randomized controlled trial. JMIR Mhealth Uhealth 2018;6(6):e141 [FREE Full text] [doi: 10.2196/mhealth.9487] [Medline: 29945861]
27. Zhang Y, Li X, Luo S, Liu C, Xie Y, Guo J, et al. Use, perspectives, and attitudes regarding diabetes management mobile apps among diabetes patients and diabetologists in China: national web-based survey. JMIR Mhealth Uhealth 2019;7(2):e12658 [FREE Full text] [doi: 10.2196/12658] [Medline: 30735147]
28. Zhang Y, Chu L. Effectiveness of systematic health education model for type 2 diabetes patients. Int J Endocrinol 2018;2018:6530607 [FREE Full text] [doi: 10.1155/2018/6530607] [Medline: 30147724]
29. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app-based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
30. Dong Y, Wang P, Dai Z, Liu K, Jin Y, Li A, et al. Increased self-care activities and glycemic control rate in relation to health education via WeChat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018;97(50):e13632 [FREE Full text] [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
31. Nathan DM, Buse JB, Davidson MB, Heine RJ, Holman RR, Sherwin R, Professional Practice Committee, American Diabetes Association, European Association for the Study of Diabetes. Management of hyperglycaemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy. A consensus statement from the American Diabetes Association and the European Association for the Study of Diabetes. Diabetologia 2006;49(8):1711-1721. [doi: 10.1007/s00125-006-0316-2] [Medline: 16802130]
32. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2020. Diabetes Care 2020;43(Suppl 1):S66-S76. [doi: 10.2337/dc20-S006] [Medline: 31862749]
33. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
34. Katulanda P, Jayawardena R, Jayawardana R, Ranasinghe P, Rezvi Sheriff MH, Matthews DR. Physical activity patterns and correlates among adults from a developing country: the Sri Lanka Diabetes and Cardiovascular Study. Public Health Nutr 2013;16(9):1684-1692. [doi: 10.1017/S1368980012003990] [Medline: 22995708]
35. American Diabetes Association. Economic costs of diabetes in the U.S. In 2007. Diabetes Care 2008;31(3):596-615. [doi: 10.2337/dc08-9017] [Medline: 18308683]

36. Rao Kondapally Seshasai S, Kaptoge S, Thompson A, Di Angelantonio E, Gao P, Sarwar N, Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. N Engl J Med 2011;364(9):829-841 [FREE Full text] [doi: 10.1056/NEJMoa1008862] [Medline: 21366474]
37. Pearson-Stuttard J, Blundell S, Harris T, Cook DG, Critchley J. Diabetes and infection: assessing the association with glycaemic control in population-based studies. Lancet Diabetes Endocrinol 2016;4(2):148-158. [doi: 10.1016/S2213-8587(15)00379-4] [Medline: 26656292]
38. Wang S, Ma P, Zhang S, Song S, Wang Z, Ma Y, et al. Fasting blood glucose at admission is an independent predictor for 28-day mortality in patients with COVID-19 without previous diagnosis of diabetes: a multi-centre retrospective study. Diabetologia 2020;63(10):2102-2111 [FREE Full text] [doi: 10.1007/s00125-020-05209-1] [Medline: 32647915]
39. Haffner SJ, Cassells H. Hyperglycemia as a cardiovascular risk factor. Am J Med 2003;115 Suppl 8A:6S-11S. [doi: 10.1016/j.amjmed.2003.09.009] [Medline: 14678859]
40. Gummesson A, Nyman E, Knutsson M, Karpefors M. Effect of weight reduction on glycated haemoglobin in weight loss trials in patients with type 2 diabetes. Diabetes Obes Metab 2017;19(9):1295-1305. [doi: 10.1111/dom.12971] [Medline: 28417575]
41. Sjöström CD, Peltonen M, Wedel H, Sjöström L. Differentiated long-term effects of intentional weight loss on diabetes and hypertension. Hypertension 2000;36(1):20-25. [doi: 10.1161/01.hyp.36.1.20] [Medline: 10904007]
42. Cheng Y, Zhang H, Chen R, Yang F, Li W, Chen L, et al. Cardiometabolic risk profiles associated with chronic complications in overweight and obese type 2 diabetes patients in South China. PLoS One 2014;9(7):e101289 [FREE Full text] [doi: 10.1371/journal.pone.0101289] [Medline: 24992024]
43. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, VADT Investigators. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med 2009;360(2):129-139. [doi: 10.1056/NEJMoa0808431] [Medline: 19092145]
44. Zoungas S, Arima H, Gerstein HC, Holman RR, Woodward M, Reaven P, Collaborators on Trials of Lowering Glucose (CONTROL) group. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. Lancet Diabetes Endocrinol 2017;5(6):431-437. [doi: 10.1016/S2213-8587(17)30104-3] [Medline: 28365411]
45. Callaghan BC, Xia R, Banerjee M, de Rekeneire N, Harris TB, Newman AB, Health ABC Study. Metabolic syndrome components are associated with symptomatic polyneuropathy independent of glycemic status. Diabetes Care 2016;39(5):801-807 [FREE Full text] [doi: 10.2337/dc16-0081] [Medline: 26965720]

Abbreviations

BG: blood glucose
BP: blood pressure
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DM: diabetes mellitus
FBG: fasting blood glucose
HbA_{1c}: glycated hemoglobin
HDL-C: high-density lipoprotein cholesterol
LDL-C: low-density lipoprotein cholesterol
T2DM: type 2 diabetes mellitus
TC: total cholesterol
TG: triglycerides

Edited by L Buis, A Mavragani; submitted 21.05.21; peer-reviewed by G Dermody; comments to author 02.11.21; revised version received 20.12.21; accepted 20.02.22; published 30.03.22.

Please cite as:
Xia SF, Maitiniyazi G, Chen Y, Wu XY, Zhang Y, Zhang XY, Li ZY, Liu Y, Qiu YY, Wang J
Web-Based TangPlan and WeChat Combination to Support Self-management for Patients With Type 2 Diabetes: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(3):e30571
URL: https://mhealth.jmir.org/2022/3/e30571
doi:10.2196/30571
PMID:35353055

Original Paper

Use of an mHealth Ketogenic Diet App Intervention and User Behaviors Associated With Weight Loss in Adults With Overweight or Obesity: Secondary Analysis of a Randomized Clinical Trial

Kaja Falkenhain[1], BSc; Sean R Locke[2], PhD; Dylan A Lowe[3], PhD; Terry Lee[4], PhD; Joel Singer[4], PhD; Ethan J Weiss[5], MD; Jonathan P Little[1], PhD

[1]School of Health and Exercise Sciences, The University of British Columbia Okanagan, Kelowna, BC, Canada

[2]Department of Kinesiology, Brock University, St. Catherines, ON, Canada

[3]Eureka Research Platform, San Francisco, CA, United States

[4]Centre for Health Evaluation and Outcome Sciences, School of Population and Public Health, The University of British Columbia, Vancouver, BC, Canada

[5]Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Jonathan P Little, PhD
School of Health and Exercise Sciences
The University of British Columbia Okanagan
1147 Research Road
Kelowna, BC, V1V 1V7
Canada
Phone: 1 250 807 9876
Email: jonathan.little@ubc.ca

Abstract

Background: Low-carbohydrate ketogenic diets are a viable method to lose weight that have regained popularity in recent years. Technology in the form of mobile health (mHealth) apps allows for scalable and remote delivery of such dietary interventions and are increasingly being used by the general population without direct medical supervision. However, it is currently unknown which factors related to app use and user behavior are associated with successful weight loss.

Objective: First, to describe and characterize user behavior, we aim to examine characteristics and user behaviors over time of participants who were enrolled in a remotely delivered clinical weight loss trial that tested an mHealth ketogenic diet app paired with a breath acetone biofeedback device. Second, to identify variables of importance to weight loss at 12 weeks that may offer insight for future development of dietary mHealth interventions, we aim to explore which app- and adherence-related user behaviors characterized successful weight loss.

Methods: We analyzed app use and self-reported questionnaire data from 75 adults with overweight or obesity who participated in the intervention arm of a previous weight loss study. We examined data patterns over time through linear mixed models and performed correlation, linear regression, and causal mediation analyses to characterize diet-, weight-, and app-related user behavior associated with weight loss.

Results: In the context of a low-carbohydrate ketogenic diet intervention delivered remotely through an mHealth app paired with a breath acetone biofeedback device, self-reported dietary adherence seemed to be the most important factor to predict weight loss (β=–.31; t_{54}=–2.366; *P*=.02). Furthermore, self-reported adherence mediated the relationship between greater app engagement (from c=–0.008, 95% CI –0.014 to –0.0019 to c'=–0.0035, 95% CI –0.0094 to 0.0024) or higher breath acetone levels (from c=–1.34, 95% CI –2.28 to –0.40 to c'=–0.40, 95% CI –1.42 to 0.62) and greater weight loss, explaining a total of 27.8% and 28.8% of the variance in weight loss, respectively. User behavior (compliance with weight measurements and app engagement) and adherence-related aspects (breath acetone values and self-reported dietary adherence) over time differed between individuals who achieved a clinically significant weight loss of >5% and those who did not.

Conclusions: Our in-depth examination of app- and adherence-related user behaviors offers insight into factors associated with successful weight loss in the context of mHealth interventions. In particular, our finding that self-reported dietary adherence was the most important metric predicting weight loss may aid in the development of future mHealth dietary interventions.

Trial Registration: ClinicalTrials.gov NCT04165707; https://clinicaltrials.gov/ct2/show/NCT04165707
International Registered Report Identifier (IRRID): RR2-10.2196/19053

(JMIR Mhealth Uhealth 2022;10(3):e33940) doi:10.2196/33940

KEYWORDS

acetone; biofeedback; psychology; diet; ketogenic; mobile apps; overweight; technology; telemedicine; weight loss; mobile phone

Introduction

Background

Overweight and obesity are strongly associated with a number of chronic health conditions, including cardiovascular disease and type 2 diabetes [1]. Approximately 39% and 13% of adults worldwide are living with overweight and obesity, respectively [2], and it is estimated that 40% of adults living in the United States are attempting a weight loss diet every year [3]. Although various diets can be used to achieve a reduction in body weight, successful weight loss requires a sustained decrease in caloric intake that necessitates behavior modification for long-term dietary adherence. For example, self-monitoring of dietary intake and body weight has been shown to aid in achieving weight loss [4]. Mobile technology, such as smartphone mobile health (mHealth) apps, now offers a platform for innovative and highly scalable interventions to be delivered to a broad audience. Apps can offer dietary guidance and assist in implementing behavior changes; for example, through interaction-enabled (eg, tracking) features, notification reminders, and educational content.

A low-carbohydrate, high-fat ketogenic diet is a popular weight loss diet that has robust impacts on metabolism [5]. A ketogenic diet aims to restrict carbohydrate intake to enable the liver to produce ketone bodies (acetoacetate, acetone, and beta-hydroxybutyrate) from free fatty acids, which can be used as an alternative fuel source and which give the diet its name. Although such diets have been shown to be a viable method to successfully lose weight [6], they require knowledge on the part of the dieter related to the macronutrient composition of foods to restrict carbohydrate intake appropriately to achieve ketosis. The diet's complex nature may present a burden of entry to people trying to lose weight by means of a ketogenic diet because it can be difficult for people wishing to lose weight to know which foods to eat to maintain a state of ketosis.

Keyto App

Keyto Inc (*Keyto*) is a company offering an app that provides a Mediterranean-based ketogenic diet intervention. With in-app resources such as recipes, meal plans, and informative articles, the program guides users to follow a low-carbohydrate diet with a focus on fats that fit the Mediterranean guidelines [7] to promote ketosis while counteracting detrimental blood lipid changes known to commonly arise with ketogenic diets high in saturated fat [8]. The app is paired with a hand-held device that measures acetone levels in the breath and serves as a noninvasive biofeedback measure of ketosis, while allowing for self-monitoring of dietary compliance and providing insight into how different foods affect fat burning and ketosis. In a previous randomized clinical trial, the app was shown to be effective for inducing weight loss and improving cardiometabolic health (ie, markers of glycemic control and liver damage) without worsening blood lipid profiles [9]. However, it is currently unknown how adults who are seeking to lose weight engage with the app and what might predict success of such a weight loss intervention. The purpose of this study is therefore to examine user behavior of adults using the *Keyto* weight loss app in a real-world setting and to identify intervention-specific (ie, app use– and adherence-related) behaviors that would predict successful weight loss.

Methods

Design

This study is a secondary analysis of app use and outcome data from the intervention group of a previous mHealth-based randomized clinical trial [9]. The original trial examined weight loss and cardiometabolic risk between participants receiving the *Keyto* app paired with a breath acetone biofeedback device and those receiving WW International Inc's WW app (formerly Weight Watchers International Inc is now WW International Inc) as an active comparator group over the course of 12 (primary end point) and 24 (secondary end point) weeks, as per registration on ClinicalTrials.gov (NCT04165707) and the published protocol (DERR1-10.2196/19053) [10]. The trial was conducted remotely out of Canada with participating individuals living in California. This paper aims to (1) examine characteristics and behaviors of users in this trial as they relate to dietary adherence and app use and (2) identify predictors of weight loss.

Ethics Approval

This study was approved by the University of British Columbia's clinical research ethics board (H19-01341) and all participants provided written informed consent digitally prior to data collection.

Participant Recruitment and Study Flow

As described previously [9,10], participants were recruited through web-based advertisements and an email listserv. Interested participants completed a web-based questionnaire to determine eligibility and, if deemed eligible, scheduled a phone call with a research team member to confirm eligibility and clarify any remaining questions. After the phone call, participants provided informed consent and, if randomized to the *Keyto* intervention, downloaded the app using a study-provided username and password to access the intervention and ensure anonymity. In addition, participants were sent a Bluetooth scale (iHealth Lina) that transmitted weight data to a cloud-based server to be accessed by the research team. The primary intervention phase was 12 weeks, with a secondary end point at 24 weeks, during which

participants were asked to follow the dietary intervention based on guidance through the app.

Mobile Weight Loss Intervention

The mobile intervention program was delivered entirely remotely through the app and without in-person interaction with the research team. The app provided to the intervention group was developed by a multidisciplinary team, including a cardiologist, an engineer, and a physician, and is commercially available. Details of the intervention have been previously described [9,10], but briefly, users are encouraged to follow a low-carbohydrate ketogenic diet with a focus on consuming fats from plant- (eg, olive oil and avocado) and fish-based (eg, salmon) sources that fit the Mediterranean diet pattern [7]. Users are not required to track their food intake but are encouraged to stick to recommended portion sizes and eat until satiety. In-app articles provide meal ideas, suggested shopping lists, advice to avoid common pitfalls, and background information about the ketogenic diet. Users can also join support groups to engage with other users.

An accompanying biofeedback device that pairs with the app measures acetone in the breath as a biomarker of ketosis [11]. The pen-sized device contains a nanostructured gas sensor with a semiconducting metal oxide core selective to acetone and each sensor is individually stabilized and calibrated during the production process. Upon breathing into the device, users receive a *Keyto* Level ranging from 0 (lowest) to 6+ (highest) as an indicator of the degree of ketosis and as a surrogate for fat loss [12]. In case of a lower score (0-3), participants are instructed to further restrict carbohydrate intake and prioritize high-fat foods, whereas in case of a higher score (≥4), participants are encouraged to continue with their current dietary habits. Participants in this trial were asked to use the accompanying breath acetone biofeedback device 3 times daily (first thing in the morning and before lunch and dinner).

Measures

Weight Loss Data

Participants were asked to weigh themselves daily on the study-provided Bluetooth scale. As described previously [9,10], baseline weight was considered the first weight measurement on the start day of the trial or, if no weight was recorded on that date, the weight measurement closest to 8 AM of the start date. The follow-up weight measurement was calculated as the average (mean) weight recorded across the final (ie, 12th and 24th) week of the intervention period to minimize the influence of daily weight fluctuations. Change in body weight and percentage of baseline body weight lost were calculated daily, at the primary end point after 12 weeks, and at the secondary end point after 24 weeks.

Adherence

Adherence to the intervention was quantified through the following metrics: (1) compliance with daily weight measurements expressed as the number of days with a weight measurement divided by the total number of days across the intervention period and (2) self-reported dietary adherence assessed weekly through a web-based survey as the response to the question: “To what extent were you able to stick to your diet in the past week?” on a 5-point Likert scale ranging from 0 (not at all) to 4 (completely). Anonymized scale data and questionnaire responses were exported at the end of the trial and summed at weekly intervals to generate adherence metrics for each participant across the intervention period.

Engagement With Dietary Intervention

Further to the aforementioned adherence metrics, engagement with the dietary intervention was assessed through (1) *Keyto* Levels obtained through use of the breath acetone biofeedback device and (2) the number of engagements with the *Keyto* app.

Questionnaires

Participants were sent a baseline questionnaire upon enrollment in the trial and weekly and monthly questionnaires throughout the intervention period. The baseline questionnaire assessed socioeconomic demographics. Weekly questionnaires assessed self-reported adherence and asked about cravings, mood, and energy (Multimedia Appendix 1). These questionnaires were designed for the purpose of this study by the research team in collaboration with the *Keyto* medical director to provide simple measures of manipulation fidelity and self-reported dietary adherence. Monthly questionnaires collected dietary intake data through the Automated Self-Administered 24-Hour Dietary Recall tool [13]. At the end of the primary intervention phase at 12 weeks, participants were asked about the impact of the COVID-19 pandemic on their dietary habits. This questionnaire was self-designed because at the time of study development, COVID-19 had not been anticipated.

Statistical Analysis

All statistical analyses were performed using R software (version 3.6.2; R Foundation for Statistical Computing), except for the mediation analysis, which was performed using SPSS software (version 25.0; IBM Corp).

Our first aim is to describe characteristics and adherence- and app use–related behaviors of users in this trial across the entire primary (12 weeks) and secondary (24 weeks) intervention periods. Baseline characteristics and questionnaire responses were summarized as mean (SD) for continuous data and n (%) for categorical data, unless otherwise stated. To investigate time-based data patterns of weight loss, participants were divided into 3 groups according to percentage baseline weight lost at 12 weeks (ie, <5%, >5% to <10%, and >10%).

Our second aim is to identify predictors of weight loss. Because of the complexity of the involved analyses, the greater data availability for the primary 12-week intervention period, and the primary outcome of the trial being weight loss at 12 weeks, we focused our investigation on the predictors of weight loss across 12 weeks. To explore data patterns over time of the adherence metrics, participants were divided into 2 groups according to whether or not clinically significant weight loss (defined as >5% of baseline weight) [14] was achieved at 12 weeks. Z-scores for each predictor variable over time were calculated as the mean of the sample subtracted from the observed value divided by the SD of the sample for that variable at each time point. A linear mixed model was used to evaluate

the effect of successful group membership (ie, whether or not the participant achieved >5% weight loss) and the interaction of the effect over time on adherence metrics; the model included time (week and month), success (>5% weight loss or <5% weight loss), and the interaction of time and success as fixed factors and participant as a random factor. Effect estimates (ie, estimated difference in z-scores) for simple effects of success group (in the case of a significant effect of success) or differences in slopes of time for success averaged over the entire intervention period (in the case of a significant interaction effect) with a 95% CI were calculated.

Potential predictor variables (ie, variables of potential relevance in predicting weight loss) related to app use and dietary adherence were explored in a pairwise Spearman rank correlation matrix and included in a multiple linear regression model. Group means of weight loss were compared across tertiles of significant predictor variables of interest using a 1-way analysis of variance. All assumptions were met, and a post hoc Tukey Honestly Significant Difference test was conducted to correct for multiple comparisons.

Similarly, potential predictors of weight loss were further explored through a mediation analysis between app-specific features (ie, average *Keyto* Level and total number of app engagements across the 12-week intervention period) and weight loss to explore causal mechanisms and investigate potential explanatory pathways. The PROCESS macro (version 3.3) for SPSS was used to estimate the hypothesized mediating effects of weight loss using direct and indirect effects based on 5000 bootstrapping samples at 95% bias-corrected CIs [15]; CIs that do not cross zero (either all positive or negative) suggest that the true effect is not zero and that the null hypothesis can be rejected.

Results

Participants

A total of 77 participants were randomized to the ketogenic diet mobile app group within the larger remote randomized clinical trial, as previously described [9]. Of the 77 participants, 1 (1%) was deemed ineligible after enrollment and 1 (1%) did not start; during the primary 12-week intervention phase, 3 (4%) discontinued the intervention, 10 (13%) were lost to follow-up, and 2 (3%) did not log a follow-up weight. The remaining participants (n=60) were included in weight-related analyses for the primary intervention phase, and partial data from all participants was used for any other analyses as available. At the secondary 24-week end point, weight data from 55% (42/77) of the participants were available. Available data from all participants were included in analyses related to adherence and user behavior. Self-reported baseline characteristics of participants assessed through a questionnaire before the intervention are presented in Table 1.

Table 1. Baseline characteristics of participants (N=75).

Characteristics	Total	Female	Male	Nonbinary
Participants, n (%)	75 (100)	53 (71)	20 (27)	2 (2)
Age (years), mean (SD)	42 (11)	41 (10)	43 (13)	33 (9)
Weight (kg), mean (SD)	94.7 (17.1)	90.7 (16.7)	104.2 (14.3)	106.1 (18.8)
BMI, mean (SD)	33.5 (4.7)	33.5 (4.9)	33.1 (4.5)	34.4 (4.7)

Change in Body Weight

Mean weight loss at 12 weeks was –5.6 (SD 4.5) kg, which equated to –5.8% (SD 4.5%) of initial body weight. Of the participants logging a follow-up weight, 53% (32/60) lost >5% and 18% (11/60) lost >10% of baseline body weight (Table 2).

Table 2. Weight loss outcomes.

Outcome measure	Total	Female	Male	Nonbinary
Primary intervention end point at 12 weeks (N=60)				
Participants, n (%)	60 (100)	42 (70)	16 (27)	2 (3)
Change in body weight (kg), mean (SD)	–5.6 (4.5)	–5.1 (4.1)	–6.9 (5.3)	–7.1 (3.0)
Change in body weight (%BBW[a]), mean (SD)	–5.8 (4.5)	–5.5 (4.7)	–6.5 (4.4)	–6.5 (1.6)
Change in BMI, mean (SD)	–1.9 (1.5)	–1.9 (1.5)	–2.1 (1.6)	–2.3 (0.9)
Lost >5% initial body weight, n (%)	32 (53)	21 (50)	9 (56)	2 (100)
Lost >10% initial body weight, n (%)	11 (18)	6 (14)	5 (31)	0 (0)
Secondary intervention end point at 24 weeks (N=42)				
Participants, n (%)	42 (100)	30 (71)	11 (26)	1 (3)
Change in body weight (kg), mean (SD)	–8.5 (6.4)	–8.3 (6.4)	–9.1 (7.1)	–9.6 (—[b])
Change in body weight (%BBW), mean (SD)	–8.7 (6.9)	–8.9 (7.5)	–8.4 (5.6)	–8.1 (—)
Change in BMI, mean (SD)	–3.0 (2.3)	–3.0 (2.5)	–2.7 (2.0)	–3.0 (—)
Lost >5% initial body weight, n (%)	24 (57)	18 (60)	5 (45)	1 (100)
Lost >10% initial body weight, n (%)	15 (36)	11 (37)	4 (36)	0 (0)

[a]%BBW: percentage of baseline body weight.

[b]Not available.

Daily change in body weight (expressed as a percentage of baseline body weight) across the 12 weeks separated by weight loss groups (ie, weight loss of <5%, >5% to <10%, and >10%) is shown in Figure 1. Body weight decreased over time in all groups; however, the slope of the decrease was steepest (ie, the change was greatest) in the group of participants who lost >10% of their initial body weight and the magnitude of change was smallest in the group of participants who lost <5% of their starting weight, suggesting overall consistency of weight loss success across the intervention period. Mean weight loss at the secondary 24-week time point was –8.5 (SD 6.4) kg or –8.7% (SD 6.9%) of initial body weight, indicating durability of the observed decrease in body weight (Table 2).

Figure 1. Individual change in body weight (calculated as daily percentage change from baseline based on measurements recorded from an at-home Bluetooth scale) are shown for each participant over time throughout the duration of the study. Daily mean values over time for each group based on end-of-study weight loss at 12 weeks are represented in solid lines (dark orange, >10% weight loss; light orange, >5% to <10% weight loss; and yellow, <5% weight loss).

Aim 1: User Behavior and Adherence Over Time

Participants were instructed to use the app and accompanying biofeedback device as dietary guidance throughout the intervention and to report dietary adherence on a weekly basis through a web-based questionnaire; self-reported adherence across 12 weeks was moderately high, with an average score of 2.6 (SD 0.2) on a 5-point Likert scale. The average number of engagements with the app across the entire primary intervention period was 123 (SD 91), or 1.5 (SD 0.6) engagements daily, and the average *Keyto* Level was 3.9 (SD 0.3). On average, participants weighed themselves 4.7 (SD 1.2) times per week. Across both the primary and secondary intervention phases, the average self-reported adherence was 2.5 (SD 0.3), *Keyto* Levels were 4.0 (SD 0.2), number of weekly weight measurements was 3.7 (SD 1.4), and the number of engagements with the app was 171 (SD 155), or approximately once per day (Figure 2).

To further assess how different factors might affect self-reported dietary adherence, we asked participants how strongly a given situation would affect their ability to stick to their diet on a scale of 1 (not at all) to 4 (every day). Descriptive data from these questionnaires are presented in Figure 3.

Figure 2. App use metrics (yellow, average *Keyto* Level obtained through breath acetone biofeedback device; dark orange, average daily number of engagements with the *Keyto* app), self-reported dietary adherence (light orange), and average number of weekly weight measurements (brown) averaged across all participants throughout the intervention. Mean values are shown.

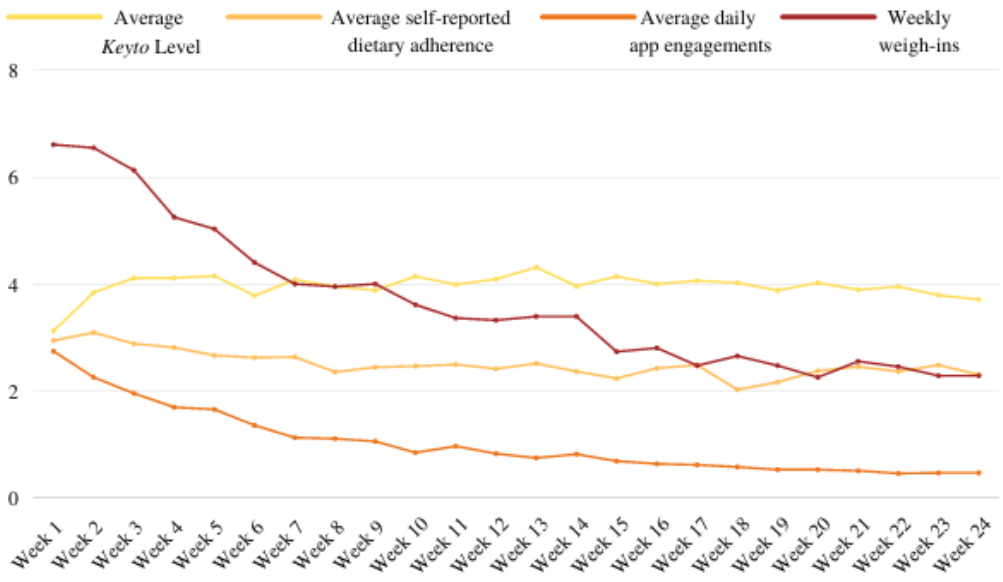
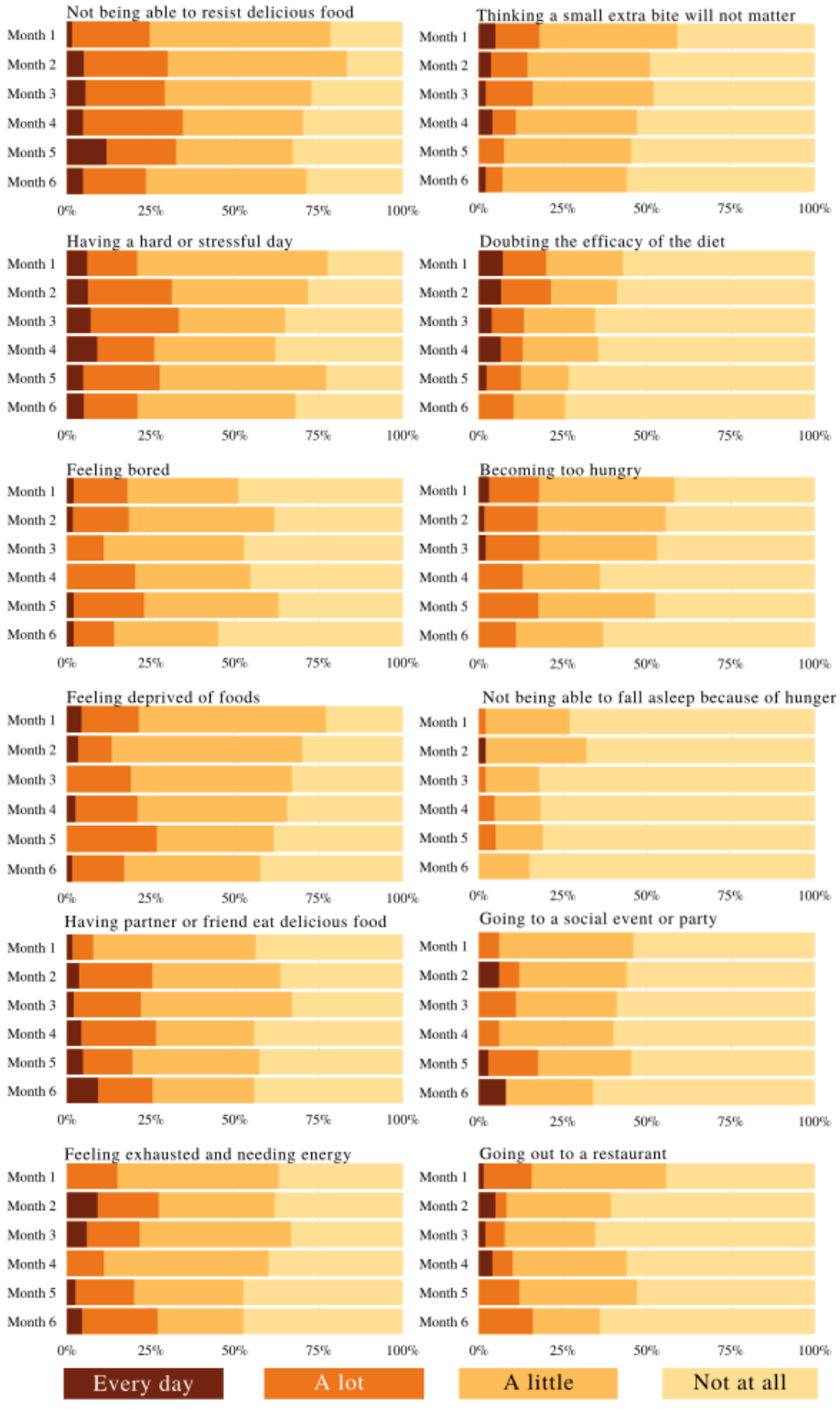


Figure 3. Aggregated responses to questionnaires sent through email, asking participants, "How does the following affect your ability to stick to your diet?" on a 4-point Likert scale, ranging from "Not at all" to "Every day.".



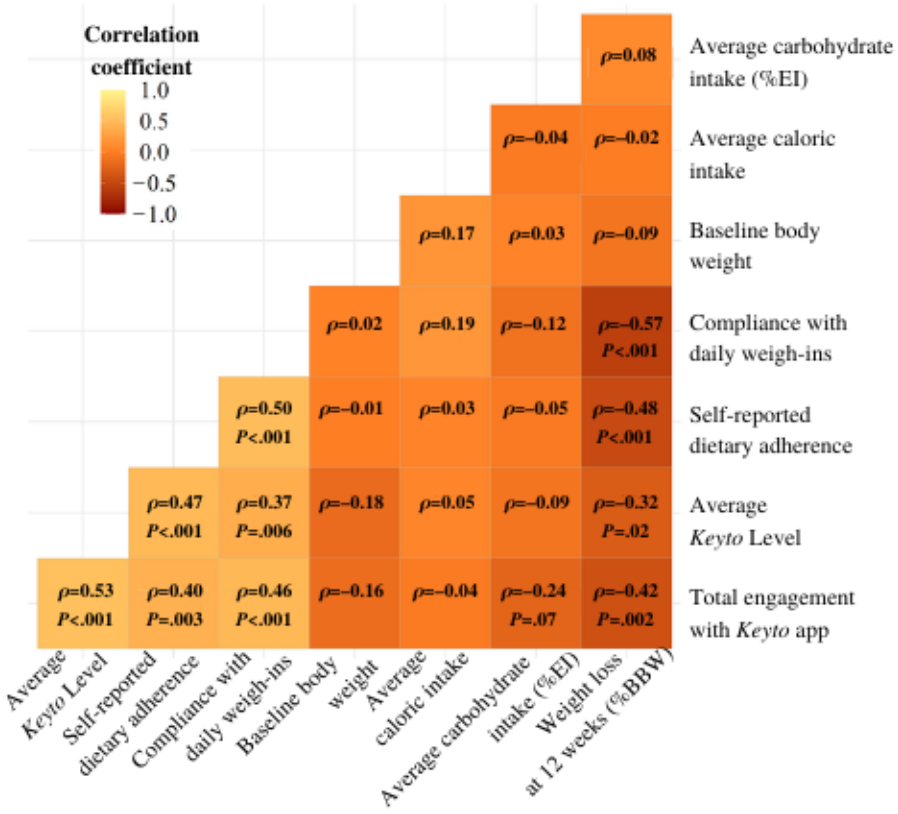
Aim 2: Predictors of Weight Loss at 12 Weeks

Correlation of Potential Predictor Variables

A correlation matrix of the assessed app use–, adherence-, and diet-related variables averaged across the primary 12-week intervention period (average *Keyto* Level, app engagement, self-reported caloric intake, carbohydrate intake as percentage of daily energy intake, compliance with daily weigh-ins, and self-reported dietary adherence), baseline body weight, and weight loss as a percentage of baseline body weight lost is shown in Figure 4. The strongest negative correlations were observed between weight loss and compliance with daily weight measurements (ρ=–0.57; P<.001), self-reported dietary adherence (ρ=–0.48; P<.001), and app engagement (ρ=–0.42; P=.001), suggesting that participants who weighed themselves more regularly, reported greater dietary adherence, and engaged

more often with the app also achieved greater weight loss. In addition, average *Keyto* Levels were significantly correlated with weight loss (ρ=–0.32; *P*=.02), indicating greater weight loss with higher *Keyto* Levels. The strongest positive correlations were observed between average *Keyto* Levels and engagement with the *Keyto* app (ρ=0.53; *P*<.001), suggesting that participants who engaged more with the app also achieved higher *Keyto* Levels, and between self-reported dietary adherence and compliance with daily weight measurements (ρ=0.50; *P*<.001), indicating high agreement between these 2 adherence measures.

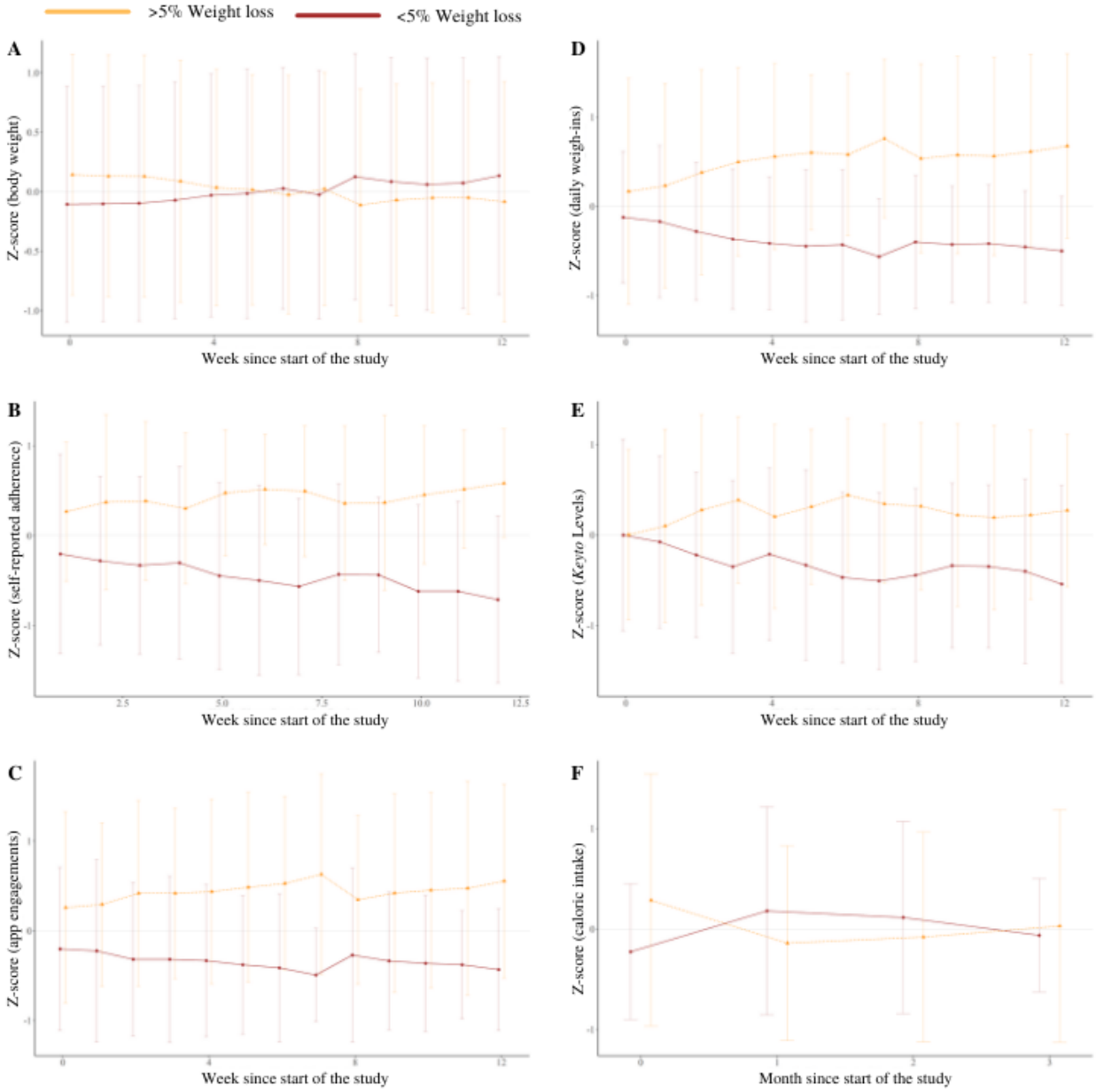
Figure 4. Pairwise Spearman rank correlation matrix of adherence-, app use–, and diet-related variables. %BBW: percentage of baseline body weight; %EI: percentage of energy intake.



Time-Based Data Patterns

To investigate more detailed patterns of assessed variables across time (as opposed to values averaged over the entire intervention period), we plotted the average z-scores (a measure of deviation from the average value at each time point) of adherence-, diet-, and app use–related variable means between the group of participants who achieved clinically significant weight loss of >5% initial body weight (termed as *successful* participants) and the group of participants who did not (ie, participants who lost <5% initial body weight). Variable trends are shown in Figure 5. As expected, there was a significant interaction between success and time for body weight (Figure 5A), with the slope for the successful group decreasing significantly more over time (–0.03, 95% CI –0.034 to –0.027; *P*<.001). Likewise, there was a significant interaction for self-reported dietary adherence (Figure 5B), app engagement (Figure 5C), and compliance with daily weight measurements (Figure 5D): self-reported dietary adherence (0.07, 95% CI 0.04-0.10; *P*<.001), engagement with the *Keyto* app (0.03, 95% CI 0.01-0.05; *P*=.005), and compliance with daily weight measurements (0.06, 95% CI 0.03-0.08; *P*<.001) were higher over time in the successful group of participants. Furthermore, there was a significant effect of success group for these variables: successful participants reported higher dietary adherence (0.47, 95% CI 0.13-0.81; *P*=.007), engaged with the app more often (0.62, 95% CI 0.28-0.97; *P*<.001), and weighed themselves more consistently (0.55, 95% CI 0.25-0.86; *P*<.001). A significant interaction effect was detected for average *Keyto* Levels (Figure 5E): participants in the successful group had higher *Keyto* Levels over time (0.08, 95% CI 0.05-0.10; *P*<.001). No significant interaction or main effect was observed for caloric intake (Figure 5F).

Figure 5. Average z-scores (the mean of the sample subtracted from the observed value divided by the SD of the sample at each time point) comparing time-series patterns of variable means between participants who lost >5% of initial body weight (light orange) and those who lost <5% of baseline body weight (dark orange). Mean values (bold lines) and SD (shaded lines) are shown for (A) body weight, (B) self-reported dietary adherence, (C) app engagement, (D) compliance with daily weight measurements, (E) average *Keyto* Levels, and (F) self-reported caloric intake.



Regression Model

Following up on the aforementioned correlative analyses, we next sought to identify what process variables might be associated with (or be predictive of) weight change; to this end, we built a regression model that included the adherence- and app use–related variables, namely (1) compliance with weight measurements, (2) average self-reported dietary adherence, (3) total number of app engagements, (4) average *Keyto* Levels, and (5) baseline body weight (Table 3).

Table 3. Regression model identifying potential predictor variables of weight loss.

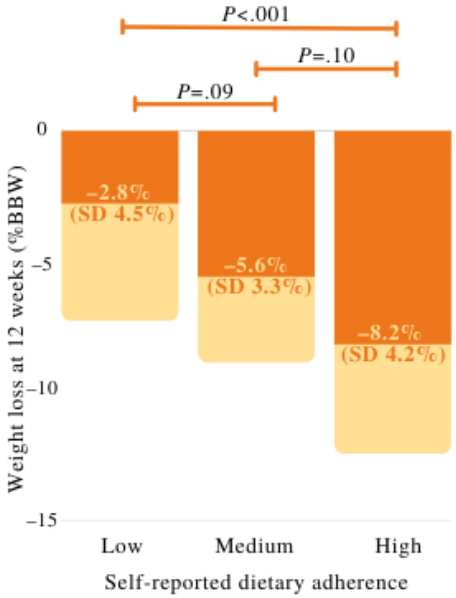
Predictor variable	Values, β (SE)[a]	*t* test (*df*)	*P* value
Compliance with daily weight measurements	–.20 (0.13)	–1.580 (54)	.12
Self-reported dietary adherence	–.31 (0.13)	–2.366 (54)	.02
Total number of app engagements	–.14 (0.13)	–1.069 (54)	.29
Average *Keyto* levels	–.06 (0.13)	–0.469 (54)	.64
Baseline body weight	–.36 (0.11)	–3.420 (54)	.001

[a]Standardized coefficient.

The model (adjusted R^2=0.40; $F_{5,54}$=8.937; *P*<.001) identified self-reported dietary adherence (β=–.31; t_{54}=–2.366; *P*=.02) and baseline body weight (β=–.36; t_{54}=–3.420; *P*=.001) as significant predictors. Compliance with daily weight measurements (β=–.20; t_{54}=–1.580; *P*=.12) was near statistical significance, whereas engagement with the *Keyto* app (β=–.14; t_{54}=–1.069; *P*=.29) and average *Keyto* Levels (β=–.06; t_{54}=–0.469; *P*=.64) were not identified as statistically significant predictors.

To further evaluate the relationship of weight loss with self-reported dietary adherence as a significant predictor variable, we split participants into tertiles based on low *versus* medium *versus* high average self-reported dietary adherence and compared weight loss among these groups (Figure 6). Weight loss was significantly different among the tertiles ($F_{2,57}$=8.586; *P*<.001); post hoc testing revealed a statistically significant difference in weight loss between the participants who reported lowest dietary adherence and those who reported highest dietary adherence (*P*<.001).

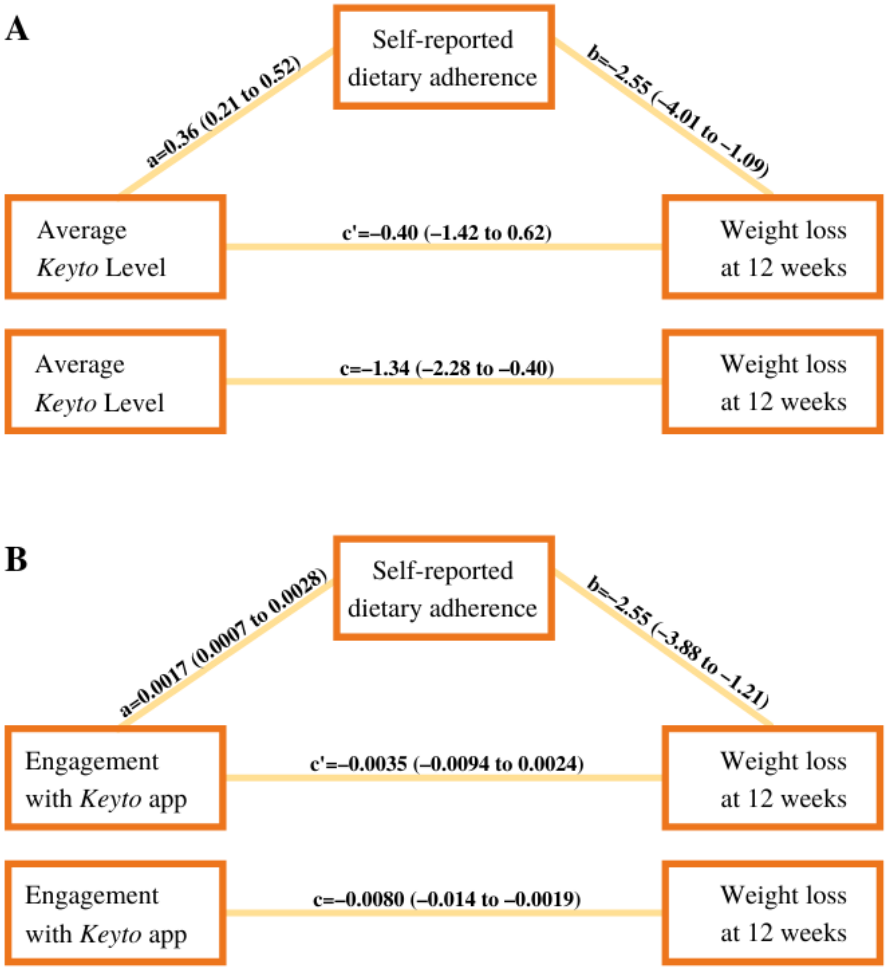
Figure 6. Mean weight loss at the primary intervention end point of 12 weeks of participants in the lowest (left), medium (center), or highest (right) tertile of average self-reported dietary adherence assessed weekly through a questionnaire. One-way analysis of variance with Tukey post hoc tests comparing the differences in group means was conducted. Mean (dark orange) and SD (light orange) are shown. %BBW: percentage of baseline body weight.



Mediation Analysis

Finally, because we were interested in the effect of features specific to the *Keyto* app, we set out to further examine explanatory pathways of the relationship between average *Keyto* Levels or overall engagement with the app, weight loss, and the previously identified predictor, that is, self-reported adherence to the dietary intervention. Figure 7 depicts the results of these 2 mediation analyses. Self-reported dietary adherence fully mediated the relationship between average *Keyto* Level and weight loss at 12 weeks (Figure 7A). When self-reported dietary adherence was added to the model, the direct relationship between average *Keyto* Level and 12-week weight loss became nonsignificant (see change from *c* to *c'*). Average *Keyto* Level accounted for 12.3% of the variance in weight loss when included on its own, and 27.8% was explained when dietary adherence was included in the mediation model. Similarly, self-reported dietary adherence fully mediated the relationship between the overall engagement with the *Keyto* app and 12-week weight loss (Figure 7B). When self-reported dietary adherence was added to the model, the direct relationship between app engagement and 12-week weight loss became nonsignificant (see change from *c* to *c'*). On its own, app engagement accounted for 10.7% of the variance in weight loss, whereas 28.8% of the variance was explained when dietary adherence was included in the mediation model.

Figure 7. Unstandardized effects with 95% CIs of the direct and indirect mediation effects of self-reported dietary adherence on (A) average *Keyto* Levels and (B) total engagement with the *Keyto* app on body weight loss at the end of the primary intervention phase at 12 weeks.



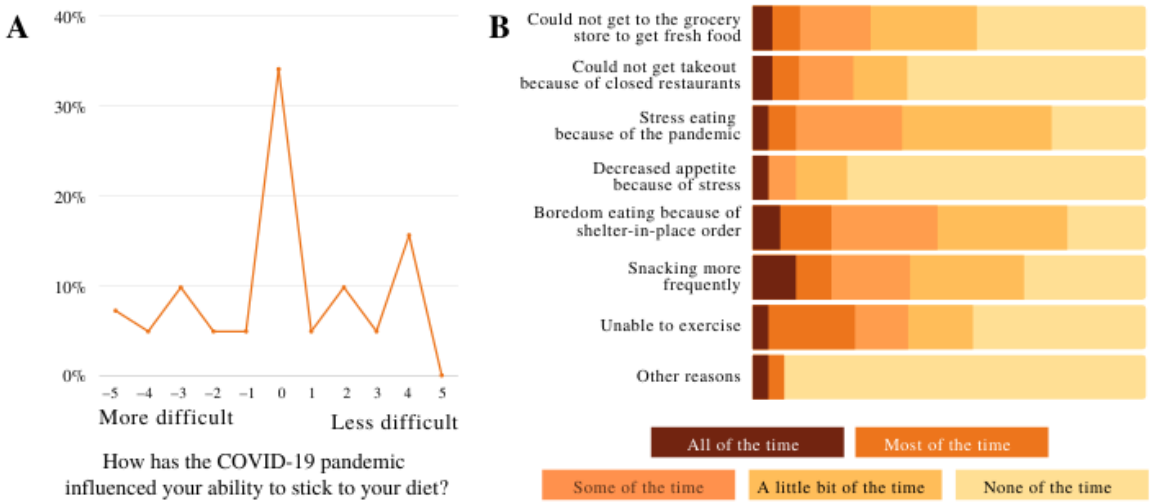
Higher average *Keyto* Levels and greater engagement with the *Keyto* app were associated with greater ability to adhere to the dietary intervention, which in turn was associated with 12-week weight loss. These findings suggest that using the *Keyto* app more frequently and achieving higher *Keyto* Levels caused individuals to better adhere to the dietary intervention, which resulted in greater weight loss.

Impact of COVID-19

Although the study was designed before the COVID-19 pandemic, most of the components of the trial were conducted during the pandemic. Therefore, at the end of the primary intervention phase, we asked participants to report how the COVID-19 pandemic had influenced their ability to stick to their diet on a scale of –5 (more difficult) to +5 (less difficult). The average response to this question was 0.2 (SD 3.1), suggesting a generally minor impact of the pandemic; however, the large variability indicates that some participants experienced significantly greater challenges because of the pandemic (Figure 8A). Regression analysis revealed that the self-reported impact of COVID-19 on participants' ability to stick to their diet explained 22.1% of the variance in self-reported dietary adherence ($F_{1,43}$=13.45; P<.001).

Figure 8B shows aggregated responses to the question of how the listed scenarios affected the participants' ability to stick to their diet; in particular, stress eating, boredom eating, and snacking more frequently were identified as posing the greatest challenges to sticking to the diet during the pandemic.

Figure 8. Response to the questionnaires sent through email, asking participants how (A) the COVID-19 pandemic has influenced their ability to stick to their diet on a scale of –5 (more difficult) to +5 (less difficult) and (B) how the listed items affected their ability to stick to their diet with respect to the COVID-19 pandemic on a 5-point Likert scale ranging from "None of the time" to "All of the time.".



Discussion

Study Overview

In this secondary analysis of a previous randomized clinical trial [9], we showed that a smartphone-based mHealth app promoting a low-carbohydrate ketogenic diet high in mono- and polyunsaturated fatty acids was effective at reducing body weight over the course of 12 and 24 weeks. Furthermore, we characterized real-world user behavior with the app and the accompanying breath acetone biofeedback device over the course of the intervention period and investigated differences between the participants who achieved clinically significant weight loss and those who did not. Finally, we identified predictor variables of weight loss and showed that self-reported adherence to the dietary intervention seemed to be the most important factor in predicting weight loss success. This suggests that a relatively simple measure of self-reported adherence could be used to help predict who will be successful with an mHealth app-based dietary weight loss intervention, which could allow for greater individualization of future interventions.

Relevance of Findings

The weight loss observed in the intervention group in this trial compared favorably with other weight loss–promoting mHealth apps [16] and previous weight reduction trials that used a low-carbohydrate ketogenic diet using more traditional and hands-on (ie, nutritional counseling and in-person group meetings) designs [17]. This suggests that an app paired with a biofeedback device that guides users to follow a low-carbohydrate ketogenic diet is a feasible and effective approach for promoting weight loss in a pragmatic real-world setting that is highly scalable and presents low burden of entry.

To our knowledge, this is the first study evaluating user behavior of a smartphone-based weight loss app using a low-carbohydrate ketogenic diet. Because of the nature of the diet, substantial knowledge is required to identify suitable food items that fit this dietary plan. In addition, a ketogenic diet offers the opportunity for self-monitoring of a fat loss biomarker (ie, breath acetone) other than, or in addition to, the weight scale, which is unique to this dietary pattern. The breath acetone biofeedback device used in this study is of particular interest because most available methods of measuring ketone bodies are invasive (eg, measuring beta-hydroxybutyrate through finger pricks) or offer only a rough proxy of ketosis (eg, measuring acetoacetate through urine sticks). Our study therefore adds valuable insight into how users interact with an app promoting such a diet without direct in-person supervision and which behaviors lead to successful weight loss.

Similar to previous studies investigating mobile weight loss interventions [18], we found that the measures of dietary adherence and intervention fidelity (ie, engagement with the app and body weight measurements) decreased over time. However, importantly, we showed that this decrease was less pronounced in participants who achieved clinically significant weight loss (ie, >5% baseline body weight) compared with those who did not. Upon further investigation of data patterns over time, we found that this difference in behaviors between *successful* and *unsuccessful* participants was apparent right from the beginning of the intervention, with successful participants reporting greater levels of dietary adherence, weighing themselves more regularly, and engaging more often with the app. Moving from correlative to more causative associations, our mediation analysis further suggests that self-reported dietary adherence was a mechanism through which engagement with the *Keyto* app and higher *Keyto* Levels affected weight loss during the primary intervention period of 12 weeks. Interestingly, the average *Keyto* Level as a direct measure of ketosis and an indirect measure of dietary adherence did not seem to decrease over time; this would suggest that for participants to maintain a state of ketosis, persistent engagement with the app (which, in contrast, did decrease over time) was not required after an initial familiarization period. However, it is important to note that the metric of average *Keyto* Levels in our study is upwardly biased because of response bias (ie, only participants who actively use their breath acetone device will

record a *Keyto* Level) and therefore a measure of people who are at least somewhat engaged with the dietary intervention, whereas engagement with the app is representative of all participants regardless of their engagement with the study (ie, a meaningful zero exists for number of app engagements).

Overall, our data support the importance of adherence to the intervention (more so than absolute carbohydrate restriction or level of ketosis) to achieve meaningful weight loss. Although the different measures used to evaluate adherence showed high correlation, indicating a large degree of concordance among them (ie, participants who weighed themselves more regularly generally engaged with the app more often and so on), more detailed analysis implied that different measures of adherence (ie, daily weight measurements, engagement with the app, and self-reported adherence) were differentially associated with weight loss success, supporting the notion of adherence as a multifaceted construct. In particular, our results suggest that self-reported dietary adherence—a subjective assessment provided by participants through weekly questionnaires—was the most important predictor of successful weight loss; future investigations should therefore aim to determine the extent to which objective adherence to the dietary intervention and self-reported dietary adherence are in agreement and which factors influence self-reported adherence, which may in turn inform the design of future dietary app-based interventions. The finding that a relatively simple self-reported assessment of dietary adherence seemed to be the most powerful predictor of successful weight loss also has practical value that could help inform future trials and knowledge translation for the design of app-based interventions. We speculate that evaluating self-reported adherence could help to determine early on in the trial which participants might need extra support (eg, through SMS text messages, push notifications, and phone calls) to maximize weight loss success.

Limitations

Importantly, our trial design was unable to separate the effect of individual intervention components (eg, breath acetone device, app engagement, and recording of weight measurements); although we attempted to statistically tease apart variables of predictive power, our study may inspire future research with other trial designs (eg, Multiphase Optimization Strategy–based [19]) to further investigate the independent importance of intervention constituents.

Furthermore, our trial was conducted in a real-world setting. Although this pragmatic trial design was chosen on purpose to evaluate the effect of the app-based intervention in a realistic setting, we were therefore limited, for the most part, to self-reported data entries; in particular, we relied on dietary self-report for measures of caloric intake, which can be biased and provide inaccurate estimates [20], and we asked users to report dietary adherence by means of a weekly questionnaire, which represents a subjective estimate of adherence to the intervention as opposed to an objective assessment. Similarly, body weight was not measured by a trained researcher in a laboratory setting but instead self-administered on an at-home Bluetooth scale by the participants, which could potentially introduce bias.

In addition, our findings on user behavior and its change over time are limited to the variables collected within the framework of this trial. It is likely that other variables, including baseline characteristics and other behaviors related to dieting and self-monitoring that were not assessed in this study, are of importance to the success that an individual sees with a given weight loss program. Similarly, our findings are constrained to the evaluated mHealth app promoting a ketogenic diet and may not be as relevant to other dietary interventions.

Notably, our study was conducted throughout the COVID-19 pandemic. Although most participants reported minor effects of the pandemic on their eating plan, some of our psychological measures may have been affected by it. For example, we assessed how going out to a restaurant or a social event affected participants' ability to stick to their diet; however, because of the statewide shelter-in-place order during the trial, which affected many participants and prohibited restaurant visits or social gatherings, this measure may not accurately reflect the challenges that adults trying to lose weight may encounter under *normal* conditions. Likewise, the greater emotional burden throughout the pandemic [21] may have altered participants' responses to questions about stress eating or boredom eating.

Finally, our study investigated user behavior and success predictors primarily over a time period of 12 weeks. This represents the effects of a short-term weight loss intervention; therefore, no conclusions can be drawn about the observed relationships on continued weight loss in the context of longer-term dietary interventions or sustained weight maintenance.

Conclusions

This study adds to the available literature on the use of mHealth technology in assisting self-guided weight loss attempts and supports the notion that even in the contexts of a low-carbohydrate ketogenic diet and mHealth technology, dietary adherence is of crucial importance to achieve the desired reduction in weight. Therefore, our study may inspire future research into how self-reported dietary adherence can be used or enhanced for long-term weight loss and maintenance and inform the design of low-carbohydrate ketogenic dietary mHealth interventions.

Acknowledgments

This work was supported by the Canadian Institutes of Health Research (MSH-141980), Michael Smith Foundation for Health Research (MSFHR 16890), and Mitacs (IT15608).

Authors' Contributions

KF, SRL, DAL, EJW, and JPL designed the analyses. KF, SRL, and DAL acquired the data. KF and SRL performed the statistical analyses with guidance from TL, JS, and JPL. KF, SRL, and JPL drafted the manuscript with critical revisions for important intellectual content from DAL, TL, JS, and EJW. EJW and JS provided administrative support. JPL had full access to all the data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

JPL is volunteer chief scientific officer for the not-for-profit Institute for Personalized Therapeutic Nutrition and holds founder shares in Metabolic Insights Inc, a for-profit company that developed noninvasive metabolic monitoring devices. EJW is an equity holder at *Keyto* Inc and Virta Health. DAL is employed as a consultant for *Keyto* Inc.

Multimedia Appendix 1
Weekly survey used throughout the trial to assess self-reported adherence and cravings, mood, and energy.
[DOCX File , 21 KB - mhealth_v10i3e33940_app1.docx]

References

1. Pi-Sunyer X. The medical risks of obesity. Postgrad Med 2009;121(6):21-33 [FREE Full text] [doi: 10.3810/pgm.2009.11.2074] [Medline: 19940414]
2. Obesity and overweight. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-09-29]
3. Santos I, Sniehotta FF, Marques MM, Carraça EV, Teixeira PJ. Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. Obes Rev 2017;18(1):32-50 [FREE Full text] [doi: 10.1111/obr.12466] [Medline: 27653242]
4. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
5. Paoli A, Rubini A, Volek JS, Grimaldi KA. Beyond weight loss: a review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. Eur J Clin Nutr 2013;67(8):789-796 [FREE Full text] [doi: 10.1038/ejcn.2013.116] [Medline: 23801097]
6. Bueno NB, de Melo IS, de Oliveira SL, da Rocha Ataide T. Very-low-carbohydrate ketogenic diet v. low-fat diet for long-term weight loss: a meta-analysis of randomised controlled trials. Br J Nutr 2013;110(7):1178-1187. [doi: 10.1017/S0007114513000548] [Medline: 23651522]
7. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018;378(25):e34. [doi: 10.1056/NEJMoa1800389] [Medline: 29897866]
8. Kirkpatrick CF, Bolick JP, Kris-Etherton PM, Sikand G, Aspry KE, Soffer DE, et al. Review of current evidence and clinical recommendations on the effects of low-carbohydrate and very-low-carbohydrate (including ketogenic) diets for the management of body weight and other cardiometabolic risk factors: a scientific statement from the National Lipid Association Nutrition and Lifestyle Task Force. J Clin Lipidol 2019;13(5):689-711.e1 [FREE Full text] [doi: 10.1016/j.jacl.2019.08.003] [Medline: 31611148]
9. Falkenhain K, Locke SR, Lowe DA, Reitsma NJ, Lee T, Singer J, et al. Keyto app and device versus WW app on weight loss and metabolic risk in adults with overweight or obesity: a randomized trial. Obesity (Silver Spring) 2021;29(10):1606-1614 [FREE Full text] [doi: 10.1002/oby.23242] [Medline: 34124856]
10. Locke SR, Falkenhain K, Lowe DA, Lee T, Singer J, Weiss EJ, et al. Comparing the Keyto app and device with weight watchers' WW app for weight loss: protocol for a randomized trial. JMIR Res Protoc 2020;9(8):e19053 [FREE Full text] [doi: 10.2196/19053] [Medline: 32804087]
11. Alkedeh O, Priefer R. The ketogenic diet: breath acetone sensing technology. Biosensors (Basel) 2021;11(1):26 [FREE Full text] [doi: 10.3390/bios11010026] [Medline: 33478049]
12. The science behind Keyto, part one – how the Keyto breath sensor works. Keyto. 2021. URL: https://getkeyto.com/science-behind-keyto-how-the-breath-sensor-works/ [accessed 2021-09-29]
13. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The automated self-administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012;112(8):1134-1137 [FREE Full text] [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]
14. Blackburn G. Effect of degree of weight loss on health benefits. Obes Res 1995;3 Suppl 2:211s-216s [FREE Full text] [doi: 10.1002/j.1550-8528.1995.tb00466.x] [Medline: 8581779]
15. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. 2nd edition. New York, NY: The Guilford Press; 2017.
16. 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;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]

17. Churuangsuk C, Kherouf M, Combet E, Lean M. Low-carbohydrate diets for overweight and obesity: a systematic review of the systematic reviews. Obes Rev 2018;19(12):1700-1718. [doi: 10.1111/obr.12744] [Medline: 30194696]
18. Kim J, Kam HJ, Kim Y, Lee Y, Lee JH. Understanding time series patterns of weight and meal history reports in mobile weight loss intervention programs: data-driven analysis. J Med Internet Res 2020;22(8):e17521 [FREE Full text] [doi: 10.2196/17521] [Medline: 32780028]
19. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
20. Park Y, Dodd KW, Kipnis V, Thompson FE, Potischman N, Schoeller DA, et al. Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018;107(1):80-93 [FREE Full text] [doi: 10.1093/ajcn/nqx002] [Medline: 29381789]
21. Pedrosa AL, Bitencourt L, Fróes AC, Cazumbá ML, Campos RG, de Brito SB, et al. Emotional, behavioral, and psychological impact of the COVID-19 pandemic. Front Psychol 2020;11:566212 [FREE Full text] [doi: 10.3389/fpsyg.2020.566212] [Medline: 33117234]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 29.09.21; peer-reviewed by P Amir, N Maglaveras; comments to author 01.11.21; revised version received 19.11.21; accepted 28.01.22; published 14.03.22.

Please cite as:
Falkenhain K, Locke SR, Lowe DA, Lee T, Singer J, Weiss EJ, Little JP
Use of an mHealth Ketogenic Diet App Intervention and User Behaviors Associated With Weight Loss in Adults With Overweight or Obesity: Secondary Analysis of a Randomized Clinical Trial
JMIR Mhealth Uhealth 2022;10(3):e33940
URL: https://mhealth.jmir.org/2022/3/e33940
doi:10.2196/33940
PMID:35285809

Original Paper

Web-Based Versus Print-Based Physical Activity Intervention for Community-Dwelling Older Adults: Crossover Randomized Trial

Claudia R Pischke[1], PhD; Claudia Voelcker-Rehage[2], PhD; Tiara Ratz[3], MSc; Manuela Peters[4], MSc; Christoph Buck[5], PhD; Jochen Meyer[6], PhD; Kai von Holdt[6], BSc; Sonia Lippke[3], DPhil

[1]Institute of Medical Sociology, Centre for Health and Society, Medical Faculty, Heinrich Heine University Duesseldorf, Duesseldorf, Germany
[2]Department of Neuromotor Behavior and Exercise, Institute of Sport and Exercise Sciences, University of Muenster, Muenster, Germany
[3]Department of Psychology and Methods, Jacobs University Bremen, Bremen, Germany
[4]Department of Prevention and Evaluation, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany
[5]Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany
[6]OFFIS – Institute for Information Technology, Oldenburg, Germany

Corresponding Author:
Claudia R Pischke, PhD
Institute of Medical Sociology, Centre for Health and Society, Medical Faculty
Heinrich Heine University Duesseldorf
Moorenstrasse 5
Duesseldorf, 40225
Germany
Phone: 49 211 81 ext 08599
Email: ClaudiaRuth.Pischke@med.uni-duesseldorf.de

Abstract

Background: Fewer than half of older German adults engage in the recommended levels of endurance training.

Objective: The study aim is to compare the acceptance and effectiveness of two interventions for physical activity (PA) promotion among initially inactive community-dwelling older adults ≥60 years in a 9-month, crossover randomized trial.

Methods: Participants were recruited in person and randomized to one of the following interventions for self-monitoring PA: a print-based intervention (PRINT: 113/242, 46.7%) or a web-based intervention (WEB: 129/242, 53.3%). Furthermore, 29.5% (38/129) of those in the web-based intervention group received a PA tracker in addition to WEB (WEB+). After randomization, the participants and researchers were not blinded. The participants' baseline intervention preferences were retrospectively assessed. All the intervention groups were offered 10 weekly face-to-face group sessions. Afterward, participants could choose to stay in their group or cross over to one of the other groups, and group sessions were continued monthly for another 6 months. 3D accelerometers to assess PA and sedentary behavior (SB) at baseline (T0), 3-month follow-up (T1), and 9-month follow-up (T2) were used. Adherence to PA recommendations, attendance of group sessions, and intervention acceptance were assessed using self-administered paper-based questionnaires. Linear mixed models were used to calculate differences in moderate to vigorous PA (MVPA) and SB between time points and intervention groups.

Results: Of the 242 initially recruited participants, 91 (37.6%) were randomized to the WEB group; 38 (15.7%) to the WEB+ group; and 113 (46.7%) to the PRINT group. Overall, 80.6% (195/242) of the participants completed T1. Only 0.4% (1/242) of the participants changed from the WEB group to the PRINT group and 6.2% (15/242) moved from the PRINT group to the WEB group (WEB-WEB: 103/249, (41.4%); PRINT-PRINT: 76/249, 30.5%) when offered to cross over at T1. Furthermore, 66.1% (160/242) of participants completed T2. MVPA in minutes per day increased between baseline and T1, but these within-group changes disappeared after adjusting for covariates. MVPA decreased by 9 minutes per day between baseline and T2 (β_{time}=−9.37, 95% CI −18.58 to −0.16), regardless of the intervention group (WEB vs PRINT: $\beta_{group*time}$=−3.76, 95% CI −13.33 to 5.82, WEB+ vs PRINT: $\beta_{group*time}$=1.40, 95% CI −11.04 to 13.83). Of the participants, 18.6% (38/204) met the PA recommendations at T0, 16.4% (26/159) at T1, and 20.3% (28/138) at T2. For SB, there were no significant group differences or group-by-time interactions at T1 or T2. Intervention acceptance was generally high. The use of intervention material was high to moderate at T1 and decreased by T2.

Conclusions: There was little movement between intervention groups at T1 when given the choice, and participation was not associated with increases in PA or decreases in SB over time.

Trial Registration: German Clinical Trials Register DRKS00016073; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00016073

(JMIR Mhealth Uhealth 2022;10(3):e32212) doi:10.2196/32212

KEYWORDS

physical activity; older adults; eHealth; print-based intervention; web-based intervention; physical activity promotion; healthy aging; preferences; randomized trial; mobile phone

Introduction

Background

Engaging in an active lifestyle with regular physical activity (PA) [1] is associated with higher physical, cognitive, and functional health across the life course [1,2], and web-based support can help individuals to adopt and maintain PA [3]. At the time of study conception, the World Health Organization (WHO) and the American College of Sports Medicine recommended that adults aged 18-64 years, as well as those aged ≥65 years, should perform moderate to vigorous endurance training for at least 150 minutes per week (in 10-minute bouts) [4]. Furthermore, performing flexibility, strength, and balance exercises twice per week is recommended [5,6]. The 2020 update of the recommendations included several modifications. For example, the authors of the update state that "all adults should undertake 150-300 min of moderate-intensity, or 75-150 min of vigorous-intensity physical activity, or some equivalent combination of moderate-intensity and vigorous-intensity aerobic physical activity, per week" [7]. An additional change is that 10-minute bouts of PA are no longer deemed relevant. Instead, bouts of moderate to vigorous PA (MVPA) of any duration count, taking new evidence into account, which suggests that the total PA volume is more important than bouts. Furthermore, high-certainty evidence summarized for the development of the new guidelines indicates that balance and functional exercises are relevant for maintaining physical function and reducing falls [7]. Hence, in the update for the age group of ≥65 years, the recommendation is to incorporate these types of exercises at moderate or greater intensity on 3 or more days per week in existing routines [7].

Less than half of the German adults aged ≥65 years meet the former recommendations for endurance training (42%), and only one-third meet the strength training recommendations [6]. However, compared with the European Union average of adults in this age segment (26.2% of women and 35.7% of men reach the recommendation of 150 minutes of MVPA per week), German men and women display slightly higher proportions of adults reaching the recommendations (45.5% and 51.2% for women and men, respectively) [8]. Furthermore, results based on the European Health Interview Survey and the Survey of Health Aging and Retirement in Europe examined associations between the proportion of European adults >65 years, reaching the recommendation of >150 minutes of PA per week and the proportion of prefrail or frail individuals suggests a negative association [9]. To prevent frailty in older adults, Haider et al [9] called for "community-based approaches aimed at achieving PA recommendations" at the population level and the creation of built environments enabling PA [9]. Previous research conducted in Germany indicated that population-based approaches to increase PA, such as mass media campaigns, community-based multicomponent interventions, and environmental approaches, can be effective in the general population [10]. In addition, individual-level interventions provide opportunities to further increase the effect of such population-level approaches [1,11-13]. However, the role of different modalities in delivering these intervention approaches to older adults remains unclear.

The results of several systematic reviews indicate that participation in interventions providing information on PA face-to-face or via printed materials leads to increased PA levels in older adults [14-16]. Engagement in web-based PA interventions is also associated with increased MVPA, walking, and a higher daily step count in the intervention group than in the control group [17,18]. Furthermore, the results of a systematic review evaluating the effectiveness of eHealth interventions compared with non-eHealth interventions or no intervention in adults ≥55 years suggest that eHealth interventions can effectively promote PA in the short term [13], but there is still a lack of evidence regarding long-term effects. Recent evidence from a review examining the effects and characteristics of PA promotion interventions aimed at community-dwelling adults >50 years indicates that increases in PA can be sustained for up to 12 months [19]. In conclusion, it is still unclear whether eHealth interventions have a greater impact on PA behavior than non-eHealth (eg, print-based) interventions in adults who are ≥60 years and whether increased levels of PA can be maintained over longer periods.

Furthermore, the influence of individual preferences for intervention modality and variances in the impact on intervention outcomes is still not well understood [20,21]. Previous studies suggest that preferences may vary by age, sex, BMI, or social or living environment [15,22,23]. For example, preference for a web-based intervention was positively related to younger age [22,23] and high internet use and was negatively associated with the female sex. Conversely, older women with obesity were more likely to choose print-based interventions [22]. These variations in sociodemographic characteristics may also explain the differences in the use of PA trackers [24]. To increase the impact of this tool that has already been shown to be effective [25], the use of trackers in PA interventions should be aligned with preferences of different target groups [15]. Both retention in intervention studies and adherence to intervention components may improve if individual preferences for interventions are considered [12,15]. Hence, in this study, a crossover design was used to examine the role of personal preferences for different delivery modes in intervention effectiveness.

This study (PROMOTE II) was funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung), as part of the Physical Activity and Health Equity: Primary Prevention for Healthy Aging research network [26]. It builds on the results of a previous study embedded in the network (PROMOTE I; [27-29]), which tested the effectiveness of 2 tailored web-based interventions for the promotion of a physically active lifestyle in adults aged 65-75 years in a community-based intervention trial against a delayed-intervention control group. In a previous study, we found relatively high baseline PA levels in the intervention participants. On the basis of this observation, individuals who had been physically active regularly for at least 2.5 hours per week for >1 year were excluded from this study. Furthermore, study dropout was higher in the group assigned to use PA trackers in addition to a website than in the website-only group, indicating that randomization to a modality that was not a preference led to participants deciding to quit the intervention [27-29].

Objectives

On the basis of these results gathered in the preceding study, this study included the following four aims:

1. To adapt and simplify the web-based intervention of a previous study to further improve usability and develop a simple print-based intervention that initially inactive participants with little affinity to technology find easy to use.
2. To investigate the acceptance and use of two interventions (web- vs print-based) and changes in PA among older adults (≥60 years) in a crossover randomized trial over the course of 9 months.
3. To examine the role of personal preferences for different delivery modes in intervention effectiveness.
4. To explore the associations between changes in PA and possible changes in physical fitness and cognitive capacity in a pooled sample of participants in both PROMOTE I and II trials.

In this paper, we report the results of the first 3 study aims. The results addressing the last aim will be reported in a subsequent paper. We hypothesized that both interventions would significantly increase MVPA and decrease sedentary behavior (SB) at the first and second follow-ups [30].

Methods

Participants and Procedures

Recruitment

A random sample of 3492 adults aged ≥60 years from 14 districts in Bremen, Germany, were invited to participate in the study via mail. The names and addresses were provided by the residents' registration offices. This included individuals who resided in districts that met the following requirements:

1. Districts that were not part of the municipalities targeted in PROMOTE I
2. Districts that were in close proximity to the two study centers (one in the Northwest and one in the Northeast of the city of Bremen, Germany)
3. Districts where the project team had already established previous liaisons, including contacts with stakeholders facilitating community involvement during the implementation of the intervention

Reminders were sent out after 2 weeks in cases of no response. The study was also publicized in local newspaper articles and mentioned during talks of the research staff, sparking the interest of 168 individuals who called up the research team directly and were consequently screened for eligibility. Eligibility for study participation was determined through computer-assisted telephone interviews with trained study nurses following the inclusion and exclusion criteria outlined below. The sample size and power calculations are described in detail in a previously published study protocol [30].

Ethical Approval and Informed Consent

The study obtained ethical approval from the Medical Association of Bremen, Germany, on July 3, 2018 (RA/RE-635). The study was registered at the German Clinical Trials Register on January 10, 2019 (DRKS00016073). Potential participants were informed of the study during the initial telephone interviews and were fully informed during an introductory face-to-face briefing session and were requested to provide informed consent. They were also told that they would be randomized to one of the intervention groups and knew about their existence. At the end of the introductory session, all participants were fully informed of the study and provided informed consent. The participants, research staff conducting the study, or statistician analyzing the data were not blinded to the intervention.

Inclusion and Exclusion Criteria

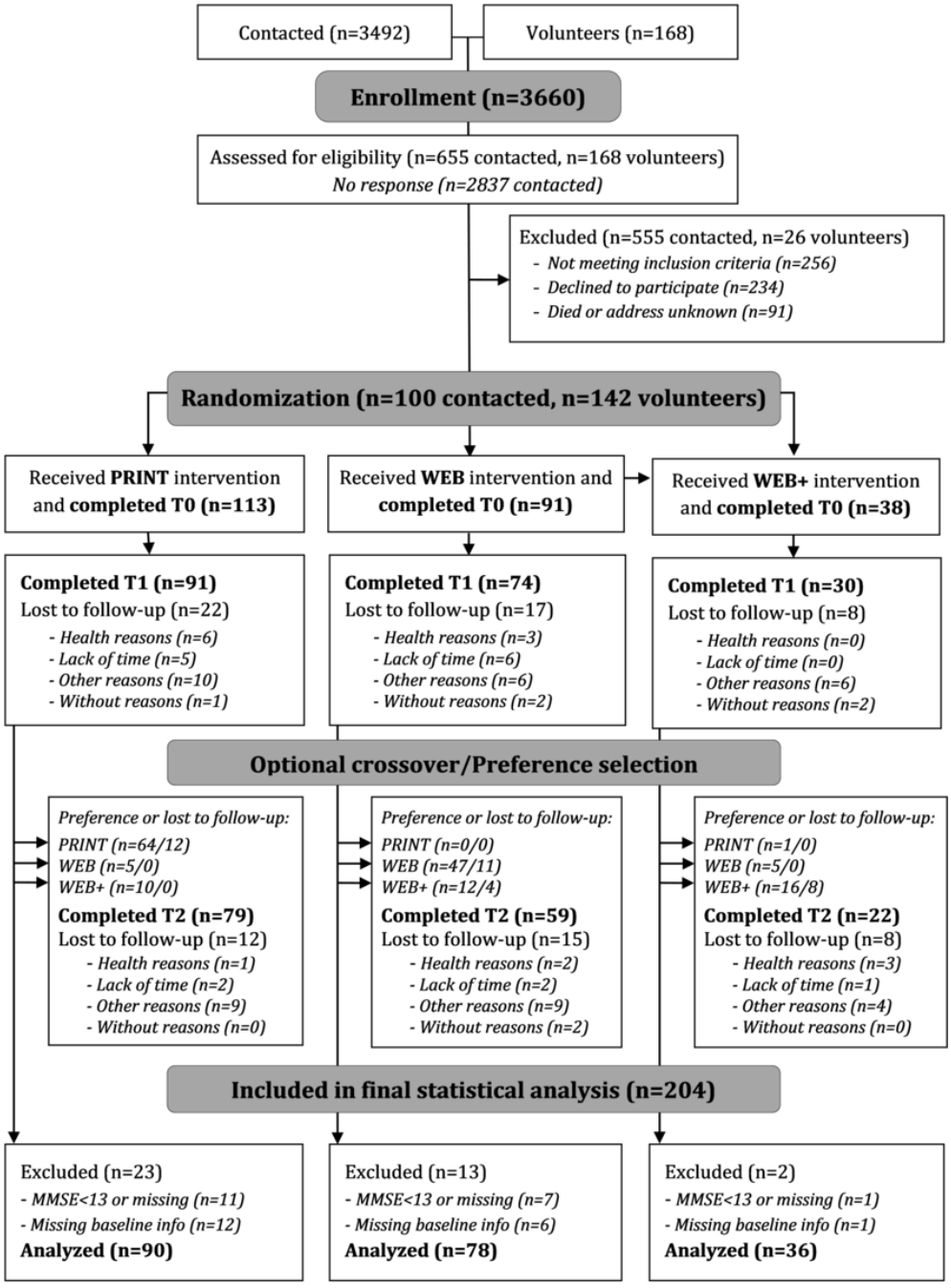
Briefly, individuals were included in the study if they were aged ≥60 years, lived independently, and provided informed consent. Individuals were excluded from the study if they reported that they had been physically active regularly for at least 2.5 hours per week for >1 year. Furthermore, having participated in the previous trial, a planned vacation during the intervention period exceeding 2 weeks, a medical condition or diagnosis prohibiting PA, severe visual or other impairments, implanted cardiac devices, or occasional syncopal episodes led to exclusion (see the study protocol by Pischke et al [30] for further details). Cognitive state was measured using the Mini-Mental State Examination 2–brief version (MMSE-2-BV) [31], and the exclusion criterion was initially set to an MMSE-2-BV score of ≤14. As the manual for the MMSE-2-BV does not define a cutoff value for the determination of cognitive impairment, the initially chosen cutoff value was re-evaluated during the study and was found to be too conservative. On the basis of previous studies [32,33], the cutoff value was adapted, and individuals with an MMSE-2-BV score <13 were excluded.

Randomization and Allocation

Of the 3660 older adults invited, 823 (22.49%) individuals were assessed for eligibility during computer-assisted telephone interviews (Figure 1). In total, 70.6% (581/823) of the potential

participants were excluded. After determination of eligibility, 29.4% (242/823) of the study participants were randomized to one of two groups by the study nurse applying an allocation ratio of 1 to 1: (1) a print intervention with subjective PA self-monitoring via printed PA-pyramid (PRINT: 113/242, 46.7%) and (2) a web-based intervention with subjective PA self-monitoring via a web-based PA-pyramid (WEB: 129/242, 53.3%). Furthermore, 29.5% (38/129) of those in the web-based intervention group were randomly selected and received a PA tracker (objective PA self-monitoring) in addition (WEB+ group). Weekly time slots were randomly assigned to the 3 intervention groups. The first 30% of the timeslots reserved for the WEB group received Fitbit devices (WEB+). Participants were blinded to the intervention group during randomization (ie, they were free to choose from available time slots during the telephone interview with the study nurse, without knowing which intervention group they were assigned to).

Figure 1. Participant flow. MMSE: Mini-Mental State Examination; T0: baseline assessment; T1: 3-month follow-up; T2: 9-month follow-up.

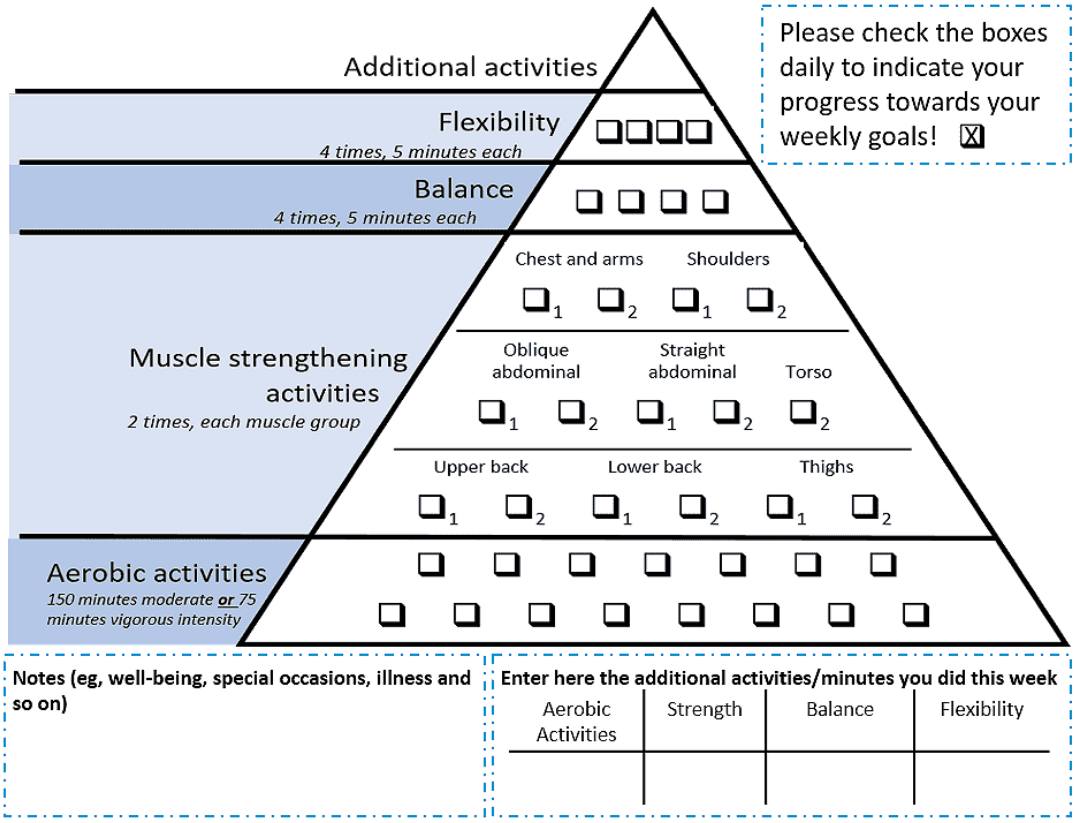


Interventions

The design process of the interventions and their contents are described elsewhere [30]. They were designed based on the state-of-the-art research on PA and the results of focus group discussions conducted with the target group. The intervention content was based on self-regulation theory and various behavior change techniques facilitating regular self-monitoring of PA [34,35]. Participants in both groups received PA recommendations according to the WHO, and brochures (web based and in print) were provided outlining exercises for

different difficulty levels, showing pictures of male versus female older adults modeling these exercises [30]; they additionally received a diary to track their PA. Depending on the group assignment, all intervention materials were either provided as printed materials or made available on the website. The smartphone app additionally provided access to the exercises and PA diary for individuals in the WEB and WEB+. On the website, in the Android web app, as well as in the printed diary, weekly feedback regarding whether PA goals were reached was provided (Figure 2), the number of minutes or units exercised and the units required to reach the goal were displayed. The WEB+ group used a PA tracker (Fitbit Zip, Fitbit Inc) in addition to the website or app, and the daily step count tracked with the device was synchronized with the website. No prompts or reminders were used on the website or in the app.

Figure 2. Intervention material (PRINT).



In tandem with the 10-week PRINT or WEB and WEB+ interventions, all 3 intervention groups were offered weekly face-to-face group sessions (facilitated by trained student assistants) with up to 25 participants per group, who were encouraged to attend the sessions. The 90-minute group sessions included performing the exercises in groups and going for joint walks and discussing weekly health education topics, and the participants were encouraged to ask open questions regarding the exercises (also see the study protocol by Pischke et al [30]). During their first weekly group meeting, participants received the necessary equipment (printed material or access information for the website and a Fitbit device) and a comprehensive introduction on how to use the equipment and materials. After 10 weeks, group meetings were continued monthly for another 6 months. After the last weekly group sessions, participants chose to continue using the material from their intervention group or to start using material from one of the other groups (Figure 1).

Outcomes and Measures

Data Collection

Two weeks before the intervention started, at the introductory event, participants received the questionnaires for the baseline assessment (T0) and were instructed on how and when to wear the accelerometer to measure their baseline PA behavior. They were asked to bring this data collection material to their first weekly group session, where they completed a short version of the MMSE-2-BV, which was conducted individually in a separate room by research staff [31]. During the first sessions of the 3- and 9-month follow-ups (T1 and T2), study participants received the data collection material in person from the research staff and were asked to send it back within 1 week via mail.

Sociodemographic and Baseline Variables

Table 1 presents the outcome measures, validated instruments, and assessment times. Sociodemographic information (eg, age, sex, family status, and employment) was collected via self-administered questionnaires at baseline as summarized in Table 1. Need-weighted household income per capita was

derived from the number of individuals living in the household and the monthly household income according to the German Microcensus [36]. The variable was then classified into low-, middle-, and high-income households. Education level was coded using the 2011 version of the International Standard Classification of Education (ISCED). Individuals with higher educational status received higher scores (range 1-8) [37]. The variable was dichotomized into low or medium level of education (ISCED score 1-4) and high level of education (ISCED score 5-8). BMI was calculated based on the self-reporting of height (T0) and weight (assessed at all time points) and dichotomized into underweight or normal weight and overweight or obese according to the WHO BMI classification for adults aged ≥20 years [38]. In addition, neighborhood, subjective general health (excellent or very good, good, less good, or poor), activity-related social support by family and friends, technology readiness (technology acceptance), technology competence beliefs (consisting of acceptance, competence belief, technology control belief, and technology willingness or readiness), and ownership and frequency of use of digital devices were measured.

Table 1. Selected measures in the self-administered study questionnaire used in the analysis for this study.

Outcome measure	Instrument or scale[a]	Time of assessment
Sociodemographic information (sex, age, education, family status, employment status, and household income)	German Health Interview and Examination Survey for Adults, questionnaire for assessing seniors' demographic and sociostructural data in Germany	T0[b]
Height and weight	Self-generated items	T0
Physical activity and neighborhood environment	Physical activity neighborhood environment scale	T0
Walking environment	Neighborhood Scales, walking environment, 1 Item (activity friendly score)	T0
Social support for engaging in physical activity	Activity-related support by family and friends (modified) and activity-related social support	T0, T1,[c] T2[d]
Subjective health status	Short-Form–12, 1 item	T0, T1, T2
Technology commitment	Technology commitment scale	T0, T1, T2
Technology use and experience	Self-generated items	T0, T1, T2
Use and acceptance of various components of the interventions (website and printed material), attendance of the offered group sessions, and overall satisfaction with the interventions	Self-generated items	T1, T2
Preference regarding intervention material at baseline (retrospective)	Self-generated items	T2
Reasons for crossing over or not crossing over after 3 months	Self-generated items	T2

[a]References for the instruments can be found in the study protocol [30].

[b]T0: baseline assessment.

[c]T1: 3-month follow-up.

[d]T2: 9-month follow-up.

PA and SB Outcomes

The main outcomes were MVPA and SB in minutes per day assessed at T0, T1, and T2 using triaxial accelerometers (GT3x+ [ActiGraph]). Participants were instructed to wear the accelerometer at the right hip over a course of 7 days for 24 hours. Accelerometer data were processed using the Actilife 6.8.0 software (ActiGraph) and R (version 3.6.1; R Foundation for Statistical Computing) [39] was used to identify nonwear times and classify PA levels into the categories described below.

Valid wear time was derived using the wear- and nonwear time classification algorithm by Choi et al [40], using a 90-minute window of consecutive zeros allowing a 2-minute interval of nonzero counts, and valid days were defined as having at least 8 hours (480 minutes) of valid wear time. There had to be at least three valid days available for each participant, including 1 weekend day, for the analysis. Using 1-second epochs, counts were categorized into SB (0-99 counts per minute [cpm]), as well as light (0-2690 cpm), moderate (2691-6166 cpm), vigorous (6167-9642 cpm), moderate to vigorous (2691-9642 cpm), and highly vigorous (>9642 cpm) PA, according to Sasaki et al [41], considering the vector magnitude.

The daily minutes for MVPA and SB were determined by dividing the total minutes by the number of days the accelerometer was worn. SB was additionally calculated in bouts of at least 30 minutes, and time spent with MVPA was calculated in bouts of at least 10 minutes. Minutes per week for MVPA and SB in the mentioned bouts were derived by multiplying the daily average minutes in 10-minute or 30-minute bouts, respectively, by 7. Furthermore, minutes of MVPA per week in bouts of 10 minutes was dichotomized as meeting the WHO recommendation (≥150 minutes per week of MVPA in bouts of at least 10 minutes) or not meeting them. The season during the accelerometer measurement was derived from the date of examination and categorized into autumn or winter for

the months of October to February and spring or summer for the months of March to September.

Adherence, Use, and Acceptance

Information on acceptance of the group sessions and intervention material was assessed with self-generated items (eg, frequency of general use, use of different components [on a 5-point Likert scale ranging from *never* to *daily*], and perceived helpfulness of intervention components [on 5-point Likert scales ranging from *not helpful at all* to *very helpful*]). The reasons for dropping out of the study and for crossing or not crossing over to the other intervention groups and preferences for intervention material were also assessed (Table 1).

Statistical Analyses

Descriptive statistics, that is, mean, SD, range, or proportion, were calculated to describe the study characteristics across intervention groups and surveys. The effects of time, group, and time by group on MVPA and SB, either in average minutes per day or minutes in bouts per week, were examined using multivariate linear mixed models that can handle unbalanced longitudinal data with varying numbers of repeated measurements per participant [42]. Analyses were adjusted for sex, age, BMI classification, level of education, family status, employment status, household income, subjective health status, built environment, activity-related support, preference, season, and valid wear time. Model diagnostics, such as residual plots and *Q*–*Q* plots, were used to check the assumptions of the linear mixed models. No violation of the assumptions of the linear mixed models was observed. In addition, outliers were checked and found to be unproblematic.

As cell counts for crossover groups were very small, linear mixed models were not run for the potential crossover combinations at the 3-month follow-up. The analyses regarding the intervention groups in this study were calculated using the group allocation at baseline as the indicator of the intervention group (ie, all analyses were conducted using the originally assigned groups). Only the numbers and proportions of individuals in the crossover combinations were reported descriptively. In addition, information assessed at follow-up (eg, preferences and reasons for crossing over or not crossing over to the other mode of delivery) and indicators of intervention adherence and acceptance were calculated. All statistical analyses were performed using SPSS 26 (IBM) [43] and SAS 9.4 [44], where the GLIMMIX procedure was used particularly for linear mixed modeling.

Results

Participant Flow and Baseline Characteristics

Of the 3660 older adults invited, 823 (22.49%) individuals were assessed for eligibility during computer-assisted telephone interviews (Figure 1). Of the 242 initially recruited participants, 91 (37.6%) were randomized to the WEB group; 38 (15.7%), to the WEB+ group; and 113 (46.7%), to the PRINT group. After 3 months, 80.6% (195/242) of the participants completed T1 (from the original group allocation; WEB: 74/91, 81%; WEB+: 30/36, 83%; and PRINT: 91/113, 80.5%). After T1, 91.8% (179/195) of the participants chose to remain in their previous intervention group, and 8.2% (16/195) decided to crossover to the other group. Finally, 66.1% (160/242) of participants completed T2 (from the original group allocation; WEB: 59/91, 65%; WEB+: 22/38, 58%; and PRINT: 79/113, 69.9%). Attrition rates from baseline to T2 were 33.9% across the groups (WEB: 35.2%, WEB+: 42.1%, and PRINT: 30.1%).

Observations from participants (T0, T1, and T2) were excluded from the analysis if they were missing information on BMI (14/501, 2.8%), subjective health, family status, or education (34/501, 6.8%), and if the MMSE-2-BV score was <13 (51/501, 10.2%). In total, 501 observations from 204 participants were included in the analysis (PRINT: 90/204, 44.1%; WEB: 78/204, 38.2%; and WEB+: 36/204, 17.7%). For follow-up samples, the exclusion criteria reduced the sample sizes to 159 at T1 and 138 at T2.

The baseline demographic characteristics of the participants included in the analysis are shown in Multimedia Appendix 1. Overall, the mean age was 68.7 (SD 5.4, range 60-82) years, with a slightly higher average age in the WEB+ group (70.5, SD 6.0 years). Fewer than half of the participants (87/204, 42.6%) had a BMI in the underweight or normal weight range according to the WHO standards. Across all groups, except for the WEB+ group, women were overrepresented. The proportion of female participants slightly differed among the study groups (PRINT: 75%, WEB: 64%, and WEB+: 47%). In the total sample of 204 participants, 112 (54.9%) had a high level of education, 110 (53.9%) were married, and 136 (66.7%) and 30 (14.7%) reported good and very good health, respectively. Participants rated their acceptance as average (mean 2.7, SD 0.83), and their competence beliefs (mean 3.9, SD 0.83), control beliefs (mean 3.9, SD 0.77), and overall willingness to deal with new technologies (mean 3.5, SD 0.63) with stronger agreement. The recommended level of MVPA was 12% (11/90) of the participants in the PRINT group, 23.1% (18/78) of the participants in the WEB group, and 25% (9/36) of the participants in the WEB+ group. Baseline differences were accounted for by including relevant variables as covariates in linear mixed models.

We analyzed potential selectivity by calculating Cohen *d* using the mean difference and pooled SD between the recruited and analyzed samples for continuous baseline characteristics. Cohen *h* was calculated based on the proportions of categorical baseline characteristics [45]. The analysis sample (n=204) did not differ from the recruited sample (n=242) in baseline characteristics as the effect sizes (Cohen *d* and *h*, respectively) were all <0.20. The only exception was the cognitive state: the analysis sample had a slightly higher MMSE-2-BV mean score compared with the recruited sample (Cohen *d*=0.22). This was expected because of the exclusion criteria (Figure 1).

PA and SB Outcomes

Overall, the proportion of individuals reaching the MVPA recommendation did not change over time; 18.6% (38/204) of them reached the WHO recommendation at baseline: 16.3% (26/159) at T1 and 20.2% (28/138) at T2 (Multimedia Appendix 2). In all 3 intervention groups, MVPA in minutes per day seemed to increase between baseline and T1: from 84.4 (SD 33.0) to 92.3 (SD 31.5) minutes in WEB, from 84.4 (SD 39.7)

to 95.5 (SD 37.7) minutes in WEB+, and from 82.8 (SD 29.2) to 84.3 (SD 26.1) minutes in PRINT (Table 2). When adjusting for covariates, the least squares mean differences in time between baseline and T1 within the intervention groups were not significant. There was a significant decrease between baseline and T2 in the whole study sample by 9 minutes of MVPA per day (β_{time}=−9.37, 95% CI −18.58 to −0.16). Within the groups, the least squares mean decrease between baseline and T2 was significant for WEB (mean difference −13.12, 95% CI −23.40 to −2.84) and PRINT (mean difference −9.37, 95% CI −18.58 to −0.16; Table 2). Compared with PRINT, there were no significant group differences and group-by-time interactions at T1 or T2 (Table 2). Compared with PRINT and baseline, the WEB group at T2 was approximately 4 minutes per day less active in MVPA ($\beta_{group*time}$=3.76, 95% CI −13.33 to 5.82) and the WEB+ group at T2 was approximately 1 minute per day more active in MVPA ($\beta_{group*time}$=1.40, 95% CI −11.04 to 13.83).

Table 2. Results of the linear mixed models (time, group, intervention effects, and comparison of intervention effects) for moderate to vigorous physical activity (MVPA; minutes per day and 10-minute bouts).[a]

Characteristics	Indicators per time point, mean (SD)			Difference in time within group (reference T0[b]), least squares mean (95% CI)		Time difference (reference T0), β (95% CI)		Group difference (reference PRINT), β (95% CI)	Group-by-time interaction (reference PRINT at T0), β (95% CI)	
	T0	T1[c]	T2[d]	T1	T2	T1	T2		T1	T2
MVPA (minutes per day)										
WEB	84.4 (33.0)	92.3 (31.5)	81.3 (31.6)	2.68 (−6.43 to 11.78)	−13.12 (−23.40 to −2.84)	−1.46 (−9.87 to 6.95)	−9.37 (−18.58 to −0.16)	4.90 (−4.49 to 14.30)	4.13 (−4.69 to 12.96)	−3.76 (−13.33 to 5.82)
WEB+	84.4 (39.7)	95.5 (37.7)	90.0 (36.7)	1.98 (−9.90 to 13.87)	−7.97 (−20.26 to 4.32)	N/A[e]	N/A	8.73 (−3.44 to 20.89)	3.44 (−7.74 to 14.62)	1.40 (−11.04 to 13.83)
PRINT	82.8 (29.2)	84.3 (26.1)	80.8 (28.8)	−1.46 (−9.87 to 6.95)	−9.37 (−18.58 to −0.16)	N/A	N/A	N/A	N/A	N/A
MVPA in 10-minute bouts (minutes per week)										
WEB	89.1 (121.2)	79.1 (106.6)	78.2 (108.4)	−17.06 (−57.94 to 23.83)	−49.64 (−95.83 to −3.45)	−2.28 (−40.18 to 35.62)	−37.7 (−78.87 to 3.48)	18.51 (−16.50 to 53.52)	−14.78 (−55.32 to 25.77)	−11.94 (−55.96 to 32.08)
WEB+	98.3 (139.5)	119.8 (165.3)	86.2 (130.0)	15.76 (−38.31 to 69.83)	−53.48 (−109.47 to 2.51)	N/A	N/A	24.54 (−20.81 to 69.89)	18.04 (−33.38 to 69.45)	−15.78 (−72.90 to 41.34)
PRINT	76.9 (108.9)	70.0 (87.1)	73.3 (81.0)	−2.28 (−40.18 to 35.62)	−37.7 (−78.87 to 3.48)	N/A	N/A	N/A	N/A	N/A

[a]The linear mixed model was adjusted for age, sex, BMI, level of education, family status, employment status, household income, subjective health status, built environment, activity-related support, preference, season, and valid wear time.

[b]T0: baseline assessment.

[c]T1: 3-month follow-up.

[d]T2: 9-month follow-up.

[e]N/A: not applicable.

With regard to MVPA in 10-minute bouts in minutes per week, only the WEB+ group seemed to be more active at T1 with mean 119.8 (SD 165.3) minutes compared with baseline with mean 98.3 (SD 139.5) minutes (Table 2). However, the estimated difference in time within WEB+ was not significant (least squares mean difference 15.76, 95% CI −38.48 to 69.83). At T2, there was a significant estimated mean decrease in minutes of MVPA in 10-minute bouts per week within the WEB (least squares mean difference −49.64, 95% CI −95.83 to −3.45). Compared with PRINT, there were no significant group differences and group-by-time interactions at T1 or T2 (Table 2).

For sedentary time in 30-minute bouts in minutes per week, there was a significant estimated decrease at T1 within the WEB group (mean difference −212.00, 95% CI −422.14 to −3.84). However, the CI was very wide and the effect was not maintained until T2 (Table 3). Compared with the PRINT group, there were no significant group differences and group-by-time interactions at T1 or T2 (Table 3). At T2, there were no significant intergroup differences over time.

At all 3 time points, the mean sedentary time per day was between 600 and 700 minutes, that is, between 10 and 11.5 hours (minimum 420 minutes, maximum 1200 minutes). There were no significant within-group differences at T1 and T2 with regard to sedentary time in minutes per day after adjusting for

covariates. There were no significant group differences or group-by-time interactions at T1 or T2 (Table 3). There were no significant within-group differences over time at T2 (β_{time}=5.29, 95% CI −9.12 to 19.69). Compared with PRINT and baseline groups, the WEB group at T2 spent approximately 10 more minutes per day with sitting ($\beta_{group*time}$=10.41, 95% CI −4.49 to 25.31) and the WEB+ group at T2 spent approximately the same minutes per day with sitting ($\beta_{group*time}$=−0.13, 95% CI −19.49 to 19.22; Table 3).

Table 3. Results of the linear mixed regression models (time, group, intervention effects, and comparison of intervention effects) for sedentary behavior (minutes per day and 30-minute bouts).[a]

Characteristics	Indicators per time point, mean (SD)			Difference in time within group (reference T0[b]), least squares mean (95% CI)		Time difference (reference T0), β (95% CI)		Group difference (reference PRINT), β (95% CI)	Group-by-time interaction (reference PRINT at T0), β (95% CI)	
	T0	T1[c]	T2[d]	T1	T2	T1	T2		T1	T2
Sedentary time (minutes per day)										
WEB	630.2 (102.2)	633.2 (90.9)	638.1 (88.2)	−5.20 (−19.43 to 9.02)	15.7 (−0.36 to 31.76)	1.04 (−12.09 to 14.17)	5.29 (−9.12 to 19.69)	−5.88 (−21.14 to 9.39)	−6.24 (−19.98 to 7.49)	10.41 (−4.49 to 25.31)
WEB+	637.7 (74.5)	642.6 (79.0)	628.9 (94.1)	−10.83 (−29.36 to 7.69)	5.15 (−14.00 to 24.31)	N/A[e]	N/A	−2.53 (−22.31 to 17.25)	−11.88 (−29.27 to 5.52)	−0.13 (−19.49 to 19.22)
PRINT	639.0 (78.3)	649.9 (128.5)	646.8 (120.6)	1.04 (−12.09 to 14.17)	5.29 (−9.12 to 19.69)	N/A	N/A	N/A	N/A	N/A
Sedentary time in 30-minute bouts (minutes per week)										
WEB	2228.1 (905.2)	2098.7 (851.1)	2348.2 (731.5)	−212.99 (−422.14 to −3.84)	179.25 (−56.85 to 415.35)	−18.83 (−211.94 to 174.29)	72.31 (−139.36 to 283.98)	−107.81 (−327.33 to 111.70)	−194.16 (−396.46 to 8.14)	106.94 (−112.63 to 326.51)
WEB+	2368.4 (839.3)	2336.5 (785.7)	2402.2 (701.7)	−160.20 (−432.82 to 112.43)	72.65 (−209.29 to 354.59)	N/A	N/A	−16.37 (−300.69 to 267.95)	−141.37 (−397.69 to 114.96)	0.34 (−284.79 to 285.46)
PRINT	2301.3 (798.9)	2371.8 (960.6)	2402.1 (900.5)	−18.83 (−211.94 to 174.29)	72.31 (−139.36 to 283.98)	N/A	N/A	N/A	N/A	N/A

[a]The linear mixed model was adjusted for age, sex, BMI, level of education, family status, employment status, household income, subjective health status, built environment, activity-related support, preference, season, and valid wear time.

[b]T0: baseline assessment.

[c]T1: 3-month follow-up.

[d]T2: 9-month follow-up.

[e]N/A: not applicable.

Attendance, Use, and Acceptance of Intervention Components

Overall, attendance of the face-to-face components of the intervention was high, with an average of 8/10 weekly group sessions attended and 2/3 monthly group sessions attended (Multimedia Appendix 3). Regarding the use of the PA diary, there were no marked differences among the intervention groups at T1. The exercise brochure was used at least once per week or daily by 68% (50/73), 48% (29/60), and 38% (10/26) of the PRINT, WEB, and WEB+ groups, respectively. The overall use of intervention material was high to moderate at T1 (the PA diary was used by between 65% (17/26) and 78% (47/60) of the participants at least once per week or daily) and declined by T2. At T2, approximately 44% (24/55) of the participants in the PRINT group, 49% (23/47) in the WEB group, and 58% (19/36) in the WEB+ group still used the PA diary at least once per week or daily. The use of the smartphone app was very low in the WEB group but higher in the WEB+ group.

Acceptance of the interventions was generally high; approximately half of the participants agreed that the program was at least somewhat helpful for being physically active (T1), and stated that they would recommend it to others (T1; Multimedia Appendix 3). Retrospectively, between 73% (49/55) and 78% (39/47) of the participants in each group stated that their random allocation matched their initial preference (T2; Multimedia Appendix 1). The most commonly reported reasons for crossing over were “wanted to try something new,” “liked the website and wanted to use the fitness tracker in addition,” and “high affinity to technology.” The reason for not crossing over reported most often was “completely satisfied with the current material.” Further reasons listed included “did not want to lose contact to the previous group members,” “printed version

seemed impractical," or "not technology-affine and wanted to keep the printed version." No unintended effects were reported by the participants.

Discussion

Principal Findings

In summary, no intervention effects on MVPA were detected in this study, including 242 community-dwelling older adults aged ≥60 years who participated in a 9-month crossover randomized trial. MVPA did not increase but decreased over time, regardless of which group the participants were randomized to. The proportion of participants meeting the WHO recommendations for MVPA remained relatively stable, with approximately one-fifth of the participants meeting the recommendations at all 3 assessment points. The use of the intervention materials decreased slightly over time. Regarding SB, all 3 intervention groups displayed a decreasing trend in this risk behavior over time. However, no significant intergroup differences were observed in this regard. Interestingly, however, there was an indication that the reduction was most pronounced in the WEB group, which had decreased sedentary time in 30-minute bouts in minutes per week from baseline to T1. However, this effect was not maintained at follow-up, and no significant time-by-group interactions were observed. This study adds to the current knowledge that the mode of delivery (PRINT vs WEB) did not appear to affect the acceptance and effectiveness of the intervention content. Both were comparable across the groups. Unfortunately, however, none of the intervention conditions displayed increases in PA over the course of 9 months (for a comparison with previous research, see the following section).

It is conceivable that the intervention effects in certain subgroups were masked in the primary analyses. In an exploratory analysis, the presence of unobserved subgroups was investigated with regard to the latent change trajectories of MVPA and SB [46]. Regarding MVPA, latent change trajectory analysis revealed an initially sufficiently active and an initially insufficiently active subgroup, both of which remained constant over time. Regarding SB, an initially highly sedentary subgroup and moderately sedentary subgroup were identified. Although the moderately sedentary subgroup experienced slight increases in sitting time, the initially highly sedentary subgroup experienced significant decreases in SB and significant increases in PA levels [46]. This may suggest that our interventions were particularly useful for older adults with high initial SB levels.

Second, we found that despite having had the opportunity to try out another condition at the 6-month follow-up, very few participants took advantage of it. In total, only approximately 7.8% (16/204) switched conditions at follow-up (1/17, 6% from WEB [including WEB+] to PRINT; 15/17, 88% from PRINT to WEB). Two-thirds of participants in each condition stated at follow-up that their random allocation had matched their initial preference, suggesting that intervention participants felt content with the condition (and use of materials) to which they were assigned. In addition, feeling part of a group during the sessions and not wanting to leave the group may have played a role. Furthermore, 10.3% (21/204) of the participants stated that they did not want to lose contact with their groups. In fact, overall satisfaction with the group sessions and attendance rates were very high and did not differ among the modes of delivery. To conclude, these results suggest that the participants did not see any reason for switching to another mode of delivery. However, another explanation for the lack of movement between conditions at T1 could be that study participants who were randomized to a specific condition either refused to participate in the study or possibly dropped out of the study early, because they felt dissatisfied with using the intervention materials assigned to them in the condition that they were randomized to. Future pragmatic trials combining a randomized controlled trial with 2 different intervention arms in which participants can self-select will be necessary to further investigate the questions regarding the role of individual preferences raised in this study.

Strengths and Limitations

Despite the advantages of the study design applied in this project and the objective measurement of MVPA using accelerometers, this study had several limitations. First, there was no untreated or placebo control group. Second, the preference for a certain intervention delivery mode at baseline was only assessed retrospectively, which may have led to a recall bias. However, we chose to assess preferences retrospectively because we anticipated disappointment (and possibly study dropout); if an individual was not randomized to the intervention group that they preferred to be in at baseline. Another question that remains is whether the recruitment channel that participants were recruited via (ie, print media vs mailed invitations) played a role when deciding for or against an intervention condition at the 6-month follow-up. This will be the topic of a future study. Finally, we were unable to recruit the number of study participants to the study that we had aimed for [30], and loss-to-follow-up was relatively high. As we did not meet the intended goal regarding the sample size, our analyses were underpowered. This problem was addressed using linear mixed modeling. Another limitation of our study is that the primary outcome defined in this study was not a state-of-the-art recommendation at the time of study completion. Thus, participants may not have been sufficiently motivated to engage in activities amounting to >10 minutes because, on the website or using PRINT materials for self-monitoring PA, they could only complete the PA diary if the activities leveled up to 10-minute bouts. In addition, we were unable to quantify the individual intervention effects of the group sessions on PA. Finally, neither participants nor researchers were blinded to the conditions, design, and aim of the study.

Several issues potentially causing selection bias and high attrition were identified in our study. One of the exclusion criteria was that participants had to own a PC with internet access. It is possible that individuals without this equipment were disadvantaged because we could not provide the equipment to participate in the study to them. Another selection bias may be the appointments made available to them. In particular, individuals who were still employed found it difficult to keep the appointments assigned to them and may have dropped out consequently. Furthermore, as discussed previously, despite the exclusion criteria already meeting the recommendation of 150 minutes per week, we found that many participants had higher

levels of PA than were reported while entering the study and did not feel sufficiently challenged by our intervention. All of these issues have been previously discussed in the context of individual studies and meta-analyses conducted on the topics of selection, retention, and dropout in behavior change trials [47-49].

Comparison With Previous Work

Compared with previous research suggesting that eHealth interventions can effectively promote PA in older adults in the short term [1,11-13,50-53], we were not able to demonstrate intervention effects for participants in the WEB and WEB+ conditions, neither in the short term at 3 months nor in the longer term at 9 months. This is puzzling, as several recently published systematic reviews and meta-analyses [50-53] demonstrate the effectiveness of eHealth interventions, including mobile interventions [52], for improving PA levels (eg, mean steps per day and minutes of daily MVPA, weekly PA, and MVPA [50]) in predominantly healthy older adults. Similar to our study, acceptance of eHealth intervention approaches was high in the studies included in these systematic reviews and meta-analyses [50]; however, contrary to our attrition rate, studies included in [50] predominantly reported attrition below 20% [50].

Looking at the evidence for the lack of effectiveness of eHealth interventions only, Elavsky et al [52] reported that 12 out of 29 randomized controlled trials and 8 out of 21 trials reporting pre-post changes in PA did not find any significant increases in PA. However, most studies that did not demonstrate effects did not include print-based conditions as a comparison group but control groups not receiving an intervention [52]. Hence, they could not be compared with the results of our study. Furthermore, in our study, we did not observe any differential effects of the intervention modality (WEB or WEB+ vs PRINT). A systematic review conducted by Muellmann et al [13] included several studies comparing print- and web-based intervention arms with contradictory results. Two studies by Peels et al [54,55] revealed that print- and web-based interventions were equally effective in promoting PA at the 6-month follow-up, but at the 12-month follow-up, only participation in the print-based interventions was associated with significant changes in PA. Van Stralen et al [56-58] compared a web-based intervention to a no-intervention control group and print-based intervention. Contrary to the findings by Peels et al [54,55], participants receiving the print-based intervention did not display any increases in PA at the 12-month follow-up, whereas participants who received the web-based intervention did. The contents of both WEB and PRINT interventions in our study were very similar and according to the evidence described above, could or could not have resulted in an intervention effect. The fact that participants in all intervention arms also received regular sessions in groups may have served as an equalizer, but it cannot explain the lack of an overall intervention effect. A possible reason for the lack of an effect may be that, despite the exclusion criterion "already meeting the WHO recommendations for one year preceding baseline," potential participants may have underreported MVPA to be able to participate in the study. This may have led to higher baseline PA levels than intended and feelings of frustration with intervention messages and materials targeting primarily inactive adults and possibly dropping out of the study. It is also conceivable that potential participants may have overreported PA because of social desirability and may have been excluded, leading to a lack of representation of physically inactive adults in the sample. However, both of these potential explanations are rather speculative. Nevertheless, we conclude that more sensitive strategies are needed to address social desirability concerning the self-reporting of PA during recruitment to better reach initially inactive adults. Owing to the reasons explained above, the results of our study are not generalizable to the general population in this age bracket, and external validity is limited.

Furthermore, previous research has shown that preferences for intervention modality may vary by age, sex, BMI, or social or living environment [15,22,23]. Younger individuals seem to prefer eHealth to print-based interventions [22,23], whereas older or female individuals or those with an adverse weight status appear to be more likely to favor print-based interventions [22]. Unfortunately, in our study, we were not able to investigate variations by sociodemographic characteristics because only 7% of the sample changed groups at follow-up, and preference was not assessed at baseline, but only retrospectively. However, we did not find any group differences in terms of technology readiness. Levels of acceptance were average across groups and competence and control beliefs, as well as willingness to deal with new technologies, were high across groups, suggesting no variations that may have affected decisions for WEB versus PRINT conditions later on in the study.

Implications for Practice and Research

The great heterogeneity among older adults >60 years (eg, in employment status, chronic disease status, or functional capability) is a key concern and needs to be addressed in future interventions, including differing motivations to participate in the study or to engage in interventions (eg, maintaining functional status). Interventions should include more tailoring in the future, including tailored messages addressing the aspects raised above. One lesson learned in this study was that group sessions paralleling eHealth intervention components contribute to acceptance in this target group and may prevent study dropout. Face-to-face contact with the PA instructor and fellow participants and a sense of structure because of regular weekly meetings were well received by participants in our study. More than two-thirds (131/159, 82.4%) of the participants across groups stated at T1 and T2 that they found the group sessions very or somewhat helpful.

Conclusions

In conclusion, we successfully adapted and simplified the interventions developed in a previous study. Despite the high acceptance and use of these interventions, no intervention effect was observed for MVPA. Owing to a lack of movement between groups at T1, the role of personal preferences for different delivery modes could not be investigated in full depth.

Acknowledgments

The authors would like to thank all study participants for their participation in the interventions and assessments. They would like to thank Beate Schütte, the study nurse, for her efforts during the recruitment and data collection phases of the study.

This study was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; project 01EL1822A, 01EL1822F, 01EL1822l, and 01EL1822C). The content of this paper reflects only the authors' views, and the funder is not liable for any use that may be made of the information contained therein.

Authors' Contributions

CRP, MP, and TR drafted the manuscript. CRP, MP, TR, SL, CVR, CB, JM, and KvH contributed substantially to the conception and design of the study. CB conducted the statistical analyses. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics.
[DOCX File , 35 KB - mhealth_v10i3e32212_app1.docx]

Multimedia Appendix 2
World Health Organization recommendations reached.
[DOCX File , 13 KB - mhealth_v10i3e32212_app2.docx]

Multimedia Appendix 3
Adherence, use, and acceptance of the intervention components at the three- and nine-month follow-ups.
[DOCX File , 31 KB - mhealth_v10i3e32212_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 12143 KB - mhealth_v10i3e32212_app4.pdf]

References

1. Hong SY, Hughes S, Prohaska T. Factors affecting exercise attendance and completion in sedentary older adults: a meta-analytic approach. J Phys Act Health 2008;5(3):385-397. [doi: 10.1123/jpah.5.3.385] [Medline: 18579917]
2. Byberg L, Melhus H, Gedeborg R, Sundström J, Ahlbom A, Zethelius B, et al. Total mortality after changes in leisure time physical activity in 50 year old men: 35 year follow-up of population based cohort. Br J Sports Med 2009;43(7):482. [Medline: 19581403]
3. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
4. Global recommendations on physical activity for health. World Health Organization. 2010. URL: https://www.who.int/dietphysicalactivity/global-PA-recs-2010.pdf [accessed 2021-07-12]
5. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007;39(8):1435-1445. [doi: 10.1249/mss.0b013e3180616aa2] [Medline: 17762378]
6. Finger JD, Mensink G, Lange C, Manz K. Gesundheitsfördernde körperliche Aktivität in der Freizeit bei Erwachsenen in Deutschland. J Health Monit 2017;2(2):37-44. [doi: 10.17886/RKI-GBE-2017-027]
7. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
8. Lange C, Finger JD. Gesundheitsverhalten in Europa – vergleich ausgewählter Indikatoren für Deutschland und die Europäische Union. J Health Monit 2017;2(2):3-19. [doi: 10.17886/RKI-GBE-2017-024]
9. Haider S, Grabovac I, Dorner TE. Fulfillment of physical activity guidelines in the general population and frailty status in the elderly population : a correlation study of data from 11 European countries. Wien Klin Wochenschr 2019;131(11-12):288-293 [FREE Full text] [doi: 10.1007/s00508-018-1408-y] [Medline: 30421283]

10. Rütten A, Pfeifer K. Nationale Empfehlungen für Bewegung und Bewegungsförderung. FAU Erlangen-Nürnberg. 2016. URL: https://www.sport.fau.de/files/2016/05/Nationale-Empfehlungen-f%C3%BCr-Bewegung-und-Bewegungsf%C3%B6rderung-2016.pdf [accessed 2021-07-15]
11. Hagger MS, Weed M. DEBATE: do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
12. Amireault S, Baier JM, Spencer JR. Physical activity preferences among older adults: a systematic review. J Aging Phys Act 2019;27(1):128-139. [doi: 10.1123/japa.2017-0234] [Medline: 29283793]
13. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
14. Richards J, Hillsdon M, Thorogood M, Foster C. Face-to-face interventions for promoting physical activity. Cochrane Database Syst Rev 2013(9):CD010392. [doi: 10.1002/14651858.CD010392.pub2] [Medline: 24085592]
15. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
16. Short CE, James EL, Plotnikoff RC, Girgis A. Efficacy of tailored-print interventions to promote physical activity: a systematic review of randomised trials. Int J Behav Nutr Phys Act 2011;8:113 [FREE Full text] [doi: 10.1186/1479-5868-8-113] [Medline: 21999329]
17. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
18. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
19. Zubala A, MacGillivray S, Frost H, Kroll T, Skelton DA, Gavine A, et al. Promotion of physical activity interventions for community dwelling older adults: a systematic review of reviews. PLoS One 2017;12(7):e0180902 [FREE Full text] [doi: 10.1371/journal.pone.0180902] [Medline: 28700754]
20. Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: threat or opportunity? J Health Serv Res Policy 1996;1(4):194-197. [doi: 10.1177/135581969600100403] [Medline: 10180870]
21. Preference Collaborative Review Group. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. BMJ 2008;337:a1864 [FREE Full text] [doi: 10.1136/bmj.a1864] [Medline: 18977792]
22. Short CE, Vandelanotte C, Duncan MJ. Individual characteristics associated with physical activity intervention delivery mode preferences among adults. Int J Behav Nutr Phys Act 2014;11(1):25 [FREE Full text] [doi: 10.1186/1479-5868-11-25] [Medline: 24568611]
23. Boekhout JM, Peels DA, Berendsen BA, Bolman C, Lechner L. A web-based and print-delivered computer-tailored physical activity intervention for older adults: pretest-posttest intervention study comparing delivery mode preference and attrition. J Med Internet Res 2019;21(8):e13416 [FREE Full text] [doi: 10.2196/13416] [Medline: 31464186]
24. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
25. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
26. Forberger S, Bammann K, Bauer J, Boll S, Bolte G, Brand T, et al. How to tackle key challenges in the promotion of physical activity among older adults (65+): the AEQUIPA network approach. Int J Environ Res Public Health 2017;14(4):379 [FREE Full text] [doi: 10.3390/ijerph14040379] [Medline: 28375177]
27. Muellmann S, Buck C, Voelcker-Rehage C, Bragina I, Lippke S, Meyer J, et al. Effects of two web-based interventions promoting physical activity among older adults compared to a delayed intervention control group in Northwestern Germany: results of the PROMOTE community-based intervention trial. Prev Med Rep 2019;15:100958 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100958] [Medline: 31410347]
28. 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;17(1):512 [FREE Full text] [doi: 10.1186/s12889-017-4446-x] [Medline: 28545506]
29. Ratz T, Lippke S, Muellmann S, Peters M, Pischke CR, Meyer J, et al. Effects of two web-based interventions and mediating mechanisms on stage of change regarding physical activity in older adults. Appl Psychol Health Well Being 2020;12(1):77-100. [doi: 10.1111/aphw.12174] [Medline: 31332957]
30. Pischke CR, Voelcker-Rehage C, Peters M, Ratz T, Pohlabeln H, Meyer J, et al. Implementation and effects of information technology-based and print-based interventions to promote physical activity among community-dwelling older adults:

protocol for a randomized crossover trial. JMIR Res Protoc 2020;9(4):e15168 [FREE Full text] [doi: 10.2196/15168] [Medline: 32338622]
31. Folstein MF, Folstein SE, White T, Messer MA. MMSE-2®: mini-mental state examination. 2nd edition. Lutz, FL: Psychological Assessment Resources; 2010.
32. Finger RP, McSweeney SC, Deverell L, O'Hare F, Bentley SA, Luu CD, et al. Developing an instrumental activities of daily living tool as part of the low vision assessment of daily activities protocol. Invest Ophthalmol Vis Sci 2014;55(12):8458-8466. [doi: 10.1167/iovs.14-14732] [Medline: 25425306]
33. Van de Winckel A, De Patre D, Rigoni M, Fiecas M, Hendrickson TJ, Larson M, et al. Exploratory study of how cognitive multisensory rehabilitation restores parietal operculum connectivity and improves upper limb movements in chronic stroke. Sci Rep 2020;10(1):20278 [FREE Full text] [doi: 10.1038/s41598-020-77272-y] [Medline: 33219267]
34. Fleig L, Lippke S, Pomp S, Schwarzer R. Intervention effects of exercise self-regulation on physical exercise and eating fruits and vegetables: a longitudinal study in orthopedic and cardiac rehabilitation. Prev Med 2011;53(3):182-187. [doi: 10.1016/j.ypmed.2011.06.019] [Medline: 21784096]
35. Pomp S, Fleig L, Schwarzer R, Lippke S. Effects of a self-regulation intervention on exercise are moderated by depressive symptoms: a quasi-experimental study. Int J Clin Health Psychol 2013;13(1):1-8. [doi: 10.1016/S1697-2600(13)70001-2]
36. Boehle M. Armutsmessung mit dem Mikrozensus: methodische Aspekte und Umsetzung für Querschnitss- und Trendanalysen. GESIS Papers 2015/2016. Köln, Germany: Leibniz-Insitut für Sozialwissenschaften; 2015.
37. Demographische standards. Statistisches Bundesamt. Wiebaden, Germany: Statistisches Bundesamt; 2016. URL: https://www.destatis.de/DE/Methoden/Demografische-Regionale-Standards/textbaustein-demografische-standards.html [accessed 2022-03-03]
38. Body mass index - BMI. World Health Organization. URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi; [accessed 2021-11-08]
39. R: a language and environment for statistical computing. R Foundation for Statistical Computing. R Core Team. 2019. URL: https://www.R-project.org/ [accessed 2022-03-03]
40. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
41. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
42. Tilling K, Macdonald-Wallis C, Lawlor DA, Hughes RA, Howe LD. Modelling childhood growth using fractional polynomials and linear splines. Ann Nutr Metab 2014;65(2-3):129-138 [FREE Full text] [doi: 10.1159/000362695] [Medline: 25413651]
43. IBM SPSS statistics for Windows, version 26. IBM. 2019. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-26 [accessed 2022-03-03]
44. SAS 9.4. SAS. URL: https://support.sas.com/software/94/ [accessed 2022-03-03]
45. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
46. Ratz T, Pischke CR, Voelcker-Rehage C, Lippke S. Distinct physical activity and sedentary behavior trajectories in older adults during participation in a physical activity intervention: a latent class growth analysis. Eur Rev Aging Phys Act 2022;19(1):1 [FREE Full text] [doi: 10.1186/s11556-021-00281-x] [Medline: 34986783]
47. Willinger N, Steele J, Atkinson L, Liguori G, Jimenez A, Mann S, et al. Effectiveness of structured physical activity interventions through the evaluation of physical activity levels, adoption, retention, maintenance, and adherence rates: a systematic review and meta-analysis. J Phys Act Health 2021;18(1):116-129. [doi: 10.1123/jpah.2019-0349] [Medline: 33383569]
48. Carlisle S, Ayling K, Jia R, Buchanan H, Vedhara K. The effect of choice interventions on retention-related, behavioural and mood outcomes: a systematic review with meta-analysis. Health Psychol Rev 2021:1-37. [doi: 10.1080/17437199.2021.1962386] [Medline: 34423744]
49. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ 2013;346:e8668 [FREE Full text] [doi: 10.1136/bmj.e8668] [Medline: 23338004]
50. Núñez de Arenas-Arroyo S, Cavero-Redondo I, Alvarez-Bueno C, Sequí-Domínguez I, Reina-Gutiérrez S, Martínez-Vizcaíno V. Effect of eHealth to increase physical activity in healthy adults over 55 years: a systematic review and meta-analysis. Scand J Med Sci Sports 2021;31(4):776-789. [doi: 10.1111/sms.13903] [Medline: 33280182]
51. Kwan RY, Salihu D, Lee PH, Tse M, Cheung DS, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020;17:7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
52. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019;27(4):565-593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
53. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017;3:51 [FREE Full text] [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]

54. Peels DA, Bolman C, Golsteijn RH, de Vries H, Mudde AN, van Stralen MM, et al. Long-term efficacy of a printed or a web-based tailored physical activity intervention among older adults. Int J Behav Nutr Phys Act 2013;10:104 [FREE Full text] [doi: 10.1186/1479-5868-10-104] [Medline: 24228862]
55. Peels DA, van Stralen MM, Bolman C, Golsteijn RH, de Vries H, Mudde AN, et al. The differentiated effectiveness of a printed versus a web-based tailored physical activity intervention among adults aged over 50. Health Educ Res 2014;29(5):870-882. [doi: 10.1093/her/cyu039] [Medline: 24980023]
56. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The working mechanisms of an environmentally tailored physical activity intervention for older adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2009;6:83 [FREE Full text] [doi: 10.1186/1479-5868-6-83] [Medline: 19995412]
57. van Stralen MM, de Vries H, Bolman C, Mudde AN, Lechner L. Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. Ann Behav Med 2010;39(2):139-150. [doi: 10.1007/s12160-010-9166-8] [Medline: 20182833]
58. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. Health Psychol 2011;30(4):442-452. [doi: 10.1037/a0023579] [Medline: 21639638]

Abbreviations

cpm: counts per minute
ISCED: International Standard Classification of Education
MMSE-2-BV: Mini-Mental State Examination 2–brief version
MVPA: moderate to vigorous physical activity
PA: physical activity
SB: sedentary behavior
T0: baseline assessment
T1: 3-month follow-up
T2: 9-month follow-up
WHO: World Health Organization

Edited by L Buis; submitted 19.07.21; peer-reviewed by Y Xu, E Gonzalez, A McGuire; comments to author 13.09.21; revised version received 08.11.21; accepted 06.12.21; published 23.03.22.

Please cite as:
Pischke CR, Voelcker-Rehage C, Ratz T, Peters M, Buck C, Meyer J, von Holdt K, Lippke S
Web-Based Versus Print-Based Physical Activity Intervention for Community-Dwelling Older Adults: Crossover Randomized Trial
JMIR Mhealth Uhealth 2022;10(3):e32212
URL: https://mhealth.jmir.org/2022/3/e32212
doi:10.2196/32212
PMID:35319484

Original Paper

Evaluating the Impact of Adaptive Personalized Goal Setting on Engagement Levels of Government Staff With a Gamified mHealth Tool: Results From a 2-Month Randomized Controlled Trial

Raoul Nuijten[1], MSc; Pieter Van Gorp[1], PhD; Alireza Khanshan[2], MSc; Pascale Le Blanc[1], Prof Dr; Pauline van den Berg[3], PhD; Astrid Kemperman[3], PhD; Monique Simons[4], PhD

[1]Department of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[2]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands

[3]Department of the Built Environment, Eindhoven University of Technology, Eindhoven, Netherlands

[4]Department of Social Sciences, Wageningen University and Research, Wageningen, Netherlands

Corresponding Author:
Raoul Nuijten, MSc
Department of Industrial Engineering
Eindhoven University of Technology
Groene Loper 3
Eindhoven, 5612 AE
Netherlands
Phone: 31 040 247 2290
Email: r.c.y.nuijten@tue.nl

Abstract

Background: Although the health benefits of physical activity are well established, it remains challenging for people to adopt a more active lifestyle. Mobile health (mHealth) interventions can be effective tools to promote physical activity and reduce sedentary behavior. Promising results have been obtained by using gamification techniques as behavior change strategies, especially when they were tailored toward an individual's preferences and goals; yet, it remains unclear how goals could be personalized to effectively promote health behaviors.

Objective: In this study, we aim to evaluate the impact of personalized goal setting in the context of gamified mHealth interventions. We hypothesize that interventions suggesting health goals that are tailored based on end users' (self-reported) current and desired capabilities will be more engaging than interventions with generic goals.

Methods: The study was designed as a 2-arm randomized intervention trial. Participants were recruited among staff members of 7 governmental organizations. They participated in an 8-week digital health promotion campaign that was especially designed to promote walks, bike rides, and sports sessions. Using an mHealth app, participants could track their performance on two social leaderboards: a leaderboard displaying the individual scores of participants and a leaderboard displaying the average scores per organizational department. The mHealth app also provided a news feed that showed when other participants had scored points. Points could be collected by performing any of the 6 assigned tasks (eg, walk for at least 2000 m). The level of complexity of 3 of these 6 tasks was updated every 2 weeks by changing either the suggested task intensity or the suggested frequency of the task. The 2 intervention arms—with participants randomly assigned—consisted of a personalized treatment that tailored the complexity parameters based on participants' self-reported capabilities and goals and a control treatment where the complexity parameters were set generically based on national guidelines. Measures were collected from the mHealth app as well as from intake and posttest surveys and analyzed using hierarchical linear models.

Results: The results indicated that engagement with the program inevitably dropped over time. However, engagement was higher for participants who had set themselves a goal in the intake survey. The impact of personalization was especially observed for *frequency parameters* because the personalization of sports session frequency did foster higher engagement levels, especially when participants set a goal to improve their capabilities. In addition, the personalization of suggested ride duration had a positive effect on self-perceived biking performance.

Conclusions: Personalization seems particularly promising for promoting the frequency of physical activity (eg, promoting the number of suggested sports sessions per week), as opposed to the intensity of the physical activity (eg, distance or duration). Replications and variations of our study setup are critical for consolidating and explaining (or refuting) these effects.

Trial Registration: ClinicalTrials.gov NCT05264155; https://clinicaltrials.gov/ct2/show/NCT05264155

(JMIR Mhealth Uhealth 2022;10(3):e28801) doi:10.2196/28801

KEYWORDS

mHealth; health promotion; physical activity; personalization; adaptive goal setting; gamification; office workers

Introduction

Research Case

Nowadays, sedentary behavior is highly pervasive. Sedentary behavior, as distinct from physical activity, encompasses a broad range of behaviors that involve sitting or lying down and do not increase energy expenditure substantially during waking hours [1,2]. On average, adults in Western countries spend between 7 and 11 hours per day sitting [3-6]. Adults sitting >10 hours a day are expected to see their all-cause mortality rates increase [7]. Conversely, adults who participate in at least 150 minutes of moderate-intensity activity per week—an equivalent of 20 to 30 minutes per day—are expected to decrease their mortality rate significantly [8]. However, even when an adult meets these guidelines, sitting for prolonged periods can compromise health [9]. Hence, frequently interrupting periods of sitting with (short) bouts of physical activity is also essential to remain healthy [9].

Although the benefits of an active lifestyle for health are well established, it remains hard for people to engage more often in physical activity and reduce sedentary behaviors, with inactivity accounting for 9% of the premature mortality globally [10]. Mobile health (mHealth) interventions can be used to promote physical activity and reduce sedentary behavior, particularly if these tools use evidence-based behavior change strategies (eg, goal setting) [11].

Promising results have been obtained by using gamification techniques as behavior change strategies [11-13]. Gamification is a set of motivational techniques that use game mechanics outside game contexts to foster participation, engagement, and loyalty [14,15]. Gamification techniques are especially effective when they are tailored toward an individual's particular preferences and needs (ie, personalized) [12] because behavior change techniques that motivate one person may not appeal to someone else [16]. For example, it has been demonstrated that there are significant associations between specific personality traits and the types of motivational techniques that individuals prefer [17,18], as well as the type of motivational messages that they appreciate more [19]. Furthermore, it has been suggested that interventions that take into account users' individual capabilities when setting intervention goals are better at sustaining user engagement [20,21]. Similarly, a review of behavior change strategies to promote physical activity using mHealth interventions concluded that (adaptively) tailored goals seem to be more effective than static generic goals [22].

In this study, we aim to replicate these findings and focus on adaptively tailoring our gamified mHealth program to the capabilities of individual end users. On the basis of the findings by Sporrel et al [22], we hypothesized that an intervention that suggests health goals to its users based on the users' capabilities and preferences will be more engaging (ie, resulting in lower dropout rates as well as higher adoption rates of healthy routines) than an intervention that does not tailor its goals. Note that in mHealth tools, capabilities are always relative to other daily routines. Specifically, the researcher is typically not interested in the participant's actual peak capabilities for certain sports activities. Instead, a researcher typically considers the participant's capability to perform a healthy activity in accordance with the participant's professional and personal duties. When also considering that mHealth interventions aim to be scalable, researchers typically rely on participants' self-reported capabilities rather than inviting all participants for an endurance test.

We aim to extend existing literature with suggestions on how goals are most effectively tailored in digital health promotion settings. Although it has already been suggested that assigned—but personalized—goals may be more effective than having users set their goals themselves [22], it remains unclear what exact strategies are most effective in setting tailored goals in a digital health promotion setting. Of course, different strategies for tailoring goals in a digital health promotion setting exist. For example, promising results have been obtained by personalizing goals based on (1) task complexity (eg, by personalizing daily step goals [23]), (2) context (eg, by setting context-aware goals [24]), or (3) the user's autonomy to set goals (eg, by recommending goals individually instead of having users select goals from a predefined list [25]). However, the relationship between the goal target behavior (eg, to go for a walk or a run) and the impact of the goal on user engagement levels remains unclear. Hence, in this study we aim to investigate the relationship between the goal target behavior and the goal's impact on user engagement by setting personalized goals for different types of health-related activities (ie, walking, biking, and engaging in sports).

In the following sections, we first survey the literature to examine the relationship between an individual's capability and (suggested) goals as well as the impact of goals on behavior. Then, we detail our intervention, treatments, and study design. Subsequently, we present the results we obtained. Finally, we discuss the implications of our results and the weaknesses of this study as well as directions for future research.

Theoretical Background

Several *behavioral* theories (eg, the COM-B [Capability, Opportunity, and Motivation Model of Behavior] System [26] and the Fogg Behavior Model [27]) argue that, for a certain (target) behavior to occur, an individual must have the *capability* and *opportunity* to engage in the (target) behavior; in addition, the strength of *motivation* to engage in it must be greater than for any competing behaviors. The concept of capability entails a person's physical and psychological capacity to perform a target behavior [26]. Besides a person's actual capabilities,

motivation is key. Several *motivational* theories highlight that besides actual capabilities, the *perceived* ease or difficulty of performing a target behavior is an important motivating factor (ie, a concept that has been referred to as *self-efficacy* by Bandura [28] and was included as well in the Theory of Planned Behavior [29] and in Self-Determination Theory [30]).

Hence, a dilemma arises when assigning someone a behavior to perform. In particular, if the target behavior is too hard for an individual, they may feel anxious and may therefore not (continue to) engage in the behavior. In contrast, if the target behavior is too easy for them, they may feel bored and therefore may not (continue to) engage in the behavior either. Hence, an individual's level of skill and the level of complexity (ie, challenge) of a target behavior have to be in harmony. This trade-off is very well described by Flow Theory, which was formulated by Buchanan and Csikszentmihalyi [31]. This theory has inspired the design of several (gamified) mHealth tools such as Nike+, Zombies, Run!, Fitocracy, and Runkeeper [32], all of which aim at promoting physical activity through the "provision of optimally difficult challenges and feedback" [32]. The trade-off between a person's skill and the level of complexity of a suggested behavior is also described in Goal-Setting Theory, which proposes that task performance can be moderated by a number of factors, including task complexity and levels of self-efficacy [33]. Especially, from Goal-Setting Theory, it is known that task complexity should generally be at the verge of an individual's capabilities to foster engagement because difficult—but specific and still attainable—tasks generally result in better performance [33].

To summarize, although tasks that are too simple lead to dropout due to boredom and tasks that are too complex trigger dropout due to anxiety (or frustration), tasks that are difficult—but specific and still attainable—generally yield the highest levels of engagement. To adhere to this principle, we designed a procedure in this study that takes into account participants' self-reported capabilities and desired health goals in setting the tasks for them to perform.

Finally, Flow Theory points out that a person's (perception of their) capability changes over time because their skill increases whenever they complete more challenging tasks [31]. Hence, to engage individuals in a task over a longer period time, the tasks' complexity should be adaptively tailored in accordance with the skill they possess. For example, a recent review of behavior change strategies to promote physical activity using mHealth interventions concluded that increasing goal complexity by 20%-100% generally yields increased goal performance [22]. To adhere to this principle, we designed a procedure in this study that increased task complexity every 2 weeks to account for participants' increased skill levels and prevent dropout due to boredom.

Methods

Recruitment

Participants were recruited among staff members of 7 governmental organizations (ie, 6 municipalities and 1 provincial organization) in the region of Antwerp, Belgium, in October 2019. The study was introduced to these staff members as a health promotion campaign to promote physical activity and reduce sedentary behaviors. Participants were enrolled only after they gave their explicit consent, which was collected upon registration for the campaign.

Participants were recruited by representatives of the sports departments of the participating organizations. These representatives were organized in a regional committee, with the aim to promote employee health. This committee had also called for this scientific study to be conducted. Different methods for recruiting participants were used within different organizations (ie, the means of recruiting participants were not prescribed in a study protocol). Some organizations relied on word of mouth to promote the campaign, whereas others used email advertising or printed advertisement posters. Promotional wristbands had been made available for distribution by all committee members, but we did not supervise the distribution. This approach was adopted to respect organizational differences.

Ethical Approval

All operational procedures were approved by the ethical committee of Eindhoven University of Technology (experiment ID ERB2019IEIS5). The ethical review committee concluded that the potential benefits of this study outweighed its potential risks.

Intervention Context

To test our hypothesis, we used the mHealth tool GameBus. GameBus was especially designed for health promotion and provides a highly configurable gamification engine that is used for sustaining participant engagement. According to the classification of gamification elements by Hamari et al [13], GameBus implements the gamification mechanisms of challenges, points, goals, progress visualizations, leaderboards, and rewards. In addition, it allows configuring of these mechanisms for testing scientific hypotheses. The tool supports hosting multiple experimental designs on a single platform, ensuring that user experience remains similar across these different designs. Moreover, the platform enables researchers to gather rich data in a manner that is compliant with European (privacy) legislations.

Using GameBus, a health promotion campaign was especially designed to promote walks, bike rides, and sports sessions. The campaign had a duration of 8 weeks and was split into 2-week periods (so-called waves).

To foster awareness of the campaign and stimulate word of mouth, participants could track their performance on 2 social leaderboards: a leaderboard displaying the individual scores of participants within a certain organization and a leaderboard displaying the average scores of participants within a certain municipal department. At the commencement of each wave, both leaderboards were reset (ie, scores were set back to zero). The actual implementation of both leaderboards in our mHealth tool is presented in Figure 1.

To score points on these 2 leaderboards, a participant was given a set of tasks that, upon completion, were rewarded with points. At the commencement of each wave, a participant received a

set of 6 tasks (Figure 2). The first three tasks were the same across all waves: (1) go for a short walk of at least 250 m, (2) go for a short bike ride of at least 1 km, and (3) share your healthiest moment of the week. These tasks were included to provide participants with a sense of gratification relatively easily and make them feel that *all* their physical efforts were awarded.

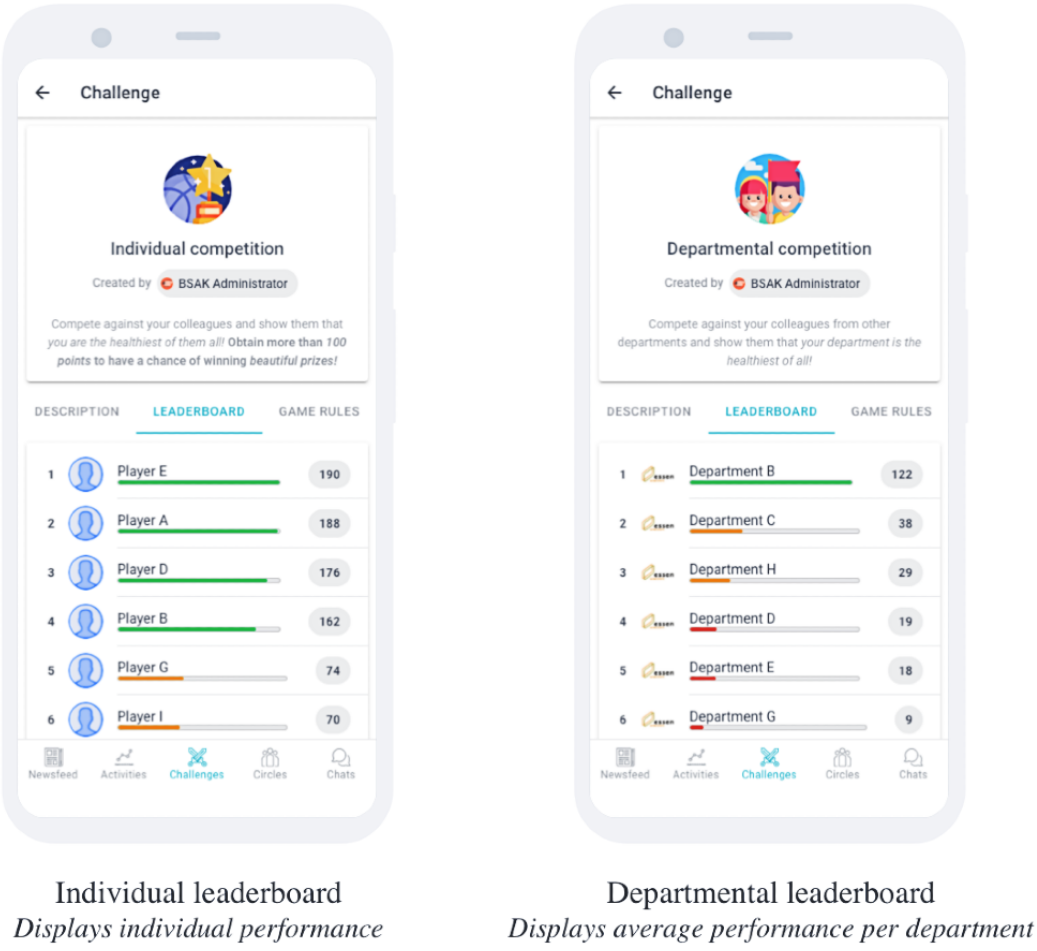
The other three tasks were dynamic (ie, updated at the commencement of each wave) and arguably more difficult to perform: (1) go for a longer walk of at least X km, (2) go for a longer bike ride of at least X km, and (3) go for a sports session lasting at least 30 minutes X times per week. In this study, these 3 dynamic tasks were either updated generically (ie, for the control group) or personalized based on the user's current self-reported capabilities and health goals (ie, for the treatment group). Specific details on how these tasks were set for the different treatment groups are presented in the *Study Design* section.

Users could either manually or automatically prove their engagement with a certain task. By means of the mobile app, users could manually register that they had performed a certain task. Alternatively, users could use an activity tracker to automatically track their efforts. The activity trackers that were supported included Google Fit, Strava, and a GPS-based activity tracker that was built into the native version of the GameBus app (available for both Android and iOS devices).

To prevent users from repeating a single task over and over, we set a maximum number of points that could be obtained per task per week, as well as a maximum number of times a task was rewarded per week with points (Table 1). Note that the sports session is rewarded X times per week, where X depends on the actual campaign wave. Note that therefore the number of points awarded per sports session needs to be calculated for a given wave by dividing 40 (the maximum number of points awarded per week) by X. Figure 2 displays the exemplar sets of tasks that users in the control or treatment groups could be assigned through GameBus.

Finally, GameBus provided a set of features for social support: a newsfeed showed when other participants had scored points, and participants could *like* and comment on each other's healthy achievements as well as chat with each other.

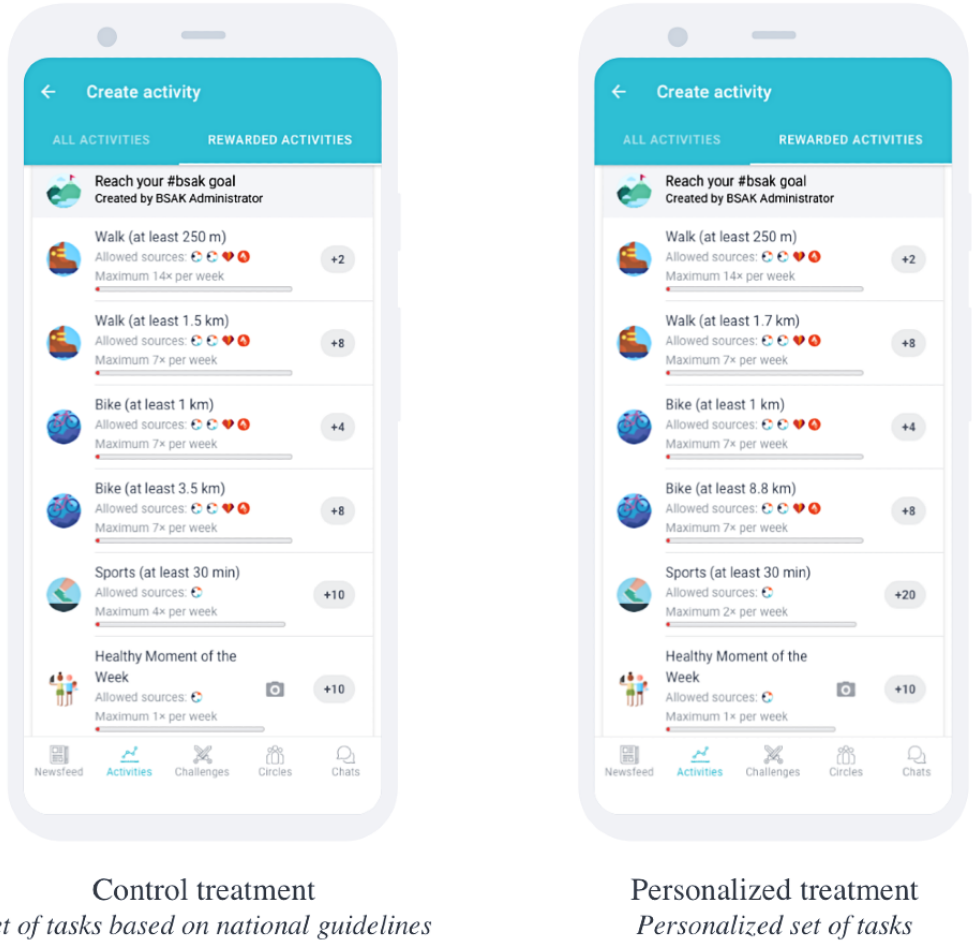
Figure 1. Display of social leaderboards.



Individual leaderboard
Displays individual performance

Departmental leaderboard
Displays average performance per department

Figure 2. Display of the different sets of tasks per treatment.



Control treatment
Set of tasks based on national guidelines

Personalized treatment
Personalized set of tasks

Table 1. Maximum number of points that could be obtained per suggested activity.

Task	Maximum number of points per week	Maximum number of times rewarded per week	Points per activity
Short walk	28	14	2
Longer walk	56	7	8
Short bike ride	28	7	4
Longer bike ride	56	7	8
Sports session	40	X	40/X
Healthy moment	10	1	10

Study Design

Overview of Study Arms

The study was designed as a 2-arm randomized intervention trial. The experimental setup was centered around setting the complexity parameters (ie, the X values) of the 3 dynamic tasks. In particular, the parameters to determine were as follows: (1) the minimum distance of a longer walk, (2) the minimum distance of a longer bike ride, and (3) the maximum number of rewarded sports sessions (and consequently the number of rewarded points per sports session). For the control group, these parameters were based on Belgian physical activity guidelines, whereas for the personalization group, these parameters were tailored to the users' self-reported capabilities and health goals.

Control Group: Tasks Based on Guidelines

For the control group, the parameter values of the dynamic tasks were based on national guidelines. The Belgian guidelines for physical activity are based on the Australian activity guidelines [34]. These guidelines recommend a minimum of 150 minutes (ie, in line with the study by Long et al [8]) of moderate-intensity activity per week, with each activity episode lasting at least 10 minutes. In addition, these guidelines suggest regularly interrupting periods of sitting with (short) bouts of physical activity (ie, in line with the study by Owen et al [9]).

On the basis of these guidelines, it was agreed with the organizing committee to suggest tasks with a duration of 10 to 30 minutes, giving participants ample opportunity to engage in at least 150 minutes of moderate-intensity activity per week. In addition, as described in Table 2, we increased the minimum

durations of tasks throughout the waves by 20%-42% because it was found that increasing goal complexity (by 20%-40%) generally yields increased goal performance [35].

Table 2. Estimated time needed to complete a dynamic task per activity type, as suggested to the control group.

Parameter	Wave 1	Wave 2	Wave 3	Wave 4
Minimum duration of the longer walk, minutes (distance; m)	17.5 (1500)	25 (2000)	27.5 (2250)	30 (2500)
Minimum duration of the longer bike ride, minutes (distance; m)	12.5 (3500)	15 (4000)	16 (4250)	17.5 (4500)
Minimum duration of the sports session, minutes (maximum times; n)	30 (4)	30 (4)	30 (5)	30 (5)

Treatment Group: Personalized Tasks

To set a value for the complexity parameters of the dynamic tasks for the treatment group, it was necessary to have some insight into the users' current capabilities and health goals. We obtained self-reports of the users' capabilities and goals by means of a short intake survey. Note that all participants (ie, even control participants) were asked to complete this intake survey to avoid introducing a bias because just the act of declaring one's goals may already foster motivation for the task at hand [33]. As an incentive to fill out this short survey, a donation of €0.25 (US $0.28) was made to charity for every completed survey.

In the intake survey, participants were asked to provide an estimation of (1) the number of steps they walked on a daily basis, (2) the number of kilometers they biked on a weekly basis, and (3) the number of sports sessions in which they participated on a weekly basis. Note that the participants' capabilities were explicitly evaluated in accordance with their existing professional and personal duties because we aimed to promote health-related activities that the participants could fit in their daily routines.

Furthermore, participants were asked whether they wanted to improve on any of these (estimated) numbers. If they wanted to improve their capabilities, they were asked to express (depending on the dimension they aimed to improve) the following: (1) the number of steps they wanted to walk on a daily basis, (2) the number of kilometers they wanted to bike on a weekly basis, and (3) the number of sports sessions they wanted to attend on a weekly basis.

Subsequently, the data on participants' capabilities and goals for walks and bike rides was transformed to fit the description templates of tasks (eg, a task has the form of *go for a longer walk of at least X kilometers*, not the form of *walk X steps per day*). The number of steps one could, and wanted to, walk per day was multiplied by 0.73 (ie, average stride length) and divided by 3 to obtain a minimum trip length (eg, to reach a goal of walking 7000 steps per day, we would suggest regularly going for a walk of at least 7000 × 0.73/3 = 1703 m). The number of kilometers one could, and wanted to, bike per week was divided by 5 (eg, to reach a goal of biking 10 km per week, we would suggest regularly going for a bike ride of at least 10/5 = 2 km).

Now we could calculate the difference between a user's current and preferred level of capability. We would update a user's task complexity at the commencement of each wave to linearly grow toward their goal. Hence, to personalize each parameter, we have used the formula that is displayed below, where i is a reference to the individual participant for whom the parameter value is calculated, t is the type of parameter (eg, walking distance, biking distance, or number of sports sessions), W is the total number of waves of the campaign (ie, 4), and w is the wave number of a given wave:

In addition, the value for *capability* was set by participants themselves (ie, by means of the intake survey). If a participant had not completed the intake survey, their *capability* was estimated to be their last performance for a particular activity type *t*. In case a participant had no recorded history on the activity type *t*, their *capability* was defined as the average performance of all other users on the activity type *t*. Note that in case there was no history of any participant on the activity type *t* yet, that *capability* was defined as a fixed value (eg, 1 km for *t* with regard to walking, 2 km for *t* with regard to biking, and 2 sessions for *t* with regard to engaging in sports).

Furthermore, a participant's *goal* was also defined by the participants themselves, again by means of the intake survey. However, if a participant had not completed the intake survey, their *goal* was derived by multiplying their *capability* with a fixed value of 1.1 for *t* equals walks and bike rides (ie, indicating a 10% improvement) or by increasing their *capability* with a fixed value of 1 for *t* equals sports sessions.

Finally, the different parameter values were capped by a predetermined minimum and maximum. The minimum and maximum for walking distance were 1 km and 10 km, respectively; the minimum and maximum for the distance of a bike ride were 2 km and 17.5 km, respectively; and the number of sports sessions that were rewarded per week was capped between 2 and 10. For instance, if the aforementioned formula would suggest to reward 0 sports sessions, this final check would override that value, and instead allow a participant to claim points for their sports sessions twice per week.

Treatment Allocation

Users were allowed to join (and drop out) at any moment throughout the campaign. Whenever a user joined the campaign, they would always be given a default set of tasks until the end of the then-active wave (ie, the default set of tasks was displayed as the control treatment; Figure 2). After the wave had ended (and at the start of a new wave), a user would be allocated to either the control group or the treatment group and receive a new set of tasks accordingly.

The control and treatment samples were stratified such that each sample included the same number of people who had set a goal to improve their current capabilities (eg, new participants were immediately requested to express their current capabilities and goals through the intake survey). Obviously, the intention to improve one's current capabilities is an important covariate because people who have a certain goal in mind are likely more motivated to engage with the campaign because this desire may influence their engagement and performance levels [33]. By stratifying our samples, the control and treatment groups were likely to be comparable.

Study Procedures

Throughout the campaign we sent some email notifications to participants. In particular, upon registration, participants received a welcome email with a request to complete the intake survey. In addition, a campaign email was sent at the start of each wave. These campaign emails included participation instructions as well as directions for obtaining (technical) support. Finally, at the end of the campaign, a closing email with a request to fill out the posttest survey was sent. As an incentive to fill out this posttest survey, a donation of €1 (US $1.13) was made to charity for every completed survey. After 4 days, we sent out a reminder to fill out the posttest survey.

Finally, some of the 7 organizations expressed some additional requests. In particular, 1 organization (ie, the municipality of Wuustwezel) expressed the need for some additional tasks (eg, ones that were more specific than the catch-all task *Share your healthiest moment of the week*). Furthermore, the municipality of Essen requested waves with a duration of 4 weeks each (instead of a duration of 2 weeks each). For them, the social leaderboards were reset every 4 weeks (ie, twice over the entire campaign). However, note that—and this applied to the participants from Essen too—the personal set of healthy tasks was still updated every 2 weeks.

Measurements

In mHealth, engagement is most commonly captured by means of measures of app use [36]. Using the GameBus platform, the engagement of participants was repeatedly measured as follows: (1) the number of days a participant visited the app (ie, the distinct days the participant opened the mobile app) and (2) the number of activities a participant registered. These variables complement each other because the former may be limited to passive engagement, whereas the latter requires active participation (ie, performing the suggested tasks).

Both measurements were recorded per participant per wave. In addition, for each record, the wave number relative to the participant's participation date was recorded. Hence, a record for a particular participant who joined the campaign only in the fourth wave would have a relative wave number of zero for that record. This relative wave number was used to model time in this study to ensure that time effects (eg, novelty effects) were equal among participants.

In addition, the type of goal that the participants set in the intake survey was recorded. A participant's goal was either *unknown* (ie, if they did not complete the intake survey), *maintain* (ie, if they did not want to improve their current capabilities on any dimension), or *improve* (ie, if they expressed an intention to improve their current capabilities on at least one dimension).

Finally, participants filled out a posttest survey (presented in Multimedia Appendix 1) in which we especially assessed the perceived impact of the campaign on their walking, biking, and sports performance, as well as their perception of their capability to perform the prescribed tasks (ie, self-efficacy).

Statistical Analysis

The first set of statistical analyses focused on the evaluation of dropouts. A participant was labeled as a (provisional) dropout if they had not visited the app during a specific wave and was therefore assumed to have lost interest (ie, dropped out) during the previous wave. Several multiple regression models were fit to determine whether the number of dropouts changed over time and were different per treatment. In particular, we tested for significant second-order interaction effects of time (ie, the wave number) and treatment.

The second set of analyses focused on the evaluation of engagement levels of the participants. To evaluate treatment differences, further analyses were performed on participants who actually had an opportunity to receive exposure to the treatment. Hence, from the entire data set, a subset was derived preserving the combination of a particular participant and wave only if they had ever checked the app during that wave and if they had participated for a duration of at least two waves because during the wave in which a participant signed up, they were not actually receiving a treatment yet. Subsequently, several hierarchical linear models were estimated for the 2 outcome variables (ie, the number of days a participant visited the app and the number of activities a participant had registered) using time (ie, the relative wave number), participant's goal, and treatment as predictors. We tested whether significant second-order interaction effects existed among these variables. In all models we allowed random intercepts for both individuals and the governmental organizations they were part of. The final model was selected based on the Akaike information criterion [37]. The Akaike information criterion estimates the relative quality of statistical models for a given set of data. The measure rewards goodness of fit and includes a penalty for increasing the number of predictors (ie, to prevent overfitting because increasing the number of predictors generally improves the goodness of the fit).

In addition, a third set of analyses zoomed in on the experimentally controlled tasks (ie, the longer walk, the longer bike ride, and the sports sessions) to evaluate treatment differences at the level of individual activity types. Specifically, for each activity type, a hierarchical linear model was built to predict the number of times a participant registered a task for that particular activity type. Again, these models included time (ie, the relative wave number), participant's goal, and treatment as predictors. In addition, we tested whether significant second-order interaction effects existed among these variables. In all models we allowed random intercepts for both individuals and the governmental organizations they were part of. The final model was again selected based on the Akaike information criterion [37].

Finally, the fourth set of analyses focused on the evaluation of subjective measures that were derived from a posttest survey. This final set of analyses was performed on a subset of the data set that only included participants who filled out the posttest survey and were using the mHealth app in more than one wave. A set of 3 separate linear models was used to estimate the perceived impact of the campaign on walking performance, biking performance, and sports performance. An additional linear model was used to estimate participants' perception of their capability to perform the tasks they were prescribed (ie, self-efficacy). Again, in all 4 models, time (ie, the total number of waves a participant had been visiting the app), participant's goal, and treatment were used as predictors, and we tested whether significant second-order interaction effects existed among these variables. To obtain the final models, a backward elimination selection procedure was used [38]. Backward elimination starts with all predictors included in the model, with variables subsequently being eliminated one at a time. At each step, the predictor with the highest $P>.05$ is deleted [38]. This method of deletion continues until all predictors are significant (ie, $P<.05$).

Ethics Approval

All operational procedures were approved by the ethical committee of Eindhoven University of Technology (experiment ID ERB2019IEIS5). The ethical review committee concluded that the potential benefits of this study outweighed its potential risks.

Results

User Statistics

In total, 176 unique participants joined the study, and they were randomly assigned to a treatment: 82 (46.6%) were assigned to the control treatment and 84 (47.7%) were assigned to the personalized treatment, whereas 10 (5.7%) were not assigned to a treatment at all because they only signed up during the last wave and therefore only experienced the default set of tasks. Of the 176 participants, 83 (47.2%) completed the intake survey (26/83, 31%, set themselves a maintenance goal and 57/83, 69%, set themselves an improvement goal), whereas 93 (52.8%) did not complete the intake survey and hence their goal was unknown. These data are summarized in Figure 3, which displays a cohort diagram that details the number of participants engaged in different study phases.

Figure 3. Cohort diagram that details the number of participants engaged in different study phases.

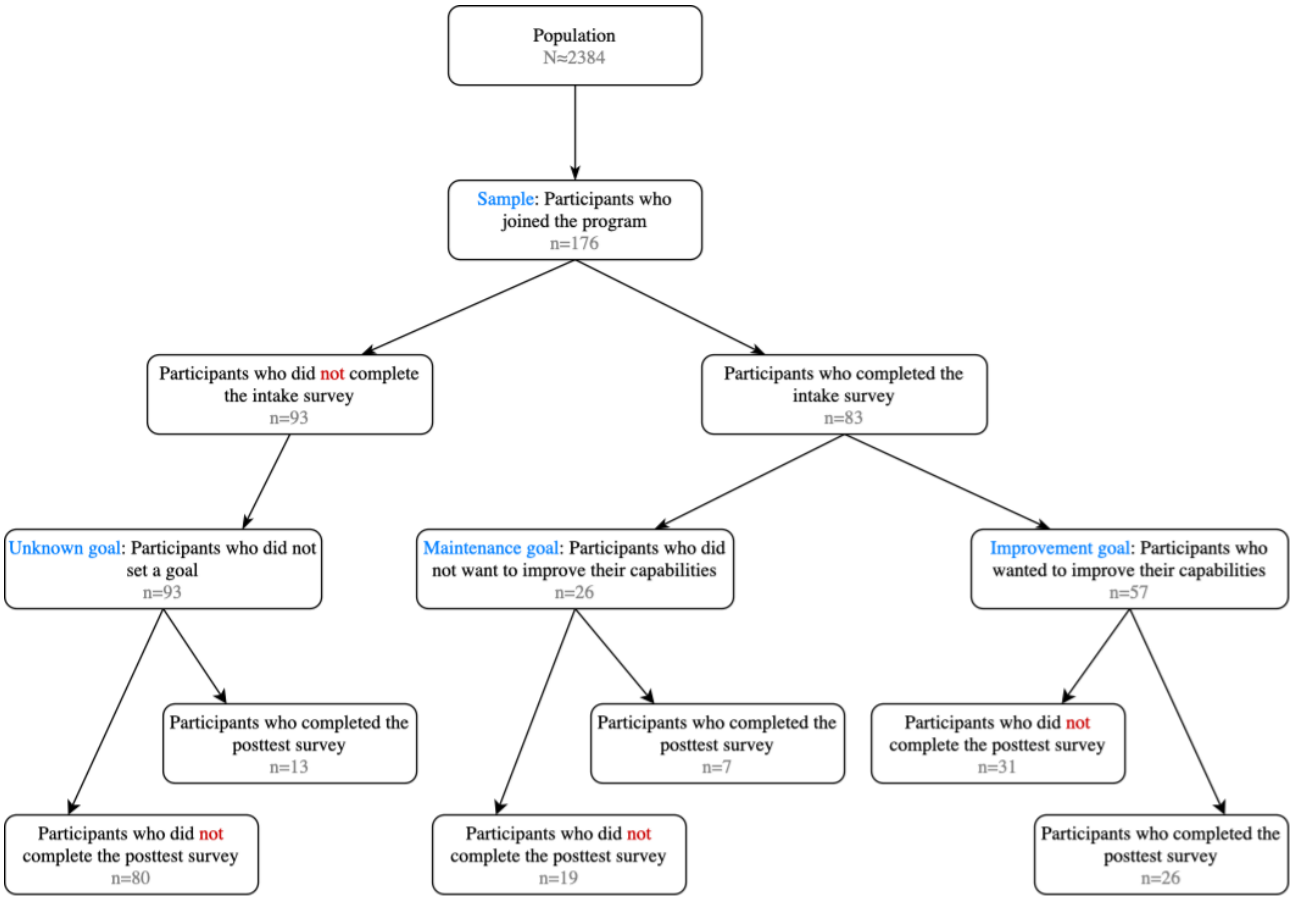


Table 3 displays sample demographics based on the results of the posttest survey, which was filled out by 26.1% (46/176) of the participants. Gender, age group, and personality scores are displayed for the entire sample as well as per treatment. The demographic characteristics in the control group and treatment group are distributed similarly. Hence, it is assumed that these groups were comparable at baseline.

Figure 4 displays the decrease in the number of participants who visited the mobile app during a given wave. The number of participants who joined the campaign for the first time during

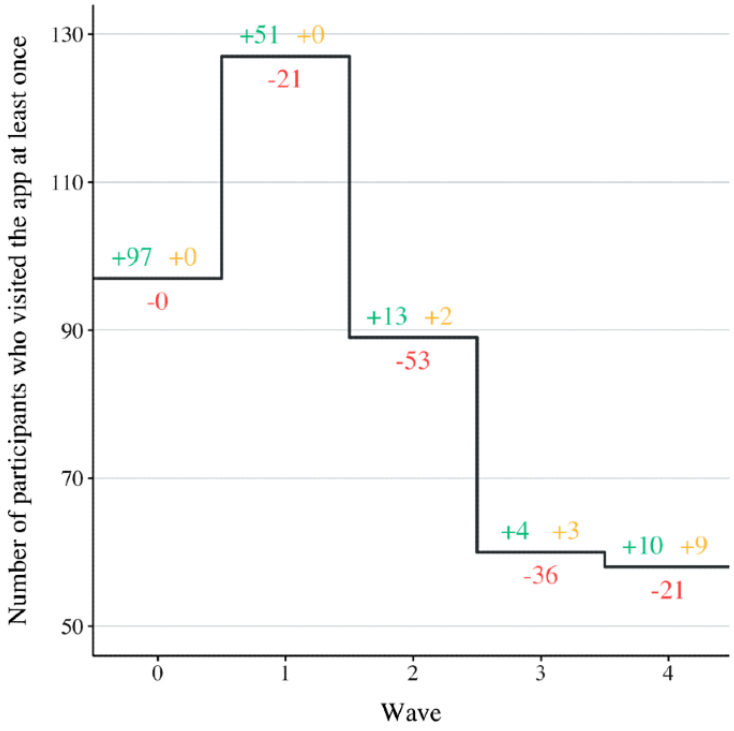
a given wave are displayed in green. The number of participants who dropped out during a specific wave are displayed in red. The number of participants who checked the mobile app during a specific wave, although they dropped out during an earlier wave (ie, reclaimed users) are displayed in yellow. Using multiple regression analysis, it was found that participants tended to drop out over time (ie, the wave number is a significant factor for predicting dropouts at P=.03; Table S1 in Multimedia Appendix 2). No significant differences in dropout rates between treatments could be detected. In addition, no significant interaction effect between time (ie, the wave number) and treatment was detected. Hence, it is assumed that dropouts were spread equally over treatments.

Table 3. Sample demographics (N=46).

Characteristic	Sample, n (%)	Control group, n (%)	Treatment group, n (%)	No treatment, n (%)
Gender (n=46)				
Male	13 (28)	7 (54)	5 (38)	1 (8)
Female	33 (72)	15 (45)	18 (55)	0 (0)
Age group (years; n=44)				
21-30	12 (27)	6 (50)	5 (42)	1 (8)
31-40	16 (36)	7 (44)	9 (56)	0 (0)
41-50	3 (7)	2 (67)	1 (33)	0 (0)
51-60	8 (18)	4 (50)	4 (50)	0 (0)
61-70	5 (11)	1 (20)	4 (80)	0 (0)
Personality scores (n=41), μ , σ				
Openness	2.573, 0.610	2.500, 0.473	2.643, 0.723	—[a]
Conscientiousness	2.329, 0.616	2.250, 0.579	2.405, 0.654	—
Extraversion	2.780, 0.645	2.763, 0.599	2.798, 0.701	—
Agreeableness	2.006, 0.476	1.975, 0.499	2.036, 0.463	—
Neuroticism	3.299, 0.710	3.362, 0.681	3.238, 0.748	—

[a]Posttest personality scores were not available for the participant who was not assigned a treatment.

Figure 4. Number of participants who visited the app at least once per wave.



Descriptive Statistics of Complexity Parameters

The complexity parameters of the dynamic tasks that the control participants were assigned are presented in Table 2. However, the complexity parameters for the treatment group were different for each individual in that group and were only determined at the start of a new wave. The mean (SD), minimum, and maximum values of the 3 complexity parameters are displayed per wave in Table 4.

Table 4. Mean (SD), minimum, and maximum values of the complexity parameters per dynamic task as presented to the treatment group.

Parameter	Wave 1	Wave 2	Wave 3	Wave 4
Minimum distance of the longer walk (m)				
Mean (SD)	1777 (600)	2025 (500)	2091 (488)	2054 (526)
Minimum	1000	1000	1000	1004
Maximum	4365	4471	4578	4684
Minimum distance of the longer bike ride (m)				
Mean (SD)	7822 (3704)	7439 (3194)	7718 (3386)	9127 (3728)
Minimum	2000	2000	2000	2000
Maximum	17,500	17,500	17,500	17,500
Suggested sports sessions				
Mean (SD)	2.41 (1.02)	2.97 (0.83)	3.67 (0.85)	3.09 (0.93)
Minimum	2	2	2	2
Maximum	7	7	7	8

Evaluation Outcomes

Evaluation of Engagement Levels

Description of the Data Set

Of the 176 participants, 10 (5.7%) only joined the study during the last wave; hence, they were not assigned a treatment and were therefore excluded from further statistical analysis, leaving 166 (94.3%) participants in the data set. In addition, of these 166 participants, 55 (33.1%) only visited the app at their registration (ie, during their first wave) and hence were also excluded from further statistical analysis, leaving a total of 111 (66.9%) participants in the data set for evaluation of engagement levels (ie, 51/111, 45.9%, assigned to the control treatment and 60/111, 54.1%, assigned to the personalized treatment).

Impact on Passive Engagement Levels

Figure 5 displays the number of days participants visited the app on average per wave per treatment. Figure 6 displays the number of days participants visited the app on average per type of goal they set.

From the second set of statistical analyses, it was found that the number of days participants visited the app dropped over time (ie, –1.174 days per relative wave; $P<.001$; Figure 5 and Table S2 in Multimedia Appendix 2). No significant difference between treatments was detected, although it did matter whether participants completed the intake survey. In particular, participants who completed the intake survey—and hence set themselves a goal to either maintain or improve their current capabilities—visited the app on more distinct days than those who did not set themselves a goal (ie, +2.176 days for participants with a maintenance goal; $P<.001$; and +1.625 days for participants with an improvement goal; $P=.005$; Figure 6). Finally, no significant interaction effects were detected; all treatments were affected equally by the impact of time.

Figure 5. Mean plot of the number of days participants visited the app per wave, per treatment.

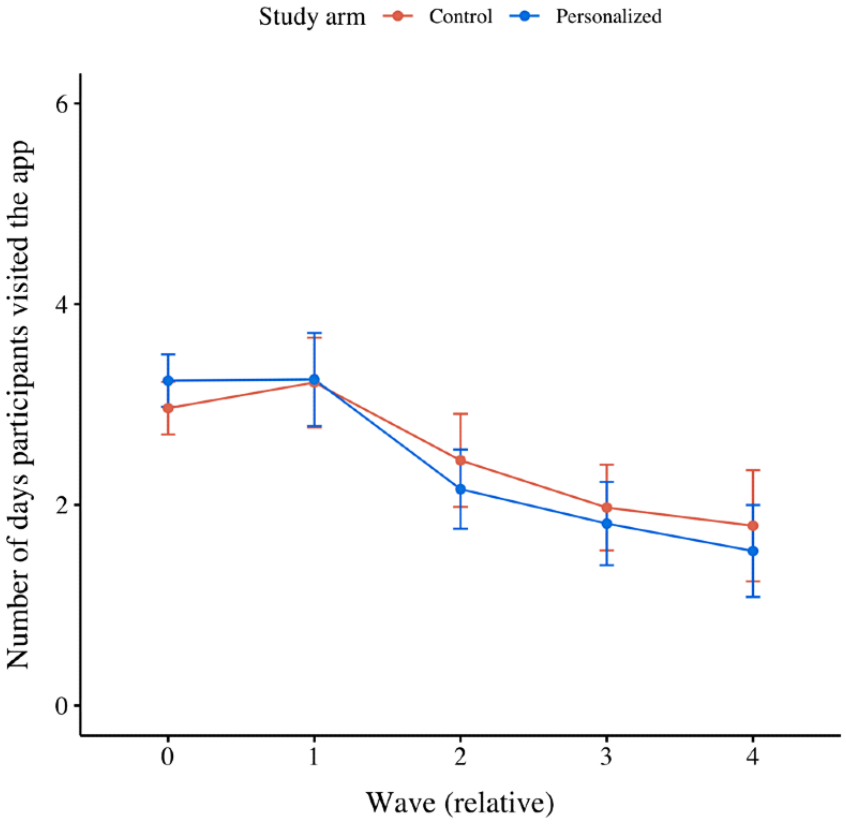
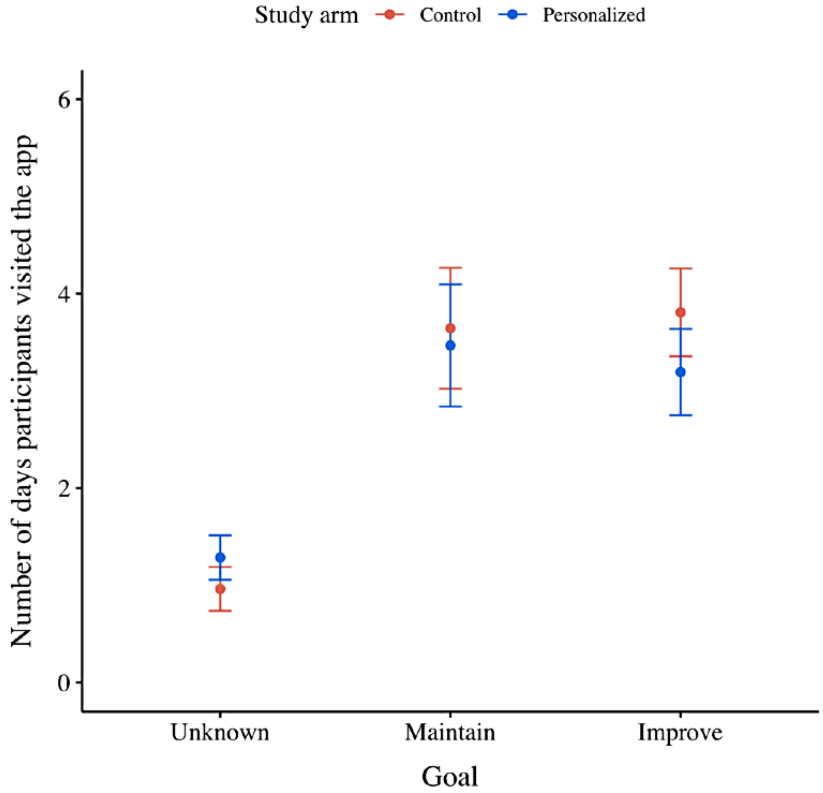


Figure 6. Mean plot of the number of days participants visited the app per their ambition to improve their current capabilities, per treatment.



Impact on Active Engagement Levels

Figure 7 displays the average number of activities participants registered per treatment. Figure 8 displays the average number of activities participants registered per type of goal they set. Multimedia Appendix 3 displays an overview of the number of times a particular suggested task was registered per organization.

Moreover, from the second set of statistical analyses, it was found that the number of activities participants registered

decreased over time (ie, –0.080 activities per wave; *P*<.001; Figure 7 and Table S3 in Multimedia Appendix 2). No significant difference between treatments was detected, although it did matter whether participants completed the intake survey. In particular, participants who set themselves a maintenance goal registered more activities than those who did not set themselves a goal (ie, +1.535 activities; *P*=.03; Figure 8). Moreover, participants who set themselves an improvement goal registered even more activities (ie, +3.258 activities; *P*<.001; Figure 8). Finally, no significant interaction effects were detected; again, all treatments were affected equally by the impact of time (ie, relative wave number).

Figure 7. Mean plot of the number of activities participants registered per wave, per treatment.

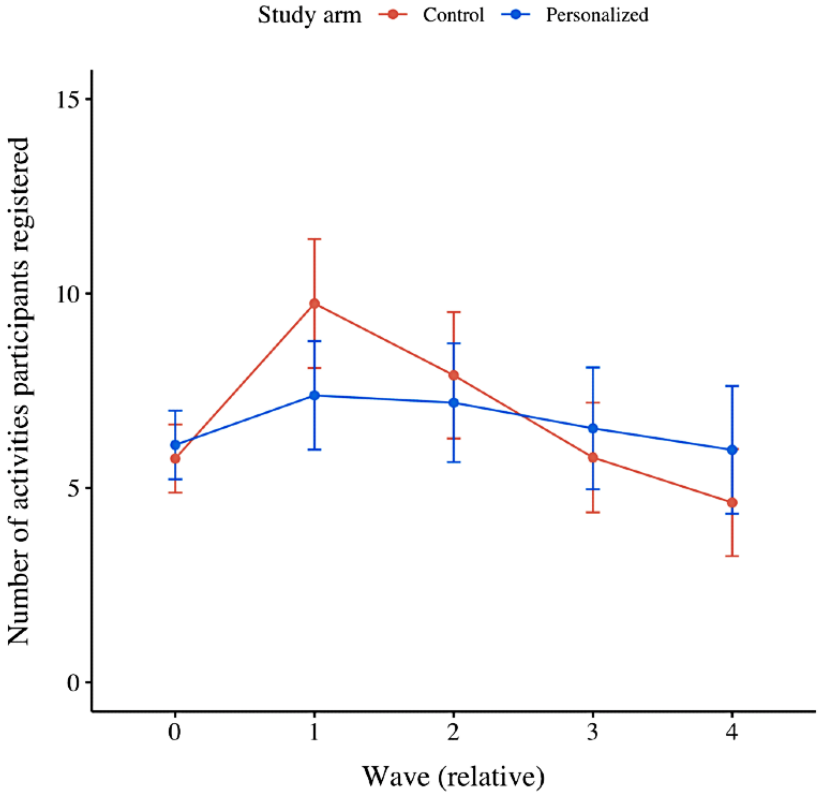
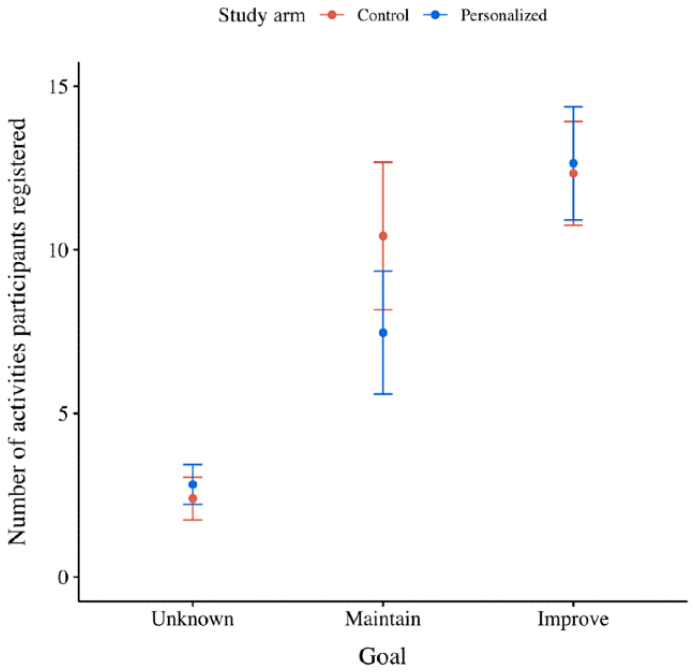


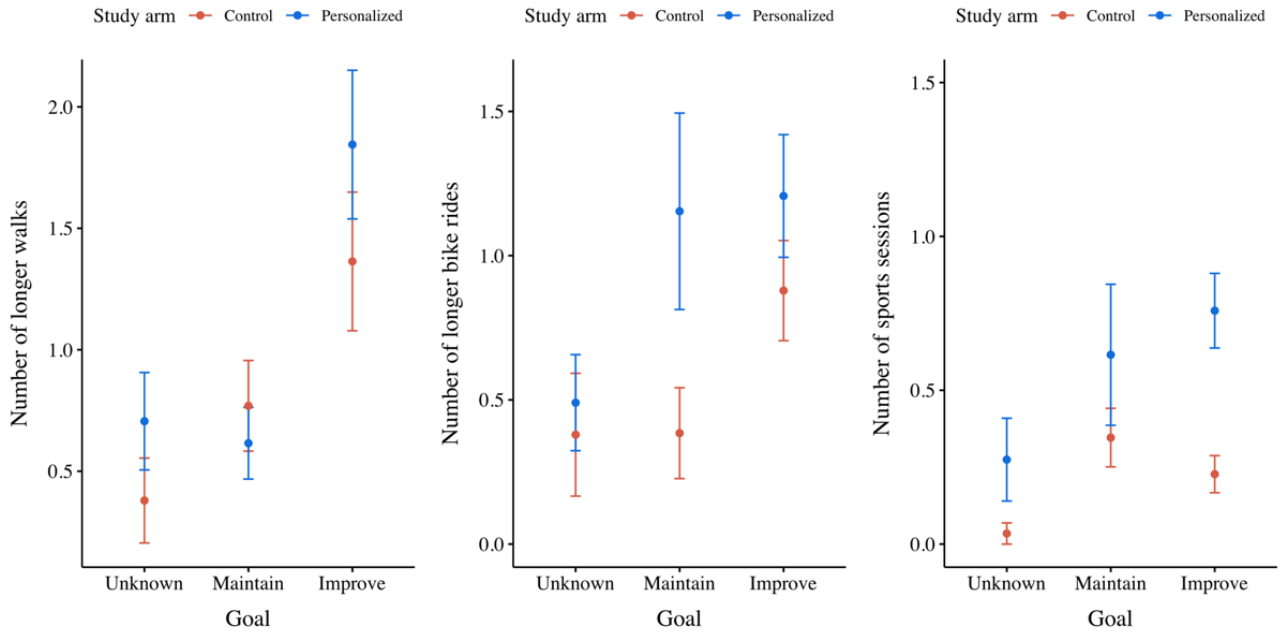
Figure 8. Mean plot of the number of activities participants (per treatment) registered after they were grouped based on their ambition to improve their current capabilities.



Impact on the Execution of Particular Activities

The third set of analyses zoomed in on the experimentally controlled tasks (ie, the longer walk, the longer bike ride, and the sport sessions) to evaluate treatment differences at the level of individual activity types (Figure 9). For each activity type, a hierarchical linear model was built to predict the number or times a participant registered a task for that particular activity type. No significant predictors were found for estimating the number of longer bike rides a participant registered. However, the number of longer walks a participant registered depended particularly on the goal they had set (ie, +0.261 walks for maintenance goals; *P*=.53; and +0.917 walks for improvement goals; *P*=.004; Table S4 in Multimedia Appendix 2). Moreover, the number of sports sessions a participant registered was dependent not only on the goal they had set (ie, +0.405 sports sessions for maintenance goals; *P*=.05; and +0.318 sports sessions for improvement goals; *P*=.04), but also on the treatment they had been assigned to (ie, +0.276 sports sessions if personalized; *P*=.05; Table S5 in Multimedia Appendix 2).

Figure 9. Mean plots of the number of longer walks, longer bike rides, and sports sessions participants registered per ambition to improve their current capabilities, per treatment.



Perception Analysis

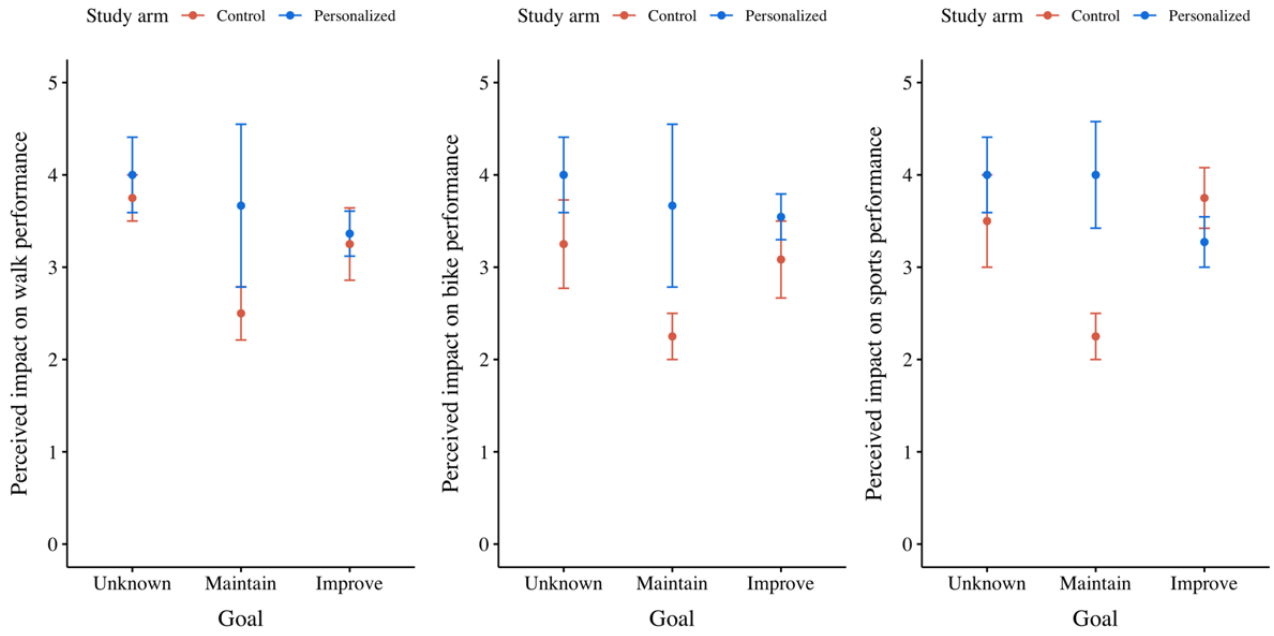
Description of the Data Set

Finally, we analyzed the participants' perception of their performance as well as capability to complete the program's suggested tasks (ie, self-efficacy). This fourth set of analyses was performed on a subset of the data set that only included participants who (1) filled out the posttest survey and (2) were using the mHealth app in >1 wave. This resulted in a data set of 38 participants (ie, 20, 53%, assigned to the control treatment and 18, 47%, assigned to the personalized treatment).

Perceived Impact on Performance

When zooming in on the perceived impact on performance of individual activity types (ie, walks, bike rides, and sports sessions), no significant predictors were found for estimating the perceived impact on walk performance (Figure 10). Nevertheless, the perceived impact on bike performance depended particularly on the treatment a participant received (ie, +0.047 if personalized; *P*=.04; Table S6 in Multimedia Appendix 2). In addition, the perceived impact on sports performance was dependent on a significant interaction effect between the treatment a participant received and the goal they had set (ie, +0.500 if personalized and a goal to maintain their current capabilities and –0.227 if personalized and a goal to improve their performance; *P*=.04; Table S7 in Multimedia Appendix 2).

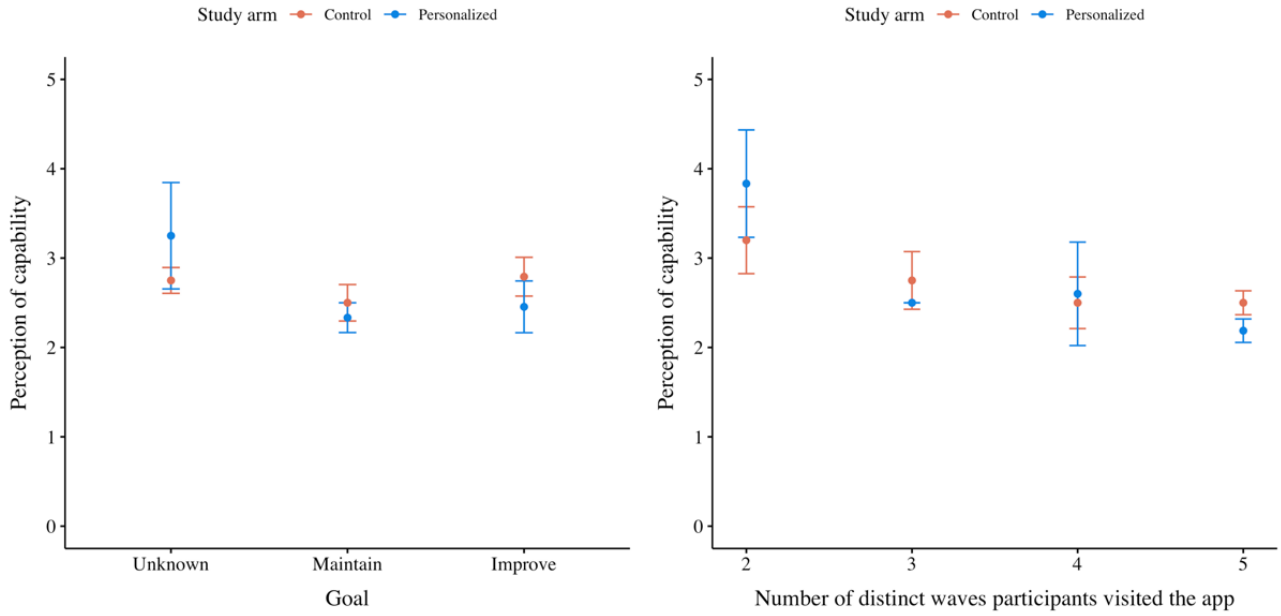
Figure 10. Mean plots of the perceived impact on walk, bike, and sports performance per participants' ambition to improve their current capabilities, per treatment.



Impact on Perception of Capability

Finally, the fourth set of analyses yielded a linear model to estimate the participants' perception of their capability to perform the prescribed activities (ie, self-efficacy; Figure 11). The treatment did not have a significant impact on the participants' perception of their capability. Nevertheless, for both the control and treatment groups, the perception of capability diminished over time (ie, –0.329; Table S8 in Multimedia Appendix 2) because the parameter measuring the total number of waves in which a participant had been visiting the app was reported significant at P=.001.

Figure 11. Mean plots of the perception of capability (ie, self-efficacy) per treatment. The chart on the left groups participants based on their ambitions to improve their current capabilities, and the chart on the right groups participants based on the number of waves during which they remained active.



Discussion

Principal Findings

The aim of this study is to evaluate the impact of personalized goal setting in a gamified health promotion program on participant engagement levels. Our results show that engagement with the program inevitably dropped over time, both in the personalized condition and in the control condition. Although this pattern is common in digital health promotion programs [39], several factors may be relevant for explaining this tendency in this particular context. First, it must be noted that only a limited number of participants had explicitly set a goal to maintain or improve their current capabilities (ie, 83/176, 47%).

According to the Transtheoretical Model of Behavior Change, there are 5 sequential *Stages of Change* that characterize one's readiness for change [40]. Hence, a great proportion of our sample seemed to be still in the precontemplation or contemplation phase, phases in which they were actually not (yet) planning for a more active lifestyle. Second, it must be noted that the participants' autonomy was limited during this program (eg, they were not rewarded for improving their dietary intake but instead only received suggestions for improving their levels of physical activity), which—according to Self-Determination Theory—may have harmed their intrinsic motivation levels [30].

Still, the participants who had set themselves a goal (ie, by completing the intake survey) were more engaged than those who had not. In particular, these participants visited the app more frequently and also registered more of the healthy tasks they were prescribed. Hence—as proposed by Goal-Setting Theory—setting a goal is in itself a motivating task [33]. Nevertheless, improvement goals—which are arguably more difficult to achieve than maintenance goals—did not seem to be significantly more motivating in general than maintenance goals. This finding seems to contradict both Flow Theory and Goal-Setting Theory, which propose that difficult—but still attainable—goals are more engaging than easier goals [31,33]. Then again, it should be noted that the descriptive means were mostly in the expected direction (ie, improvement goals were more engaging than maintenance goals) and the impact of improvement goals was actually significantly larger for promoting sports sessions: if a participant explicitly expressed a need to improve their current performance, they perceived their sports performance to be improved significantly.

Finally, the impact of the personalized treatment on engagement levels seemed to be generally limited. However, descriptive means were mostly in the expected direction (ie, personalized goals were more engaging than generically suggested goals). The seemingly limited impact of personalized goal setting may be explained by the actual strategy for personalizing the set of tasks. Moreover, we found that personalizing the suggested minimum number of sports sessions did stimulate participants to perform significantly more sports sessions, as well as significantly improved their perception of their sports performance. Upon close examination of this complexity parameter, we found that it can be characterized as a frequency parameter, whereas the parameters for personalizing walks and bike rides are typically characterized as intensity parameters. A frequency parameter defines *how many times* a particular activity should be performed in a given time frame, whereas an intensity parameter defines *how* a particular activity should be executed (eg, for how long and how far). We are unaware of context-specific factors that could have influenced this effect. However, we cannot claim generalizability yet either.

Finally, it must be noted that the treatment group participants did not feel more capable of completing the program's tasks than the participants in the control group. Although no significant differences between the treatment groups could be detected with respect to the participants' perception of capability to complete the program's tasks (ie, self-efficacy), the treatment group participants who set themselves a goal reported the lowest levels of self-efficacy on average among all participants. Hence, our personalization strategy may have suggested tasks that were perceived as too difficult or too easy by our target users, thereby potentially compromising self-efficacy and engagement with the program [30,31].

Limitations

The execution of this study was subject to several limitations. First, participants could take part without completing the intake survey. As a result, it was unknown in the case of some participants whether they explicitly choose not to set goals for themselves or whether they actually did aim to maintain or improve their current capability levels.

Second, participants may have felt that the number of points they were awarded for their activities, which affected their position on the social leaderboard, was unfair. By nature of the personalized treatment, each participant's intervention program was unique (ie, the intervention program was tailored to participants' individual capabilities and goals). Although, objectively speaking, this tailoring strategy makes the whole competition actually more fair, we received reports from several participants perceiving it as unfair that they had to (seemingly) expend more effort than their colleagues to be awarded the same number of points.

Third, an additional design choice that participants may have perceived as unfair was the decision to reward walks and bike rides on a per-trip basis, instead of, for example, on a daily aggregate basis. As a result, participants who went out for multiple shorter walks may not have been sufficiently rewarded for their effort. Then again, our aim was to promote activities with a minimum duration of 10 minutes, but perhaps it is worthwhile exploring this trade-off in more depth.

Fourth, the study outcomes were largely based on self-reported measures. Although participants could automatically (ie, objectively) prove their engagement with a certain task using Google Fit, Strava, or a built-in GPS-based activity tracker, they were also allowed to manually (ie, subjectively) claim that they had engaged in a certain task. This design choice could have introduced fraudulent activity registrations.

Fifth, the posttest survey suffered from low response rates (ie, 46/176, 26%). This low response rate on the posttest survey may have introduced a selection bias in the fourth set of analyses of subjective measures.

Finally, this study evaluated the impact of our intervention on a particular target group (ie, government staff) within a specific context (ie, the work environment). It is likely that the results will be generalizable to other audiences and contexts—because both Flow Theory [31] and Goal-Setting Theory are universal theories [33]—but it remains unclear what the intervention's exact impact on health behavior would be in different settings.

Future Work

A follow-up study should better control how participants set goals for themselves (ie, by means of the intake survey). For example, participants could be required to complete the intake survey before they are allowed to engage in the (gamified) program. Moreover, the intake survey could be extended to also

assess participants' Stage of Change according to the Transtheoretical Model of Behavior Change [40]. It seems natural to set different goals for participants who are in the precontemplation or contemplation phase (ie, the phase in which participants are not [yet] planning for a more active lifestyle) and for participants who are already actively improving their lifestyle (ie, participants in the action phase). Perhaps these 2 groups need to be assigned a different (gamified) program altogether.

In addition, future work should focus on evaluating different strategies for personalizing goal parameters. A particular opportunity is exploring in more detail the potential impact of personalizing the frequency parameters, rather than the intensity parameters. Focusing on promoting activity frequency particularly satisfies physical activity guidelines, which suggest that frequently interrupting periods of sitting with (short) bouts of physical activity is essential to remain healthy because sitting for prolonged periods can in itself compromise health [9]. Does personalization based on frequency parameters also have a larger impact on engagement levels in general? And if so, why? Finally, future work could explore the impact of allowing participants to add personalized goals for other types of activities too (eg, healthy dietary intake).

Recommendations

Although we have not yet been able to generalize our findings to support the claim that personalizing activity frequency fosters engagement levels better than personalizing activity intensity, we still suggest that practitioners focus on setting personalized goals based on activity frequency, in particular, because focusing on activity frequency implies performing physical activity more often (instead of for longer duration or performing more intense physical activity). This focus adheres especially well to physical activity guidelines, which suggest that frequently interrupting periods of sitting with (short) bouts of physical activity is essential to remaining healthy because sitting for prolonged periods can in itself compromise health [9]. Meanwhile, we encourage scholars to replicate our study setup to gain a deeper understanding of the potential impact of different strategies for tailoring health goals. To this end, we recommend that scholars (also) apply Goal-Setting Theory [33] and Flow Theory [31] when designing their studies. Similarly, we encourage scholars to evaluate the relationship between strategies of adaptive goal setting and contextual factors (eg, whether outcomes can be replicated with other target audiences).

Conclusions

In this study, we evaluated a gamified program that was designed to promote engagement in physical activity with sedentary government staff. Our aim is to investigate the impact of adaptive goal-setting strategies on end-user engagement levels with the program. In particular, through the program, study participants were stimulated to engage in a set of health-related activities (eg, to go for a walk, run, or sports session). Of these activities, we tailored the suggested intensity (ie, the minimum walking or biking distance) and frequency (ie, for sports sessions) based on the end users' self-reported current capability (eg, current walking capability) and desired capability (eg, desired walking capability). Our results indicated that end-user engagement with the program inevitably decreased over time. However, compared with a control group, it was found that tailoring the frequency of suggested activities (ie, as opposed to tailoring the intensity of activities) does promote engagement in that activity (ie, engaging in sports sessions). This effect was reported to be especially strong in participants who expressed an intention to improve their health-related capabilities at the beginning of the program. In fact, engagement was generally higher in participants who expressed an intention to improve their capabilities on at least one health dimension. Hence, when designing a gamified health promotion program, end-user engagement levels may be fostered by having end users explicitly state their current and desired capabilities and by setting health goals that tailor the suggested frequency of engaging in activities that constitute these goals.

Acknowledgments

This work is part of the research program Gamification for Overweight Prevention and Active Lifestyle (443001101), which is partly financed by the Netherlands Organization for Health Research and Development.

Conflicts of Interest

PVG, RN and AK were involved in the development of the GameBus mHealth platform.

Editorial notice: This randomized study was only retrospectively registered. The authors explained that this is due to them not being aware that it was classified as a randomized controlled trial. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Posttest survey.
[PDF File (Adobe PDF File), 89 KB - mhealth_v10i3e28801_app1.pdf]

Multimedia Appendix 2

Detailed output of statistical analysis.
[PDF File (Adobe PDF File), 131 KB - mhealth_v10i3e28801_app2.pdf]

Multimedia Appendix 3
Overview of the number of times a particular suggested task was registered per organization.
[PDF File (Adobe PDF File), 24 KB - mhealth_v10i3e28801_app3.pdf]

Multimedia Appendix 4
CONSORT eHealth Checklist (v1.6.1).
[PDF File (Adobe PDF File), 339 KB - mhealth_v10i3e28801_app4.pdf]

References

1. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [doi: 10.1097/00005768-200009001-00009] [Medline: 10993420]
2. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl Physiol Nutr Metab 2012 Jun;37(3):540-542 [FREE Full text] [doi: 10.1139/h2012-024] [Medline: 22540258]
3. Health Survey for England - 2008: Physical activity and fitness. UK: The NHS Information Centre for Health and Social Care; 2009.
4. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):7-14 [FREE Full text] [Medline: 21510585]
5. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 01;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
6. Van Dyck D, Cardon G, Deforche B, Owen N, Sallis JF, De Bourdeaudhuij I. Neighborhood walkability and sedentary time in Belgian adults. Am J Prev Med 2010 Jul;39(1):25-32. [doi: 10.1016/j.amepre.2010.03.004] [Medline: 20547277]
7. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. PLoS One 2013;8(11):e80000 [FREE Full text] [doi: 10.1371/journal.pone.0080000] [Medline: 24236168]
8. Long G, Watkinson C, Brage S, Morris J, Tuxworth B, Fentem P, et al. Mortality benefits of population-wide adherence to national physical activity guidelines: a prospective cohort study. Eur J Epidemiol 2015 Jan;30(1):71-79 [FREE Full text] [doi: 10.1007/s10654-014-9965-5] [Medline: 25377532]
9. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
10. 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]
11. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
12. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
13. Hamari J, Koivisto J, Sarsa H. Does gamification work? - a literature review of empirical studies on gamification. In: Proceedings of the 47th IEEE Hawaii international conference on system sciences. 2014 Presented at: 2014 47th Hawaii International Conference on System Sciences; Jan 6-9, 2014; Waikoloa, HI, USA. [doi: 10.1109/hicss.2014.377]
14. Deterding S, Dixon D, Khaled R. 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: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28 - 30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
15. Huotari K, Hamari J. Defining gamification: a service marketing perspective. In: Proceeding of the 16th International Academic MindTrek Conference. 2012 Presented at: MindTrek '12: Proceeding of the 16th International Academic MindTrek Conference; Oct 3 - 5, 2012; Tampere Finland. [doi: 10.1145/2393132.2393137]
16. Orji R, Mandryk RL, Vassileva J. Improving the efficacy of games for change using personalization models. ACM Trans Comput Hum Interact 2017 Oct 31;24(5):1-22. [doi: 10.1145/3119929]
17. Jia Y, Xu B, Karanam Y, Voida S. Personality-targeted gamification: a survey study on personality traits and motivational affordances. In: Proceedings of the 2016 CHI Conference 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. [doi: 10.1145/2858036.2858515]
18. Karanam Y, Filko L, Kaser L, Alotaibi H, Makhsoom E, Voida S. Motivational affordances and personality types in personal informatics. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. 2014 Presented at: UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication; Sep 13 - 17, 2014; Seattle, Washington, USA. [doi: 10.1145/2638728.2638800]
19. D'Hondt J, Nuijten RC, Van Gorp PM. Evaluation of computer-tailored motivational messaging in a health promotion context. In: Modeling and Using Context. Cham: Springer; 2019.
20. Khoshkangini R, Marconi A, Valetto G. Machine learning for personalized challenges in a gamified sustainable mobility scenario. In: Proceedings of the CHI PLAY '17 Extended Abstracts: Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY '17 Extended Abstracts: Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play; Oct 15 - 18, 2017; Amsterdam, The Netherlands. [doi: 10.1145/3130859.3131321]
21. Maggiorini D, Mannalà M, Ornaghi M, Ripamonti L. FUN PLEdGE 2.0: a FUNny Platformers LEvels GEnerator (Rhythm Based). In: Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter. 2015 Presented at: CHItaly 2015: 11th biannual Conference of the Italian SIGCHI Chapter; Sep 28 - 30, 2015; Rome Italy. [doi: 10.1145/2808435.2808451]
22. Sporrel K, Nibbeling N, Wang S, Ettema D, Simons M. Unraveling mobile health exercise interventions for adults: scoping review on the implementations and designs of persuasive strategies. JMIR Mhealth Uhealth 2021 Jan 18;9(1):e16282 [FREE Full text] [doi: 10.2196/16282] [Medline: 33459598]
23. Phatak S, Freigoun M, Martín CA, Rivera D, Korinek E, Adams M, et al. Modeling individual differences: a case study of the application of system identification for personalizing a physical activity intervention. J Biomed Inform 2018 Mar;79:82-97 [FREE Full text] [doi: 10.1016/j.jbi.2018.01.010] [Medline: 29409750]
24. Rabbi M, Aung M, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7 - 11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2805840]
25. Herrmanny K, Ziegler J, Dogangün A. Supporting users in setting effective goals in activity tracking. In: Persuasive Technology. Cham: Springer; 2016.
26. 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]
27. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology; Apr 26 - 29, 2009; Claremont, California, USA. [doi: 10.1145/1541948.1541999]
28. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
29. Ajzen I. From intentions to actions: a theory of planned behavior. In: Action Control. Berlin, Heidelberg: Springer; 1985.
30. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. Boston, MA: Springer; 1985.
31. Buchanan R, Csikszentmihalyi M. Flow: the psychology of optimal experience. Design Issues 1991;8(1):80. [doi: 10.2307/1511458]
32. Hamari J, Koivisto J. Measuring flow in gamification: dispositional flow scale-2. Comput Human Behav 2014 Nov;40:133-143. [doi: 10.1016/j.chb.2014.07.048]
33. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066X.57.9.705]
34. Brown WJ, Bauman AE, Bull F, Burton NW. Development of Evidence-based Physical Activity Recommendations for Adults (18-64 years). Canberra: Australian Government Department of Health; 2013.
35. Moon D, Yun J, McNamee J. The effects of goal variation on adult physical activity behaviour. J Sports Sci 2016 Oct;34(19):1816-1821. [doi: 10.1080/02640414.2016.1140218] [Medline: 26860430]
36. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]
37. Grove D, Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. J Royal Stat Soc 1988;37(4/5):477-478. [doi: 10.2307/2348776]
38. Derksen S, Keselman H. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. British J Mathematical Stat Psychol 1992:265-282. [doi: 10.1111/j.2044-8317.1992.tb00992.x]
39. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]

40. Prochaska J, DiClemente C. Toward a comprehensive model of change. In: Treating Addictive Behaviors. Boston, MA: Springer; 1986.

Abbreviations

COM-B: Capability, Opportunity, and Motivation Model of Behavior
mHealth: mobile health

Edited by L Buis; submitted 16.03.21; peer-reviewed by S Wang, L van Velsen, S Warsinsky; comments to author 27.05.21; revised version received 22.07.21; accepted 20.12.21; published 31.03.22.

Please cite as:
Nuijten R, Van Gorp P, Khanshan A, Le Blanc P, van den Berg P, Kemperman A, Simons M
Evaluating the Impact of Adaptive Personalized Goal Setting on Engagement Levels of Government Staff With a Gamified mHealth Tool: Results From a 2-Month Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(3):e28801
URL: https://mhealth.jmir.org/2022/3/e28801
doi:10.2196/28801
PMID:35357323

Original Paper

Data Collection Mechanisms in Health and Wellness Apps: Review and Analysis

Ben Joseph Philip[1], MIT; Mohamed Abdelrazek[2], PhD; Alessio Bonti[2], PhD; Scott Barnett[3], PhD; John Grundy[4], PhD

[1]Deakin University, Melbourne, Australia

[2]School of Information Technology, Faculty of Science, Engineering and Built Environment, Deakin University, Melbourne, Australia

[3]Applied Artificial Intelligence Institute, Deakin University, Melbourne, Australia

[4]Monash University, Melbourne, Australia

Corresponding Author:
Ben Joseph Philip, MIT
Deakin University
221 Burwood Highway
Melbourne, 3125
Australia
Phone: 61 426824528
Email: benjo@deakin.edu.au

Abstract

Background: There has been a steady rise in the availability of health wearables and built-in smartphone sensors that can be used to collect health data reliably and conveniently from end users. Given the feature overlaps and user tendency to use several apps, these are important factors impacting user experience. However, there is limited work on analyzing the data collection aspect of mobile health (mHealth) apps.

Objective: This study aims to analyze what data mHealth apps across different categories usually collect from end users and how these data are collected. This information is important to guide the development of a common data model from current widely adopted apps. This will also inform what built-in sensors and wearables, a comprehensive mHealth platform should support.

Methods: In our empirical investigation of mHealth apps, we identified app categories listed in a curated mHealth app library, which was then used to explore the Google Play Store for health and medical apps that were then filtered using our selection criteria. We downloaded these apps from a mirror site hosting Android apps and analyzed them using a script that we developed around the popular AndroGuard tool. We analyzed the use of Bluetooth peripherals and built-in sensors to understand how a given app collects health data.

Results: We retrieved 3251 apps meeting our criteria, and our analysis showed that 10.74% (349/3251) of these apps requested Bluetooth access. We found that 50.9% (259/509) of the Bluetooth service universally unique identifiers to be known in these apps, with the remainder being vendor specific. The most common health-related Bluetooth Low Energy services using known universally unique identifiers were Heart Rate, Glucose, and Body Composition. App permissions showed the most used device module or sensor to be the camera (669/3251, 20.57%), closely followed by location (598/3251, 18.39%), with the highest occurrence in the *staying healthy* app category.

Conclusions: We found that not many health apps used built-in sensors or peripherals for collecting health data. The small number of the apps using Bluetooth, with an even smaller number of apps using standard Bluetooth Low Energy services, indicates a wider use of proprietary algorithms and custom services, which restrict the device use. The use of standard profiles could open this ecosystem further and could provide end users more options for apps. The relatively small proportion of apps using built-in sensors along with a high reliance on manual data entry suggests the need for more research into using sensors for data collection in health and fitness apps, which may be more desirable and improve end user experience.

(JMIR Mhealth Uhealth 2022;10(3):e30468) doi:10.2196/30468

KEYWORDS

data collection; mHealth apps; app review; app analysis; mHealth; mobile apps; development; data sharing; user experience; usability; automation; data reliability; mobile phone

Introduction

Background

Mobile health (mHealth) apps support health delivery by the use of mobile devices such as mobile phones, wearables, and other wireless devices [1]. Several mHealth systems have been created for various apps, such as drug dosage reference [2,3], weight management [4], and monitoring cardiac health using wearable devices [5]. These mobile apps collect or generate health insights from three sources: external devices (Bluetooth or Wi-Fi-based sensors), built-in smartphone sensors, and manual data entry.

By 2017, it was estimated that more than 300,000 health apps were available in app stores, with a market growth of 25% each year [6,7]. The use of mobile apps along with wearables and external sensors has enabled self-monitoring of one's health. They unobtrusively collect physiological data to provide better health outcomes and can also play an important role for patients living in remote areas with limited access to health care [4]. mHealth apps have been classified either as active or passive [8]—the former generate or derive health data using sensors, whereas the latter rely on manual user input.

The mHealth domain has seen a steady rise in smart wearable and fixed devices [9] that can be used to gather more detailed and accurate insights into people's health [10]. According to Forbes, by 2022, their demand is expected to grow annually by approximately 20% [11]. This introduction of sensors has also opened up new avenues for health care where these devices can continuously monitor one's health without manual interference. This constant monitoring can also help detect anomalies that may not manifest during a visit to a health care professional and can permit caregivers to remotely monitor their patients [12-14]. Several wearables have been developed for specific support in the mHealth domain and are augmented by novel solutions, such as virtual reality implemented on mobile devices [15]. Built-in sensors such as inertial measurement units (IMUs), microphones, cameras, and GPS modules can also provide insights into one's health and have been previously used for managing conditions such as sleep apnea [16]. Bluetooth Low Energy (BLE) has been widely adopted for transferring data, and several apps have been developed that pair BLE devices with smartphones for fetching health insights. The popularity of BLE and the availability of low-cost BLE devices has opened up new avenues for continuous health monitoring in a more user-friendly manner [12]. Such sensors provide an effective platform for collecting real time metrics conveniently and less intrusively, which may be useful in medical research [17]. Recently, they have also been suggested for use in low-cost mHealth systems such as those for diagnosing pneumonia [18]. Similar suggestions have also been made for physiological measurements such as heart and respiration rates, blood oxygen saturation, and blood pressure for application in health interventions [19]. Recent studies in this area include the use of BLE devices for managing diseases ranging from asthma [20] to tissue pain and mobility issues [21]. Several studies have reviewed mHealth apps and explored them from various perspectives, such as their impact on health outcomes [22], usability [23], and even the use of integrated smartphone sensors for monitoring health conditions [16]. Despite limitations around the accuracy of the apps and peripheral such as measurement errors caused by darker skin tones and higher BMI [24] and poor energy expenditure estimations by apps [25], they remain mostly well received [26].

A study by Wisniewski et al [27] researched around the attributes of health apps where they selected 120 top-rated apps from Google and Apple app stores in different categories and evaluated them manually. Their study revealed that most apps fell under the category of *self-monitoring of health or diagnostic data by client* apps (World Health Organization classification 1.4.2) [1], indicating a higher interest in, and availability of, self-monitoring apps.

The Use of Built-in and External Sensors

Overview

Most smartphones host several built-in sensors such as IMUs and GPS modules and support different wireless communication technologies such as Bluetooth and Wi-Fi. Many mHealth apps provide features such as workout tracking, medication reminders, and general health monitoring using external or built-in smartphone sensors, whereas others offer other features that may require manual data entry and include apps such as meal trackers and weight loss coaches.

Built-in Sensors

A recent assessment of health apps from curated health app libraries indicated that cameras were the most frequently used sensors where they were used for assessing one's heart rate and even for automated skin cancer diagnosis [16]. Similarly, the use of microphones has been used in apps that provide respiratory therapy [28,29]. Algorithms have also been developed for processing IMU readings to monitor movement and activity levels in a noninvasive manner and are now widely used for applications in fall detection and gait analysis to track the progression of diseases such as Parkinson disease [30]. These algorithms and functions have been integrated with other data collection mechanisms described below to create complex and robust health apps, with a common example being popular fitness trackers that use external heart rate sensors along with the onboard IMUs and GPS modules.

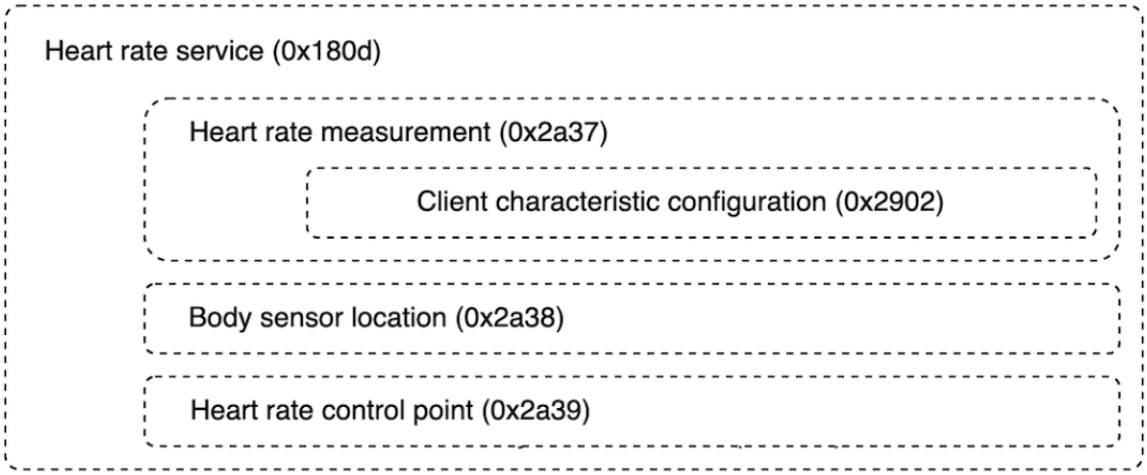
BLE Standard

BLE standard was originally designed with a focus on low cost, bandwidth, power consumption, and complexity and has allowed developers to design products that are more affordable than other wireless technologies such as Wi-Fi and Zigbee [31]. BLE uses profiles to define its functionality, which can cover operation procedures such as the Generic Attribute (GATT) profile, which describes procedures for exchanging data between devices and defines data models for the same. As several implementations can be made using GATT to exchange different types of data, the Bluetooth Special Interest Group (SIG) has defined a set of use cases and specific profiles that cover the required procedures and data structures. These have been defined using GATT services and characteristics and include profiles for securely transferring health-related metrics [32] such as

heart rate and blood pressure. Given that predefined profiles may not completely cover all apps, the Bluetooth SIG also permits device manufacturers to create their own vendor-specific profiles.

GATT provides a framework for data transfer and device operations, and apps based on BLE are required to comply with its specifications [31]. Data are exchanged between devices using the smallest addressable data units described by GATT—attributes. These are identified by 128-bit universally unique identifiers (UUIDs), which can also be represented using 16- (uuid16) or 32-bit (uuid32) shortened versions, with all currently SIG-assigned UUIDs being the uuid16 type [33]. The attributes are organized into nested blocks—services, which may contain 0 or more related characteristics, which, in turn, may also contain 0 or more descriptors [31]. As an example, Figure 1 describes the Bluetooth SIG-defined heart rate service specification [34].

Figure 1. Hierarchy of Bluetooth Low Energy heart rate service with uuid16 attribute representation (adapted from Bluetooth Specification—Heart Rate Service [34]).



As the GATT structure is strictly enforced for all BLE-compatible devices, any client app that intends to exchange data with them needs to either discover each exposed service or be aware of relevant services and characteristics. For specific use cases, apps would require UUID descriptions in their code for connecting with peripherals and identifying the services and subsequently reading exposed characteristics. Thus, an analysis of Android packages to extract these UUIDs would help us to identify not only those apps using peripherals to collect health metrics but also the use of standard and vendor-specific services. Apps have been previously analyzed to identify the use of BLE peripherals; however, the focus of existing works in the domain has been around security assessment and identification of vulnerabilities [35]. Tools such as BLEScope [35] and BLECryptracer [36] have been created for the same; however, they have not been used to identify the types of services supported by health apps.

Objectives

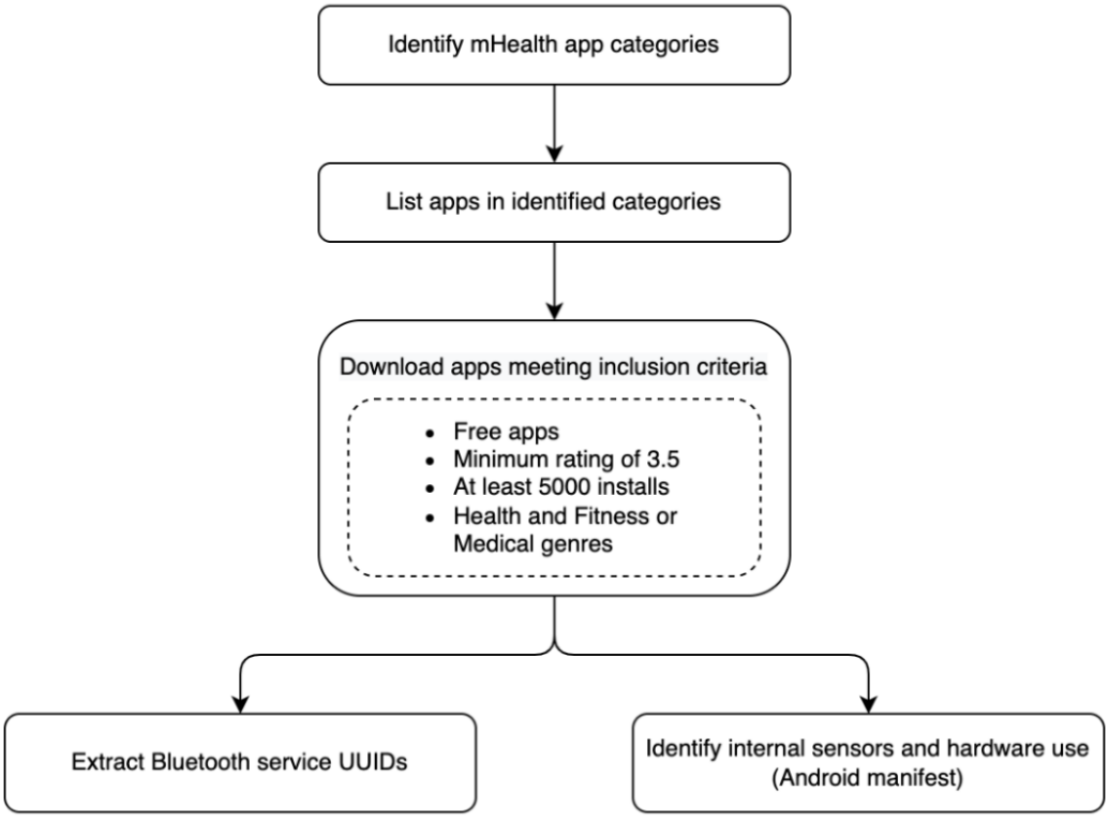
Recent exploration of the domain has also revealed interconnectivity and convenience as 2 factors impacting user experience [37]. This is even more important today, given the thousands of health apps with overlapping features and the user tendency to use more than one app [2]. Although they are mostly data driven, there is limited work around the analysis of existing mHealth apps to identify what data are collected and how, an understanding of which can help develop better health apps and eventually improve technology adoption. Therefore, our objective is to analyze a set of free mHealth apps to investigate the use of peripherals along with built-in sensors as an indicator of the collected data and provided features.

Methods

Overview

The Google Play Store is the official hub for downloading Android apps and offers over 100,000 mHealth apps [38]. Given the availability of several curated health app repositories, we explored the app categories described in one major curated app list—MyHealthApps [39]. Through a search on the Play Store using terms identified from this library, we identified apps that we then downloaded and analyzed. Figure 2 shows a high-level overview of the methodology, which is discussed in detail later.

Figure 2. Data collection methodology. mHealth: mobile health; UUID: universally unique identifier.



Identifying and Collecting mHealth Apps

We referred to the categories and subcategories of mHealth apps defined by the MyHealthApps library to guide our search, and we identified 15 categories with their respective subcategories listed in Table 1. We wrote a script around a Play Store scraper [40] that returned apps from the United States with English as the default language and gave us detailed information about the apps, and the results indicated that not all apps matching the search terms were related to health but included other genres also such as *News & Magazines* and *Tools*. Although several apps of interest fell under the *Health & Fitness* or *Medical* categories, many were classified under other groups and were excluded from our list. For example, the search term *Blood Pressure* returned a set of 250 apps with 113 health and fitness apps and 89 medical apps (as of November 2020). Our selection criteria for the apps had the following four key conditions to include a large set of more popular, accessible, and quality mHealth apps: free apps, rating >3.5, number of installs >5000, and health and fitness or medical apps

However, as we also wanted to include comparatively new apps along with well-established ones, we did not consider a minimum number of ratings.

After filtering the list and removing duplicates, the remaining apps were downloaded to our test machine from a mirror site [41], following which files except those with the .apk extension were discarded. We investigated the use of built-in sensors such as accelerometers, gyroscopes, GPS modules, and even the smartphone's camera modules in the identified apps. By analyzing app permissions, it is possible to infer to some extent what features these apps provide and how data are gathered. We were particularly interested in the use of GPS (coarse and fine locations), Bluetooth, Camera, Body Sensors, Microphones, and Activity Recognition permissions.

Table 1. App categories, the number of subcategories, and the total number of apps in each category (November 2020).

App category	Search terms (subcategories; n=157), n	Apps (N=38,780), n (%)
Bones and muscles	11	2745 (7.08)
Breathing and lungs	6	1493 (3.85)
Cancer	13	3237 (8.35)
Diabetes	5	1247 (3.21)
Endocrine	2	499 (1.29)
Heart, circulation, and blood	13	3238 (8.35)
HIV	1	249 (0.64)
Kidneys	1	250 (0.64)
Medication	3	749 (1.93)
Mental health	16	3991 (10.29)
Nervous system and brain	22	5469 (14.1)
Skin	3	748 (1.93)
Staying healthy	23	5744 (14.81)
Stomach, bowel, and continence	14	3493 (9.01)
Senses, mobility, and learning	24	5728 (14.77)

Data Extraction From mHealth Apps

App Data Set

Our query fetched a list of 38,130 apps (as of November 2020), which were then filtered to remove duplicates in each app set and those not meeting the inclusion criteria, giving us a much smaller list of apps for analysis [42] (N=3330). Of the 3330 apps, 12 (0.36%) apps were not found on the mirror site and 67 (2.01%) returned zip files that were discarded.

Extracting UUIDs From Packages

To analyze the downloaded apps, we used a popular static analysis tool—AndroGuard [43], which allowed us to decompile Android packages to extract relevant details. These apps need to be aware of the relevant services, characteristics, and descriptors to connect with peripherals. However, apart from statically defined UUID strings, apps can also construct them from a base ID and a shortened version at runtime. Although tracking these IDs may be necessary to identify all the possible uses of standard services, not all apps follow this approach. Analyzing the downloaded packages with AndroGuard helped us identify the following:

- The set of permissions and hardware features requested by the apps (to help understand how data are collected by the apps)
- Apps requesting Bluetooth permission (for identifying apps that may use external peripherals)
- Statically defined UUIDs for apps using Bluetooth (for understanding the use of predefined or vendor-specified profiles)

The Use of Internal Sensors

As access to device hardware and other features may have security implications, Android restricts access by mandating the use of permissions declared in the app's manifest file [44]. AndroGuard was used to identify built-in sensors accessed by mHealth apps through the declared permissions. Although the Android developer documentation recommends only using permissions necessary for the app to work as one of the best practices [45], some developers may request access to extra sensors and hardware without actually using them—a sign of a poorly developed app. However, such edge cases were not considered in this study.

iOS Apps

Although iOS apps also contribute to the mHealth app numbers, we limited our search to Android because of technical limitations around downloading these apps and the lack of open tools for decompiling and analyzing them. However, permission checks could be performed to indicate the types of hardware features used by these apps. We randomly selected 30 apps from the list of Android apps and searched for them on the iOS app store. Of these 30 apps, 25 (83%) were available on iOS, which were downloaded using Apple's Configurator tool and unpacked to identify the hardware features used in the apps based on the app permissions.

Overall, in each step of the exploratory analysis, custom tools were built and used to automate app downloads, static analysis, data manipulation, and management. Data were then manually checked to ensure accuracy.

Results

The Use of Internal Sensors

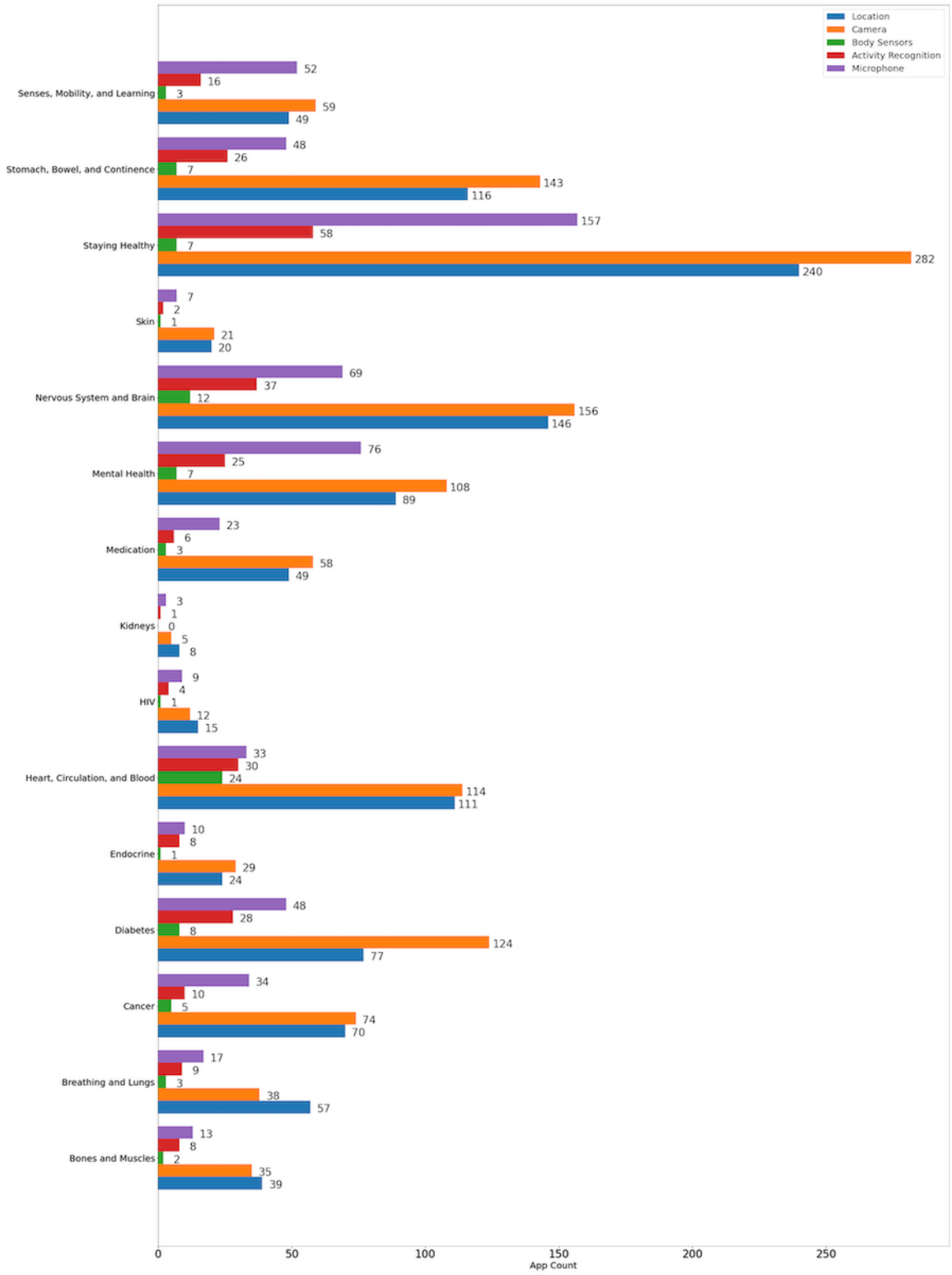
From the analyzed set of 3251 apps, we found several apps using the coarse (*ACCESS_COARSE_LOCATION*) and fine (*ACCESS_FINE_LOCATION*) locations, suggesting the use of distance tracking as a possible feature. Similarly, several instances of activity recognition for tracking step counts (*ACTIVITY_RECOGNITION*) and a few for body sensors

(*BODY_SENSORS*) were found. Smartphone cameras have also been widely used, as indicated by the presence of over 600 apps that requested the appropriate permission (*CAMERA*). Table 2 lists the number of apps using these permissions. Figure 3 shows the use of different sensors in each subset. We found that the camera being more popular across most search categories with GPS following closely, with the highest use seen in the *Staying Healthy* category. The high use of cameras is consistent with previous app reviews [16] and is discussed in the next section.

Table 2. Apps and requested permissions (N=3251).

Permissions (simplified)	Apps, n (%)
Coarse location	557 (17.13)
Fine location	598 (18.39)
Camera	669 (20.57)
Body sensors	36 (1.11)
Activity recognition	123 (3.78)
Audio recording	340 (10.45)

Figure 3. Built-in sensors used in apps across different categories.



The Use of Bluetooth Peripherals

Apps need to *know* the UUIDs of the services exposed by BLE peripherals to communicate with them and transfer data, and we found that 10.74% (349/3251) of the apps requested Bluetooth access. Table 3 lists the percentage of apps in each search category using Bluetooth. The Bluetooth SIG permits the use of vendor-specific UUIDs for different use cases, and 50.9% (259/509) of the discovered UUIDs were known and include service, characteristic, and descriptor identifiers. The unknown IDs include vendor-specific UUIDs along with those not related to Bluetooth operations; as these are not available publicly, further separation of this set was not possible.

We mapped the known UUIDs to the apps that used them, which allowed us to identify the most commonly used UUIDs and therefore, services. *Client Characteristic Configuration (00002902-0000-1000-8000-00805f9b34fb)* was found to be the most common, with 116 apps using the same UUID.

However, not many standard health-related services were identified, and we note that only the Heart Rate Measurement (41/3251, 1.26%) and the Heart Rate Service (40/3251, 1.23%) UUIDs were found in the top 10, with 1.2% (30/3251) of the analyzed apps using these services (Table 4).

We analyzed the permissions requested in the 25 iOS app versions of the selected Android apps. Given that the apps provided the same features, the permissions were not expected to differ and were mostly identical in the main categories of interest—Bluetooth, Camera, Microphone, Activity Recognition, and Location (Multimedia Appendix 1). A few minor deviations were observed on both platforms where some permissions did not match (eg, 2 Android variants requested near-field communication, which was not available on iOS). We could not analyze the BLE UUIDs in the apps because of the lack of open tools such as AndroGuard on iOS. However, given that both iOS and Android versions of an app are connected to similar hardware, the UUIDs are expected to be the same for the same group of apps.

Table 3. Percentage of apps in each category using Bluetooth.

App category	Percentage of apps using Bluetooth, n/N (%)
Bones and muscles	16/216 (7.4)
Breathing and lungs	29/188 (15.4)
Cancer	23/326 (7.1)
Diabetes	59/374 (15.8)
Endocrine	12/91 (13.2)
Heart, circulation, and blood	72/495 (14.5)
HIV	14/71 (19.7)
Kidneys	4/67 (6)
Medication	16/258 (6.2)
Mental health	53/505 (10.5)
Nervous system and brain	80/649 (12.3)
Skin	4/67 (6)
Staying healthy	145/1340 (10.8)
Stomach, bowel, and continence	58/454 (12.8)
Senses, mobility, and learning	49/391 (12.5)

Table 4. Services and the number of apps using them (N=3251).

Generic Attribute service	Apps, n (%)
Heart rate measurement	41 (1.26)
Glucose measurement	26 (0.79)
Running speed and cadence	14 (0.43)
Cycling speed and cadence	13 (0.39)
Blood-pressure measurement	13 (0.39)
Body composition measurement	12 (0.36)
Weight measurement	10 (0.3)

Discussion

Overview

End users tend to deal with multiple mHealth apps to manage their health and well-being, with even health care providers referring to more than one app as one may not provide all the details they need [2]. These apps do not share a consistent user interface, sensors, or a common mHealth data model, leading to poor overall user experience. To address this problem, there is a need for a comprehensive mHealth data collection model and catalog of sensors to develop robust app development guidelines and frameworks. This study represents the first step in this road map. We reviewed data collection in 3251 mHealth apps to understand what health data are collected and how apps collect them with a focus on built-in and external Bluetooth-based sensors.

Data Collection

Our findings indicate that although there is an increasing use of smart wearables and the increasing popularity of peripherals, not many apps use them to collect health data. Similarly, not many apps were found to use built-in smartphone sensors. Our results are consistent with a recent study by Wisniewski et al

[27], where most apps relied on manual entry with limited support for any wearables. Their reliance on manual reviews limited their study to 120 apps for mental health, which we were able to extend by automating the app review process. However, one drawback of our approach is that we cannot conclusively determine where the data were being used or what they were used for.

Our results show 20.57% (669/3251) and 18.39% (598/3251) of the apps used the camera and GPS modules, respectively, which are also the most used sensors across each app category, with the highest occurrence in the *Staying Healthy* set. This was expected as this category includes apps around diet and exercise where images and location data may be used for tracking meals and physical outdoor activities. We also expected a similar trend favoring cameras in the *Heart, Circulation and Blood* category. We were also surprised by the relatively high use of location and images in *Nervous System and Brain* apps, which may indicate the increasing acceptance of these data types in different apps. This can be an indication of useful features, such as scanning an item (eg, medication) or tracking movement. However, it may also indicate poor app design where access to sensors is requested without actually using them. Unsurprisingly, the lowest occurrence of these sensors was found in more medical apps as opposed to health and fitness apps, where categories such as HIV and Kidneys may not have any use of currently available built-in sensors at all. Apps were also found to use the microphone with the highest occurrence in the *Staying Healthy* category where its use can range from call features to speech analysis to tracking one's sleep.

Given the popularity of health wearables and peripherals, we expected to find a significant number of apps supporting them for passive data collection. However, our results indicated the opposite. We found 10.74% (349/3251) of the apps requesting Bluetooth access, of which only about half of the discovered UUIDs were found to be the standard Bluetooth services with the remaining unknown. Apart from unrelated IDs, this also indicates that most devices and apps used proprietary algorithms, limiting their compatibility and use [46]. However, of those that were known, very few were related to health, with the highest occurrence being the Heart Rate Measurement service in 40 apps (1.2%). Vendor-specific IDs (almost 50% of the reported apps using Bluetooth, n=250, 49.1%) may be used for any purpose, as defined by device manufacturers, making it difficult to identify the data transferred through those services. Besides the possibility of the UUIDs not being detected, this suggests that despite the growing popularity of wearables, they are restricted to a few manufacturers with limited apps using proprietary services and formats.

As we rejected apps with low ratings and downloads, we may also have skipped several bespoke apps used for specific cases or by small groups. These can include apps developed for research studies and specialized devices that may not be widely available. Similarly, as Google restricts search results to 250 items per search term, we were also limited in our app search. The analyzed data also indicated the presence of other known health-related services in a smaller number and showed the use of Heart Rate, Glucose, and Body Composition as the more common services provided by peripheral devices.

Data Sharing

mHealth devices and apps have been found to be useful for collecting clinical insights [47], which shows their potential not only in personal use but also in clinical apps where integration with electronic health records can help improve health outcomes. Newer apps integrate with frameworks such as Apple Health or Google Fit that allow data aggregation and sharing; however, they also require installation of more than one app—a challenge that deters end users. Here, a platform integrating a diverse set of apps, health records, and sensors could improve this aspect of mHealth apps with functionality and usability blending in seamlessly, potentially improving health outcomes.

Tools and Data Set

In addition to app analysis, our contribution also includes the raw data set, including collected app details along with the extracted data comprising app permissions and identified UUIDs. Our tool for downloading and analyzing apps is also included in our repository, which is available on GitHub [48], and would be beneficial for future studies related to mHealth app analysis.

Overall, our results suggest a more common use of manual entry (where automated data collection is possible), which, apart from being less reliable, also degrades user experience, leading to more users abandoning health apps [49]. Although usability is subjective, limited support for passive data collection with internal and external sensors can have a negative impact on app experience, which can lead to reduced adoption by end users—sidestepping any benefits the apps could offer. Therefore, it is critical to understand the importance of peripherals and built-in sensors in modern health solutions and integrating them in a clinically acceptable manner with health apps.

However, the main limitation of our work arises from automated data extraction, where we could not capture more nuanced details such as where the data from these sensors are being used and requires further investigation. Many valid health apps such as reference apps, management apps (weight and diet), and calculators (body composition, drug dosage, etc) may also be classified under other categories such as *Education* and *Books & Reference* and were rejected. Similarly, because of the difference in app numbers in each category, a comparison between them may be biased.

Given the presence of over 300,000 apps in app stores, analyzing all of them was not feasible and using a curated app list was a better approach for identifying different apps as they would be closer to the domain than manually searching through thousands of apps. As more health apps are widely available today for managing one's health, we believe that our results are relevant where accuracy of health data and wider integration would be important for better health care delivery. Although app analysis can be performed manually, we chose to automate the process, which ignored possible data sources such as developer descriptions and user reviews. Similarly, our results are based on limited app categories where the use of sensors and wearables may not be feasible (eg, *Medication* or *Mental Health*), and we acknowledge that these results may not be generalizable to the

entire domain. Therefore, there is also a need to explore new apps of current sensors in these areas to improve data collection.

Given the potential of mHealth apps to improve an end user's health, adherence to regular use is essential, which can only be ensured if such apps are intuitive and convenient to use. In the larger context of a connected, Internet of thing–enabled ecosystem, apps would play an important role as an interface. This indicates a need to integrate more peripherals with health apps to collect user data, which, along with built-in sensors, could ultimately help improve health outcomes. To that end, we envision a connected ecosystem of mini health apps, sensors, and health records as a key mHealth technology of the future. We plan to use these results to develop a single mHealth platform for aggregating several wearables and health apps as mini health services, which we believe would provide a much better experience to end users. We have built a prototype of such a platform with health micro-mHealth apps [50], an introduction to which is planned in our upcoming work followed by a study to understand its impact on user experience and technology adoption.

Conclusions

Given that user studies on app experience have highlighted convenience and data interconnectivity and aggregation as important factors, automating data collection can improve user experience, especially in apps requiring access to health metrics. However, a limited number of apps in our search were found to do so, indicating the need for more focus on integrating more peripherals and built-in sensors for health apps.

Our analysis of 3251 apps indicates that <10.74% (n=349) of the apps use smart devices and wearables to gather health metrics from users. In this set, extracted UUIDs show that very few apps used standard health-related Bluetooth services, with the most popular service being Heart Rate Measurement. Several apps have been found to use custom services that affect the interoperability of devices with different apps. Here, using standard profiles may be beneficial, as more apps would be able to interact with these devices, giving end users more options. Similarly, several apps were found to request access to device hardware features, such as GPS and camera, indicating the increasing acceptance of these devices. However, their numbers remain small, indicating the need for more research into using them in health apps.

Although manual entry may be inevitable for some apps, a significant number of apps requiring manual data entry were found in our set, highlighting the need to focus more on developing mHealth apps that automate health data collection. As several apps for research and health studies have been published, a better approach for developing and consuming mHealth apps is required. Overall, our findings can guide the design of future mHealth apps and has a positive impact on improving mHealth data collection in these apps.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Android and iOS permission comparison.
[XLSX File (Microsoft Excel File), 10 KB - mhealth_v10i3e30468_app1.xlsx]

References

1. World Health Organization. MHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
2. Anthony Berauk VL, Murugiah MK, Soh YC, Chuan Sheng Y, Wong TW, Ming LC. Mobile health applications for caring of older people: review and comparison. Ther Innov Regul Sci 2018 May;52(3):374-382. [doi: 10.1177/2168479017725556] [Medline: 29714532]
3. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
4. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
5. Yusuf A, Zulkifli FY, Mustika IW. Development of monitoring and health service information system to support smart health on android platform. In: Proceedings of the 2018 4th International Conference on Nano Electronics Research and Education (ICNERE). 2018 Presented at: 2018 4th International Conference on Nano Electronics Research and Education (ICNERE); Nov 27-29, 2018; Hamamatsu, Japan. [doi: 10.1109/icnere.2018.8642592]
6. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
7. 325,000 mobile health apps available in 2017 – Android now the leading mHealth platform. Research 2 Guidance. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2020-11-21]
8. Herron J. Bad apps: mHealth apps doubling as medical devices. J Electronic Res Med Libraries 2016 Dec 14;13(4):177-181. [doi: 10.1080/15424065.2016.1256800]
9. Previsión de los envíos de dispositivos wearables a nivel mundial de 2018 a 2022, por categoría de producto. Statista. URL: https://es.statista.com/estadisticas/637982/ventas-de-dispositivos-wearable-en-el-mundo-por-categoria/ [accessed 2020-11-21]

10. Garge GK, Balakrishna C, Datta SK. Consumer health care: current trends in consumer health monitoring. IEEE Consumer Electron Mag 2018 Jan;7(1):38-46. [doi: 10.1109/mce.2017.2743238]
11. Smartwatch popularity booms with fitness trackers on the slide. Forbes. URL: https://www.forbes.com/sites/paullamkin/2018/02/22/smartwatch-popularity-booms-with-fitness-trackers-on-the-slide/%5C#379d69447d96 [accessed 2020-11-21]
12. Zhang Y, Liu H, Su X, Jiang P, Wei D. Remote mobile health monitoring system based on smart phone and browser/server structure. J Healthc Eng 2015;6(4):717-737 [FREE Full text] [doi: 10.1260/2040-2295.6.4.717] [Medline: 27010652]
13. Banos O, Garcia R, Holgado-Terriza JA, Damas M, Pomares H, Rojas I, et al. mHealthDroid: a novel framework for agile development of mobile health applications. In: Ambient Assisted Living and Daily Activities. Cham: Springer; 2014.
14. Vandenberk T, Storms V, Lanssens D, De Cannière H, Smeets CJ, Thijs IM, et al. A vendor-independent mobile health monitoring platform for digital health studies: development and usability study. JMIR Mhealth Uhealth 2019 Oct 29;7(10):e12586 [FREE Full text] [doi: 10.2196/12586] [Medline: 31663862]
15. Grundy M, Abdelrazek M, Curumsing MK. Vision: improved development of mobile ehealth applications. In: Proceedings of the 5th International Conference on Mobile Software Engineering and Systems. 2018 Presented at: MOBILESoft '18: Proceedings of the 5th International Conference on Mobile Software Engineering and Systems; May 27 - 28, 2018; Gothenburg Sweden. [doi: 10.1145/3197231.3197263]
16. Baxter C, Carroll J, Keogh B, Vandelanotte C. Assessment of mobile health apps using built-in smartphone sensors for diagnosis and treatment: systematic survey of apps listed in international curated health app libraries. JMIR Mhealth Uhealth 2020 Feb 03;8(2):e16741 [FREE Full text] [doi: 10.2196/16741] [Medline: 32012102]
17. Munos B, Baker PC, Bot BM, Crouthamel M, de Vries G, Ferguson I, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. Ann N Y Acad Sci 2016 Jul;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]
18. Salti TE, Sykes ER, Zajac W, Abdullah S, Khoja S. NewPneu: a novel cost effective mHealth system for diagnosing childhood pneumonia in low-resource settings. In: Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2019 Presented at: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); Oct 17-19, 2019; Vancouver, BC, Canada. [doi: 10.1109/iemcon.2019.8936160]
19. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
20. 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]
21. Tilles A, Shaporev A, Reukov V. mHealth musculoskeletal applications: bluetooth low energy devices and android. In: Proceedings of the 2016 32nd Southern Biomedical Engineering Conference (SBEC). 2016 Presented at: 2016 32nd Southern Biomedical Engineering Conference (SBEC); Mar 11-13, 2016; Shreveport, LA, USA. [doi: 10.1109/sbec.2016.61]
22. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
23. Brown J, Nam Kim H. Usability evaluation of Alzheimer's mHealth applications for caregivers. Proc Human Factors Ergon Soc Annual Meeting 2018 Sep 25;62(1):503-507. [doi: 10.1177/1541931218621115]
24. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
25. Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, et al. Accuracy of 12 wearable devices for estimating physical activity energy expenditure using a metabolic chamber and the doubly labeled water method: validation study. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e13938 [FREE Full text] [doi: 10.2196/13938] [Medline: 31376273]
26. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
27. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
28. Clarence B, Julie-Anne C, Brendan K, Corneel V. Prescribable mHealth apps using built-in smartphone sensors for diagnosis and therapy. In: Proceedings of the IHBI Inspires 2019 Conference. 2019 Presented at: IHBI Inspires 2019 Conference; Aug 12-13, 2019; Brisbane, Australia.
29. Baxter C, Carroll JA, Keogh B, Vandelanotte C. Seeking inspiration: examining the validity and reliability of a new smartphone respiratory therapy exergame app. QUT ePrints. 2020. URL: https://eprints.qut.edu.au/204730/ [accessed 2021-02-28]
30. del Rosario M, Redmond S, Lovell N. Tracking the evolution of smartphone sensing for monitoring human movement. Sensors (Basel) 2015 Jul 31;15(8):18901-18933 [FREE Full text] [doi: 10.3390/s150818901] [Medline: 26263998]

31. Townsend K. Getting Started with Bluetooth Low Energy Tools and Techniques for Low-Power Networking. Sebastopol, California, United States: O'Reilly Media; 2014.
32. Specifications and test documents list. Bluetooth. URL: https://www.bluetooth.com/specifications/specs/ [accessed 2020-08-18]
33. Assigned numbers. Bluetooth. URL: https://www.bluetooth.com/specifications/assigned-numbers/service-discovery/ [accessed 2020-08-18]
34. Heart rate service. Bluetooth. URL: https://www.bluetooth.com/specifications/specs/heart-rate-service-1-0 [accessed 2020-08-18]
35. Zuo C, Wen H, Lin Z, Zhang Y. Automatic fingerprinting of vulnerable BLE IoT devices with static UUIDs from mobile apps. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019 Presented at: CCS '19: 2019 ACM SIGSAC Conference on Computer and Communications Security; Nov 11 - 15, 2019; London, United Kingdom. [doi: 10.1145/3319535.3354240]
36. Sivakumaran P, Blasco J. A study of the feasibility of co-located app attacks against BLE and a large-scale analysis of the current application-layer security landscape. 28th Unisex Security Symposium. URL: https://www.usenix.org/conference/usenixsecurity19/presentation/sivakumaran [accessed 2020-08-19]
37. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
38. Store stats 2022. 42 Matters. URL: https://42matters.com/google-play-statistics-and-trends [accessed 2021-01-06]
39. Welcome to myhealthapps.net. My Health Apps. URL: http://myhealthapps.net/about [accessed 2020-11-02]
40. Google Play store scraper. GitHub. URL: https://github.com/facundoolano/google-play-scraper [accessed 2020-11-02]
41. APK Combo homepage. APK Combo. URL: https://apkcombo.com/en-au/ [accessed 2020-11-02]
42. mHealth data collection. GitHub. URL: https://github.com/benphilip1991/mHealth-data-collection/blob/main/apps_list.txt [accessed 2021-01-08]
43. Welcome to Androguard's documentation. Androguard. URL: https://androguard.readthedocs.io/en/latest/index.html [accessed 2020-09-12]
44. Declare app permissions. Developers. URL: https://developer.android.com/training/permissions/declaring [accessed 2021-08-20]
45. Android permissions overview. Developers. URL: https://developer.android.com/guide/topics/permissions/overview [accessed 2021-08-20]
46. Godfrey A. Wearables for independent living in older adults: gait and falls. Maturitas 2017 Jun;100:16-26. [doi: 10.1016/j.maturitas.2017.03.317] [Medline: 28539173]
47. Shaw RJ, Yang Q, Barnes A, Hatch D, Crowley MJ, Vorderstrasse A, et al. Self-monitoring diabetes with multiple mobile health devices. J Am Med Inform Assoc 2020 May 01;27(5):667-676 [FREE Full text] [doi: 10.1093/jamia/ocaa007] [Medline: 32134447]
48. mHealth data collection. GitHub. URL: https://github.com/benphilip1991/mHealth-data-collection/tree/main/Results [accessed 2021-03-04]
49. 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]
50. Digital enhanced Living container app. GitHub. URL: https://github.com/benphilip1991/del-container [accessed 2021-01-08]

Abbreviations

BLE: Bluetooth low energy
GATT: Generic Attribute
IMU: inertial measurement unit
mHealth: mobile health
SIG: Special Interest Group
UUID: universally unique identifier

Edited by L Buis; submitted 16.05.21; peer-reviewed by L Emmerton, S Jalil; comments to author 21.07.21; revised version received 10.09.21; accepted 10.12.21; published 09.03.22.

Please cite as:
Philip BJ, Abdelrazek M, Bonti A, Barnett S, Grundy J
Data Collection Mechanisms in Health and Wellness Apps: Review and Analysis
JMIR Mhealth Uhealth 2022;10(3):e30468
URL: https://mhealth.jmir.org/2022/3/e30468
doi:10.2196/30468
PMID:35262499

Original Paper

Nutrition-Related Mobile Apps in the French App Stores: Assessment of Functionality and Quality

Prescilla Martinon[1*], MSc; Ina Saliasi[1*], PhD; Denis Bourgeois[1], Prof Dr, PhD; Colette Smentek[1], MPh-PhD; Claude Dussart[1], Prof Dr, PhD; Laurie Fraticelli[1*], MPh-PhD; Florence Carrouel[1*], MPh-PhD

Health, Systemic, Process UR 4129 Research Unit, University Claude Bernard, University of Lyon, Lyon, France

[*]these authors contributed equally

Corresponding Author:
Florence Carrouel, MPh-PhD
Health, Systemic, Process UR 4129 Research Unit
University Claude Bernard
University of Lyon
11 Rue Guillaume Paradin
Lyon, 69008
France
Phone: 33 478785745
Email: florence.carrouel@univ-lyon1.fr

Abstract

Background: The global burden of disease attributes 20% of deaths to poor nutrition. Although hundreds of nutrition-related mobile apps have been created, and these have been downloaded by millions of users, the effectiveness of these technologies on the adoption of healthy eating has had mixed

Objective: The aim of this study was to review which nutrition-related mobile apps are currently available on the French market and assess their quality.

Methods: We screened apps on the Google Play Store and the French Apple App Store, from March 10 to 17, 2021, to identify those related to nutritional health. A shortlist of 15 apps was identified, and each was assessed using the French version of the Mobile App Rating Scale: 8 dietitians and nutritionists assessed 7 apps, and the remaining apps were randomly allocated to ensure 4 assessments per app. Intraclass correlation was used to evaluate interrater agreement. Means and standard deviations of scores for each section and each item were calculated.

Results: The top scores for overall quality were obtained by Yazio - *Régime et Calories* (mean 3.84, SD 0.32), FeelEat (mean 3.71, SD 0.47), and *Bonne App* (mean 3.65, SD 0.09). Engagement scores ranged from a mean of 1.95 (SD 0.5) for iEatBetter: *Journal alimentaire* to a mean of 3.85 (SD 0.44) for FeelEat. Functionality scores ranged from a mean of 2.25 (SD 0.54) for Naor to a mean of 4.25 (SD 0.46) for Yazio. Aesthetics scores ranged from a mean of 2.17 (SD 0.34) for Naor to a mean of 3.88 (SD 0.47) for Yazio. Information scores ranged from a mean of 2.38 (SD 0.60) for iEatBetter to a mean of 3.73 (SD 0.29) for Yazio. Subjective quality scores ranged from a mean of 1.13 (SD 0.25) for iEatBetter to a mean of 2.28 (SD 0.88) for *Compteur de calories* FatSecret. Specificity scores ranged from a mean of 1.38 (SD 0.64) for iEatBetter to a mean of 3.50 (SD 0.91) for FeelEat. The app-specific score was always lower than the subjective quality score, which was always lower than the quality score, which was lower than the rating from the iOS or Android app stores.

Conclusions: Although prevention and information messages in apps regarding nutritional habits are not scientifically verified before marketing, we found that app quality was good. Subjective quality and specificity were associated with lower ratings. Further investigations are needed to assess whether information from these apps is consistent with recommendations and to determine the long-term impacts of these apps on users.

(JMIR Mhealth Uhealth 2022;10(3):e35879) doi:10.2196/35879

KEYWORDS

mobile apps; behavior change; diet; healthy food; nutrition; prevention; mHealth; mobile health; lifestyle; French

Introduction

Worldwide, the burden of noncommunicable diseases continues to rise [1]. The Global Burden of Disease study [2] found that 1 in 5 deaths was due to poor diet; thus, dietary factors were responsible for 11 million deaths per year, which was more than those from any other risk factor included in the study. Several forms of malnutrition, including obesity and undernutrition, can coexist in the same population and have a significant impact on health systems. Primary health care services and lifestyle behavior improvement based on education and behavior change have great potential to decrease the global burden of noncommunicable diseases, improve health throughout the life course, and enhance well-being [3]. Thus, counseling on healthy diets and proper nutrition are among the most important nutritional interventions for promotion, prevention, treatment, and rehabilitation [4].

Mobile health is defined by the World Health Organization's Global Observatory for eHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [5]." In recent years, the number of web-based mobile health apps has increased exponentially. Currently, there are more than 325,000 mobile health apps available on major app stores. These apps are in addition to web-based health apps available on other platforms such as websites, PC software, and game consoles [6].

Furthermore, the number of apps for improving nutrition and fitness continues to grow [7]. Hundreds of nutrition-related mobile apps have been created and downloaded by millions of users over the past few years [8]. The fact that some of these apps have been downloaded numerous times indicates that people want to monitor and control their diet [9]. Access to these mobile health apps is primarily via smartphones [10]. However, it has been shown that web and mobile technologies related to nutrition have a greater impact if combined with personalized advice from a dietitian [11]. Although other prevention approaches are required, the development of effective and equitable nutrition programs is a prerequisite [12]. Since the number of apps is growing exponentially every year, it is essential to update them regularly [13]. The main industry-wide challenge is to provide credible evidence for these apps [14]. To date, little usability testing of these apps has been conducted [6]. Only a small number of English-language digital health apps have reported their usability evaluation results [6]. Although the usefulness of technologies has been demonstrated, results on the effectiveness of technology integration on the adoption of healthy eating habits are conflicting [15]. In 2018, French was spoken in 29 countries on all continents, by approximately 300 million people; 235 million people use it daily, and 90 million people are native speakers [16]; however, no overall evaluation of French-language nutrition apps has been identified in the literature.

The aim of this study was to review which nutrition-related mobile apps were available on French App stores and to evaluate their quality.

Methods

Selection of the French Mobile Health Apps

Two academic researchers searched for nutritional health–related apps from March 10-17, 2021 on the French Apple App Store (for iOS) and the French Google Play Store (for Android) using the following search terms: "*nutrition*" (nutrition), "*diététique*" (dietetics), "*alimentation*" (food intake), "*régime alimentaire*" (diet), and "*manger sain*" (healthy eating). Because the use of truncation and logic operators (such as AND, OR, and NOT) were not possible in the App Store and Google Play Store, each search term was provided separately.

The 2 researchers individually eliminated duplicate apps by cross-checking name of the app and the developer before comparing their respective lists. The download pages of the remaining apps were screened, and then, apps were downloaded for in-depth screening using the inclusion criteria: (1) French language, (2) targeting adult users, (3) nutrition, diet or eating habits as subject matter, (4) self-personalized programs, and (5) free (or free for at least 14 days). Mobile health apps focusing on the following topics were excluded: sports, shopping, water alert notifications only, special diet (diabetic, baby, pregnancy, vegan, religious, abstain from eating, weight gain), recipes, apps created specifically for nutritionists' patients, meal delivery, pollution trackers, allergy and intolerance trackers, and barcode scanners.

Selection of a Standardized Rating Scale for Mobile Apps

We used the French version of the Mobile App Rating Scale (MARS-F). The MARS-F includes 19 objective items rated with a 5-point Likert scale that are divided into 4 sections [17-21]: the *engagement* section (5 items) evaluates if the app is fun, interesting, customizable, and interactive (eg, sends alerts, messages, reminders, feedback, or allows sharing); the *functionality* section (4 items) focuses on app operation, easy to learn, navigation, flow logic, and gestural design of the app; the *aesthetic*s section (3 items) evaluates the graphic design, the overall visual appeal, the color scheme, and the stylistic consistency; and the *information quality* section (7 items) determines if the app contains high-quality information (eg, text, feedback, measurements, and references) from a credible source. The mean scores and distributions for each section were calculated. The overall MARS-F mean score was the mean score of the engagement, functionality, aesthetics, and information quality sections. Additionally, there is a *subjective* quality section (4 items), which evaluates the user's interest for the app, and a *specificity* section, which assesses perceived effect on the user's knowledge, attitudes, and intentions to change as well as likelihood of changing the identified targeted behaviors (we used daily habits).

Evaluation

Training the Raters for Evaluation

We asked 8 dieticians and nutritionists (Multimedia Appendix 1) to rate the apps. All raters viewed a training video in French (available upon request to the corresponding author) developed

for the MARS-F [22], adapted from the English-language training video [20]. To train, all raters evaluated 2 apps that had been excluded. For this, the raters downloaded and tested each app for at least 15 minutes and fulfilled the questionnaire of MARS-F. When an individual item's rating score differed by at least 2 points, raters discussed until consensus was reached to ensure similar understanding of the item.

App Selection

Among the 15 apps that were included, we randomly selected 7 apps for evaluation by all raters (*Compteur de calories* FatSecret, Yazio - *Régime et Calories*, MyFitnessPal, Macros - *Compteur de calories*, Foodvisor, Lose It! - *Compteur de calories*, and *Compteur de calories*), and the remaining 8 apps were assigned to 4 raters (Lifesum: *Compteur de calories*, Naor, iEatBetter: *Journal alimentaire*, *Le secret du poids*, *Compteur de calories* ScanFood, FeelEat, Kalipi, and *Bonne App*).

The evaluation process took place from April to May 2021. The raters independently used each app for 15 minutes, and then immediately evaluated the app using a web-based MARS-F questionnaire.

Statistical Analysis

To evaluate the interrater reliability, intraclass correlations (2-way random, average measures, absolute agreement) [23,24] and their 95% confidence intervals were calculated for the 7 common apps (for each item, section, and overall). The mean values and standard deviations were calculated for each item and for each section. Item 19 was excluded from all analyses due to missing values.

Scatter plots were used to compare differences between the quality of the apps (for each item and for each section).

The correlation between the overall quality mean and subjective item 23 ("What is your overall star rating of the app?") was evaluated through the Pearson correlation coefficient.

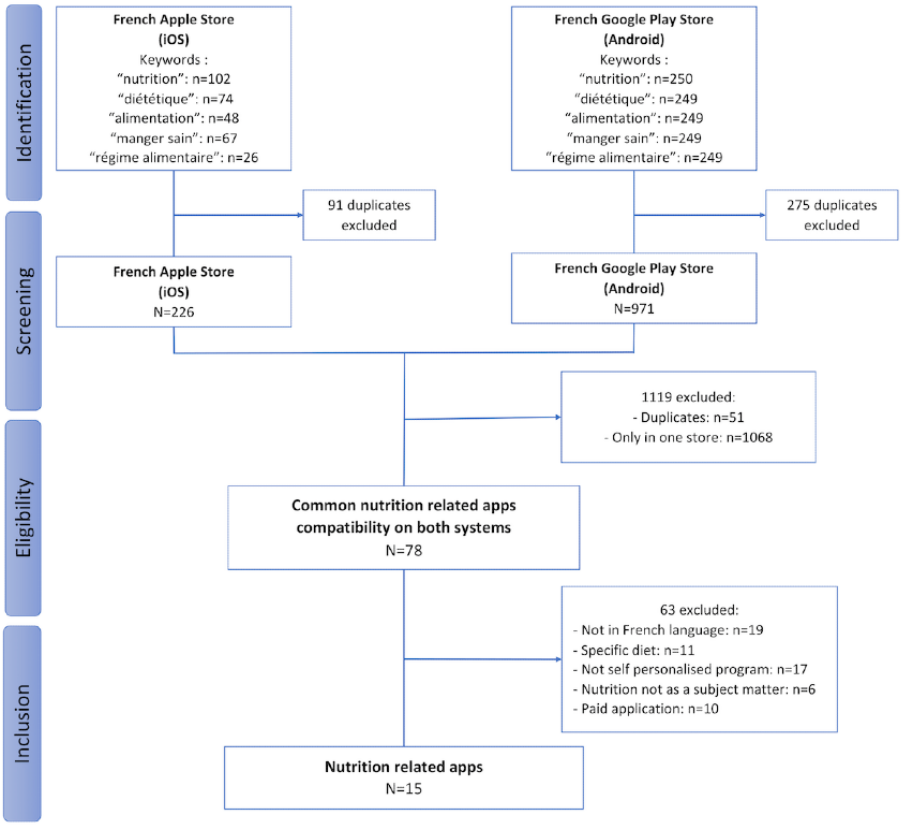
Statistical analyses were performed using Stata (version 15; StataCorp LLC) and using dplyr (version 1.0.8) and ggplot2 (version 3.3.5) packages with R software (version 4.1.1; The R Project for Statistical Computing).

Results

Selection of Mobile Apps

A total of 226 apps in the Apple App Store and 971 apps in the Google Play Store were identified (Figure 1), with 78 apps available on both systems. After screening, 18 apps were preliminarily identified. After downloading, 15 apps were included.

Figure 1. Selection flowchart.



Characteristics of Mobile Apps

No common developer was identified for the 15 apps (Multimedia Appendix 2). Only 2 apps were fully free of charge; 1 app was free for 30 days, and the others required in-app purchases to function completely. MyFitnessPal was the most downloaded app (n=26,804 in the Apple App store and n=2,397,052 downloads in the Android App store), followed by Yazio (n=51,674 in the Apple App store and n=373,162 downloads in the Android App store) and FatSecret (n=3711

in the Apple App store and n=394,958 downloads in the Android App store).

All 15 apps targeted behavior change, goal setting, and physical health (Table 1). Most apps (10/15, 67%) focused on increasing happiness and well-being. The theoretical background and strategies used were (1) information and education, (2) monitoring and tracking, and (3) goal setting. The apps were designed for adults (15/15, 100%), young adults (15/15, 100%), adolescents (13/15, 87%), and children under 12 years (11/15, 73%). All 15 apps (100%) sent reminders, and 10 apps (10/15, 67%) required internet access to function.

Table 1. Characteristics of the 15 nutrition mobile apps.

Characteristic	App (n=15), n (%)[a]
Focus or target	
Increase happiness or well-being	10 (67)
Mindfulness, meditation, or relaxation	3 (20)
Anxiety or stress	3 (20)
Behavior change	15 (100)
Goal setting	15 (100)
Relationships	1 (7)
Physical health	15 (100)
Theoretical background or strategies	
Assessment	10 (67)
Feedback	10 (67)
Information or education	15 (100)
Monitoring or tracking	15 (100)
Goal setting	15 (100)
Advice, tips, strategies, and skills training	9 (60)
Cognitive behavioral therapy - Behavioral (positive events)	5 (33)
Cognitive behavioral therapy - Cognitive (thought challenging)	5 (33)
Acceptance commitment therapy	4 (27)
Mindfulness or meditation	1 (7)
Relaxation	1 (7)
Gratitude	0 (0)
Strengths based	6 (40)
Other	0 (0)
Age group	
Children (under 12 years)	11 (73)
Adolescents (13-17 years)	13 (87)
Young adults (18-25 years)	15 (100)
Adults	15 (100)
Technical aspects of app	
Allows sharing (Facebook, Twitter, etc)	4 (27)
Has an app community	5 (33)
Allows password-protection	9 (60)
Requires log-in	2 (13)
Sends reminders	15 (100)
Needs web access to function	10 (67)

[a]More than one could be applicable; therefore, percentages do not add to 100%.

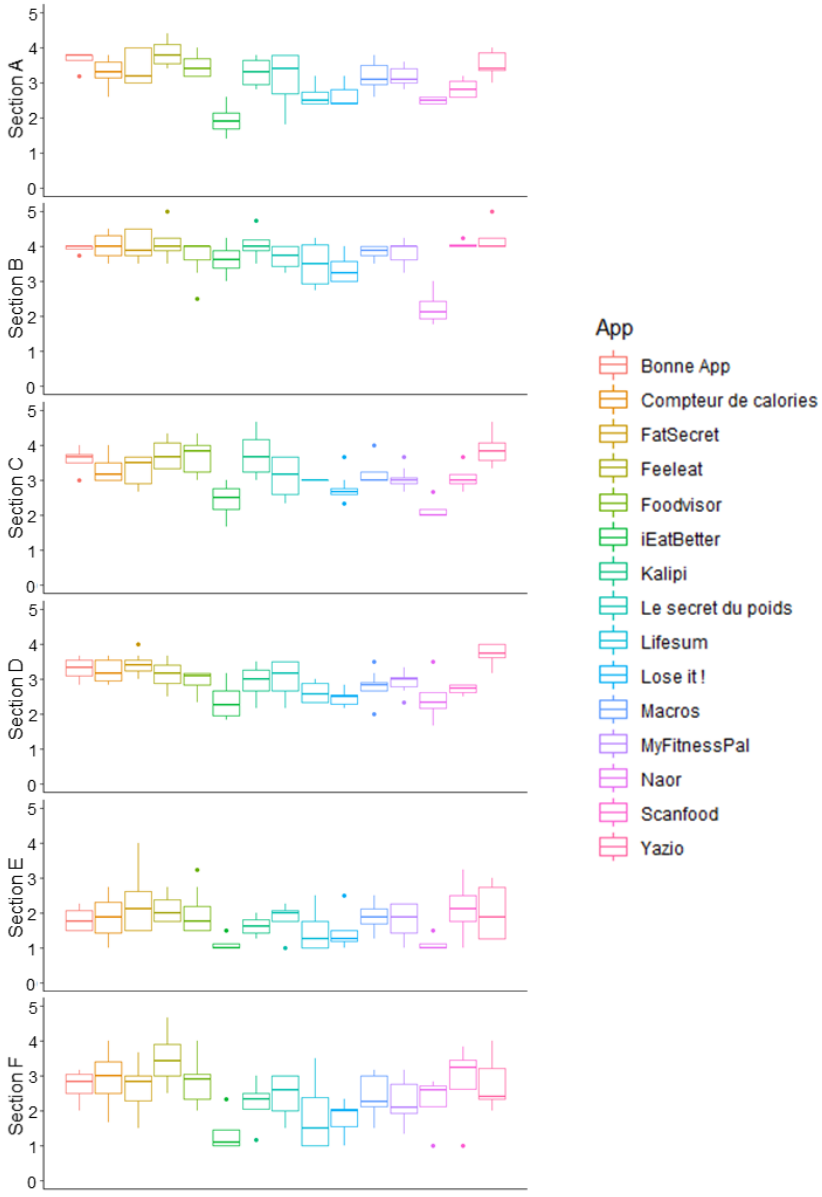
Reliability of the Evaluation

The reliability of the evaluations of the 7 common apps was considered good for overall quality (ICC 0.89, 95% CI 0.70-0.98) and for engagement (ICC 0.83, 95% CI 0.57-0.96), functionality (ICC 0.77, 95% CI 0.45-0.95), and aesthetics (ICC 0.83, 95% CI 0.57-0.97) sections individually. The reliability was excellent for the information quality section (ICC 0.92, 95% CI 0.78-0.98).

Quality of the Content of the Nutrition-Related Mobile Apps

The best quality scores (Figure 2; Multimedia Appendix 3) were obtained by Yazio (mean 3.84, SD 0.32), FeelEat (mean 3.71, SD 0.47), and *Bonne App* (mean 3.65, SD 09); whereas, the worst quality scores were obtained by Naor (mean 2.34, SD 0.39), iEatBetter (mean 2.59, SD 0.40), and Lose It! (mean 2.79, SD 0.29).

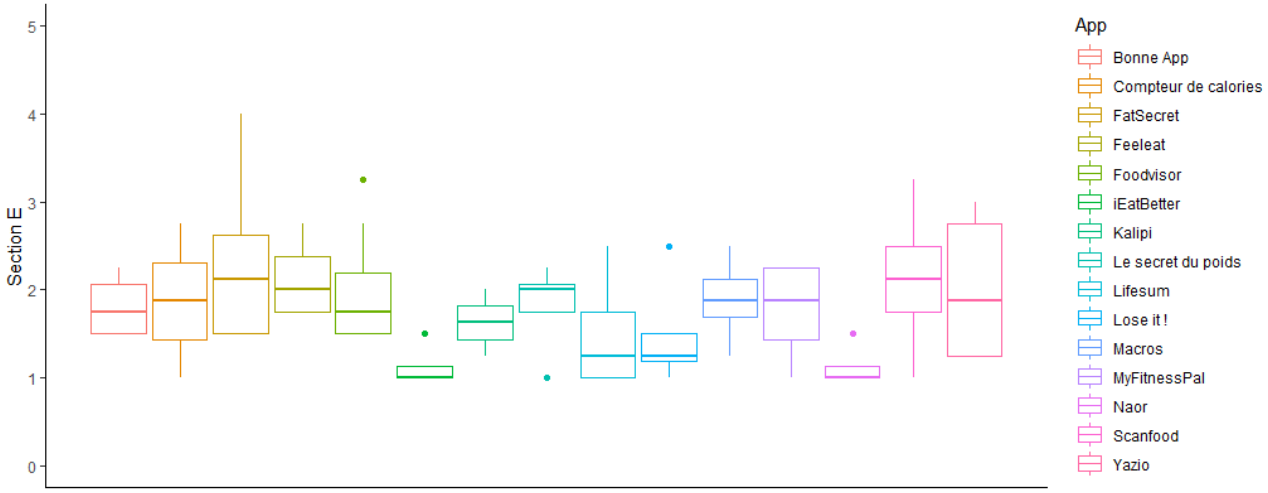
Figure 2. Qualitative evaluation of nutrition-related apps. Section A: Engagement; Section B: Functionality; Section C: Aesthetics; Section D: Information; Section E: Quality.



The engagement scores ranged from a mean of 1.95 (SD 0.5) for iEatBetter to a mean of 3.85 (SD 0.44) for FeelEat. The functionality scores ranged from a mean of 2.25 (SD 0.54) for Naor to a mean of4.25 (SD 0.46) for Yazio. The aesthetics scores ranged from a mean of 2.17 (SD 0.34) for Naor to a mean of 3.88 (SD 0.47) for Yazio. The information quality scores ranged from a mean of 2.38 (SD 0.60) for iEatBetter to a mean of 3.73 (SD 0.29) for Yazio. For all apps, except Naor, the functionality mean score was always higher than the engagement mean score.

The subjective quality scores (Figure 3) ranged from a mean of 1.13 (SD 0.25) for iEatBetter to a mean of 2.28 (SD 0.88) for FatSecret. The best subjective quality scores were obtained by FatSecret (mean 2.28, SD 0.88), FeelEat (mean 2.13, SD 0.48), and ScanFood (mean 2.13, SD 0.92); whereas, the worst quality scores were obtained by iEatBetter (mean 1.13, SD 0.26), Naor (mean 1.13, SD 0.25), and Lose It! (mean 1.41, SD 0.48).

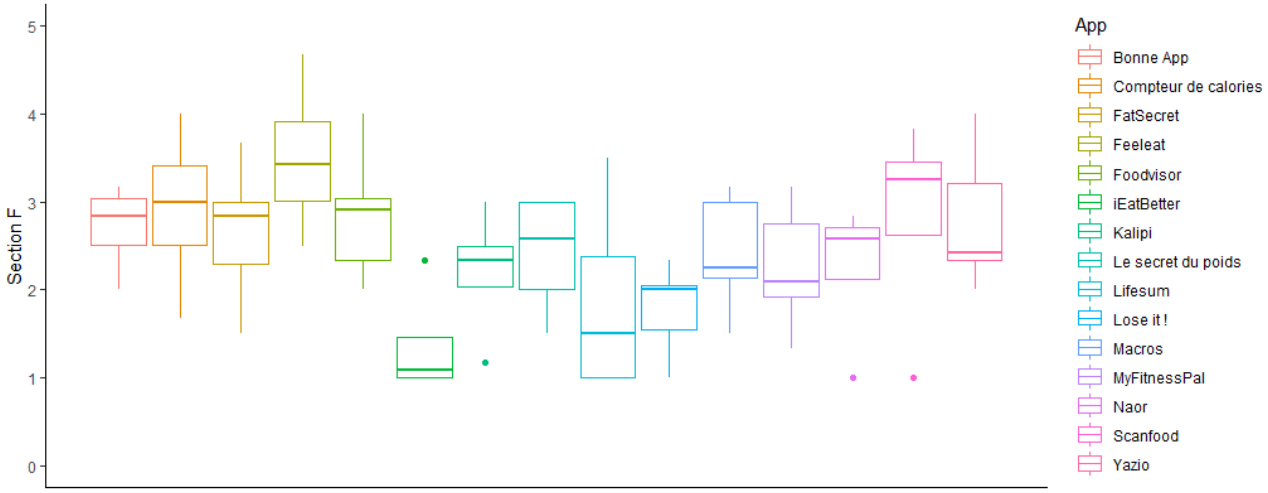
Figure 3. Subjective qualitative evaluation of nutrition-related apps (Section E).



Specificity of the Content of the Nutrition-Related Mobile Apps

Scores for specificity of the content of the apps (Figure 4) ranged from a mean of 1.38 (SD 0.64) for iEatBetter to a mean of 3.50 (SD 0.91) for FeelEat. The best subjective quality scores were obtained by iEatBetter (mean 1.38, SD 0.64), *Compteur de calories* (mean 2.92, SD 0.79), Foodvisor (mean 2.83, SD 0.62), and ScanFood (mean 2.83, SD 1.25); whereas, the worst quality scores were obtained by Lifesum (mean 1.88, SD 1.18), Lose It! (mean 1.79, SD 0.48), and iEatBetter (mean 1.38, SD 0.64).

Figure 4. App-specific scores (Section F).



Strengths and Weaknesses of Each App

The app-specific score was always lower than the subjective quality score, which was always lower than the quality score. This score was lower than the rating score from the iOS or Android app stores (Multimedia Appendix 2; Table S1 in Multimedia Appendix 3).

Low overall quality scores (Figure 5) were due to the information quality scores, for all apps except FatSecret, iEatBetter, Naor, and Yazio. In the information quality section, the worst score was observed for the item regarding the credibility of the app for all apps except for *Bonne App*, which obtained the worst score for goals and quality of information items, and Lifesum, which obtained the worst score for the goals item. In the subjective quality section, low scores were the result of the item indicating whether people be willing to pay for this app. The specificity scores were very close between items for the same app.

Figure 5. Heatmap of the average scores per item and per app, from yellow (1: worst score) to green (5: best score).

	Bonne App	Compteur de calories	FatSecret	Feeleat	Foodvisor	iEatBetter	Kalipi	Le secret du poids	Lifesum	Lose it !	Macros	MyFitnessPal	Naor	Scanfood	Yazio
Section A															
Item 1 - Entertainment	3.75	2.75	3.5	3.5	3.5	1.25	3.5	3	3.25	2.13	3.25	2.63	2.25	3.25	3.63
Item 2 - Interest	3.75	3.38	3.38	4.25	3.5	1.5	3.25	3.25	2.25	2.38	3.38	3.13	2.25	3.5	3.75
Item 3 - Customisation	3.25	3.63	3.38	3.75	3.38	2.25	3	3.25	2.5	3.13	3.25	3.38	3	3	3.38
Item 4 - Interactivity	3.75	3.25	3.38	4	3.38	2	3.25	3	2.25	2.5	2.5	3.38	2.5	1.75	3.25
Item 5 - Target group	3.75	3.5	3.5	3.75	3.75	2.75	3.5	3	3	2.88	3.63	3.38	2.5	2.75	3.63
Section B															
Item 6 - Performance	4	4.38	4.38	4	3.88	3.25	3.75	3.75	3.5	3.38	3.5	4.25	2.5	3.75	4.25
Item 7 - Ease of use	4	3.88	4.25	4.25	3.5	4	4.25	3.75	3.5	3.25	4	3.63	2.25	4.5	4.25
Item 8 - Navigation	3.75	3.75	3.88	4.25	3.75	3.75	4.25	3.75	3.5	3.38	4	3.75	2	4	4.25
Item 9 - Gestural design	4	4.13	3.63	4	3.63	3.5	4	3.5	3.5	3.38	3.88	3.63	2.25	4	4.25
Section C															
Item 10 - Layout	3.75	3.63	3.5	4.25	3.38	3.5	4.25	3.25	3	3	3.38	3.38	2.5	3.75	4.13
Item 11 - Graphics	3.75	3.25	3.63	3.5	3.88	2	3.75	3	3	2.75	3.25	3.13	2.25	2.25	3.88
Item 12 - Visual appeal	3.25	3.13	2.75	3.5	3.75	1.75	3.25	3	3	2.5	3.13	2.63	1.75	3.25	3.63
Section D															
Item 13 - Accuracy	3.75	3.63	4.38	4	3.5	3.25	3.5	3.5	3.5	3.13	3.25	3.5	3	3	3.88
Item 14 - Goals	2.75	3.5	3.88	2.75	2.75	2.5	3	3.5	2	2.13	2.75	3.13	2.75	3	3.63
Item 15 - Quality of information	2.75	3.5	3.5	3.75	2.88	2.25	3	3.25	2.5	2.38	3	3.38	2.5	2	3.75
Item 16 - Quantity of information	3.5	3.38	3.38	3.75	3.63	2.5	3.25	3.25	2.5	2.88	3	2.75	3	2.75	3.88
Item 17 - Visual information	4	4.38	4.38	3.5	3.75	2.75	3.75	3.5	2.25	3.25	3.88	3.75	2.5	4.5	4.25
Item 18 - Credibility	3	1	1	1	1	1	1	1	3	1	1	1	1	1	3
Section E															
Item 20 - Recommendations	1.75	2	2.38	2.5	2.25	1	1.5	2	1.75	1.38	2.13	1.88	1	2.5	2.25
Item 21 - Usage	1.5	1.75	2.5	1.5	1.5	1	1.25	1.75	1.25	1.13	1.5	1.88	1	2.25	1.88
Item 22 - Price	1	1	1.25	1	1.25	1	1	1	1	1	1	1	1	1	1
Item 23 - Overall rating	3	2.75	3	3.5	3	1.5	2.75	2.5	2	2.13	3	2.38	1.5	2.75	2.88
Section F															
Awareness	2.5	2.63	2.63	3.25	2.88	1.25	2.5	2.25	1.75	1.88	2.38	2.25	2.25	3	2.75
Knowledge	2.5	2.63	2.38	3.5	2.63	1.25	2	2.5	2	1.75	2.13	1.88	2	2.5	2.5
Attitude	2.75	3.13	3.13	3.75	3	1.25	2.25	2.25	2	1.75	2.63	2	2.5	2.75	3.13
Intention to change	3	3.13	2.63	3.75	3.13	1.25	2.25	2.5	2	2	2.38	2.63	2	3	3.13
Help seeking	2.5	3	2.63	3.25	2.38	1.75	2	2.5	1.5	1.75	2.75	2.5	2.5	3	2.5
Behaviour change	3	3	2.63	3.5	3	1.5	2.25	2.5	2	1.63	2.25	2.25	2.25	2.75	2.75

Correlation Between MARS and Stars Ratings

The correlation between the quality mean and the subjective item 23 (“What is your overall star rating of the app?”) was considered to be good (r=0.67, P<.001) and indicated that the quality score (overall) was generally higher than that of subjective item 23.

Correlation analysis between overall MARS-F scores of the apps and their respective store ratings was limited by the availability of store ratings and the discrepancies among the number of raters. The store ratings were higher than overall MARS-F mean scores. Store ratings ranged from 3.0 (*Le secret du poids*) to 4.9 (ScanFood) for the iOS store and from 3.0 (ScanFood) to 4.7 (FatSecret) for the Android store.

Discussion

Principal Findings

The increasing public consciousness and high comorbidity burden related to unhealthy nutrition has highlighted the necessity of a healthy diet [25]. Nutrition behaviors can be improved by using mobile health apps, which have become very popular [26]. For diabetes [27], renal disease [28], weight loss [29], and age-related macular degeneration [30], health and nutrition professionals have used mobile health apps to monitor and encourage better lifestyle and dietary choices. The use of mobile health apps has also been found to increase adherence to dietary monitoring [29,31].

Screening of nutrition-related apps available in the French Apple App and Google Play App stores yielded 15 apps. In another study, screening of the Korean Apple app and Google Play stores yielded 29 nutrition-related apps [32]; the study [32] used 2 criteria—including only apps rated 4 stars or higher and the top 100 most reviewed apps. Another study, screening of the US Google Play Store yielded 86 apps, but the criterion for inclusion (only apps rated 4 stars or higher) was less restrictive [33].

All 15 apps targeted behavior change, physical health, and goal setting via information and education and monitoring and tracking. A previous study [34] showed that diet monitoring and education were the most frequently used functions in diet and nutrition apps [34]. All ages were targeted for 73.3% (11/15) of the apps; this finding is consistent with that from another study [33], which found that 94% of diet and nutrition apps appealed to users of all ages.

Ratings in the iOS and Android stores were higher than the MARS-F quality scores. Star ratings and user comments are valuable to users because they provide insight into the effectiveness and popularity of apps [33], but star ratings do not provide objective assessment of quality. In contrast to other studies [32,33], our study did not used star ratings as an inclusion criteria; however, for the 15 apps included (except

Bonne App for which the number of raters or downloads was not sufficient) the star score was greater than 4.

Quality scores were greater than 2.5, except for Naor (mean 2.34, SD 0.39). Functionality was the strength for all apps, except for Naor. The high scores could be explained by the inclusion of scroll and zoom features to increase readability. The maximum score of 5 was obtained for the ease of use, navigation, and gestural design (FeelEat, iEatBetter, Yazio, and FatSecret). In contrast, Naor navigation was rated low (mean 2, SD 0.82), which could be explained by difficulty in accessing the menu, the amount of data, and the design. The weakness of all the apps, except FatSecret, iEatBetter, Naor, and Yazio, was information quality. The worst score in the information quality section was typically for credibility of the app. This corresponded to the fact that the source of information was identified, but the source's validity or reliability was questionable (eg, commercial enterprise with vested interest). Moreover, the level of scientific evidence was difficult to evaluate. The evaluators selected "N/A The app has not been tested" in most cases; therefore, this item was not included in the statistical analysis. To the best of our knowledge, only 5 of the 15 apps (FatSecret [8,35-38], Lifesum [8,35,38], MyFitnessPal [8,35-45], Yazio [35,36,38], and Lose It! [8,35,37]) are indexed in PubMed. On the other hand, the information contained in these nutrition-related apps may have errors. For example, FatSecret, Lifesum, MyFitnessPal, and Yazio tended to underestimate total energy intake [38].

The subjective quality score were always lower than the star rating scores from the iOS and Android stores. This can be explained because the evaluations in the stores are made by all the users; whereas, in our study, dieticians or nutritionists assessed the apps using the MARS-F. Indeed, the use of user version of MARS can show different results [21].

Moreover, subjective quality scores were also lower than quality scores. This indicates that even if engagement, functionality, aesthetics, and information quality for an app were good, professionals did not think that they would use the app often in the next 12 months, and they would not be willing to pay for the app. This finding can be compared with the results of an international survey of health care professionals' opinions on nutrition and diet apps [46]. Among 1001 health care professionals questioned, only 45.5% recommended these types apps to their patients. Surprisingly, 22.5% of people who had not yet recommended the use of these types of apps did not know of their existence. Health care professionals who have recommended apps may have used them as supplementary tools to broaden their daily practice, engage patients, enhance care, and possibly contribute to the reduction in health care costs [47]. Additionally, patients living with diseases such as diabetes or obesity may use apps for self-monitoring of their diet and physical activity [48].

Generally, raters shared a common negative opinion on the potential impact of the nutrition-related apps on the behavior change (macro micronutrients intakes), even if these apps have already demonstrated positive results in with respect to the prevention of being overweight or other chronic disease [34,49-51].

Limitations

This study has several limitations. First, only nutrition-related apps available on both Apple and Android French stores, were included. Other stores, such as the Huawei store, the Samsung store, the Windows phone store, or BlackBerry, could have been investigated. Second, we chose to use the French version of the MARS because this scale is the most commonly used in scientific literature for mobile health app evaluation to date [52-57]. However, other scales, such as ENLIGHT or Application Quality Evaluation, which was initially specifically developed for the evaluation of mobile health app linked to nutrition purpose [58], could have been used. Third, the assessment was conducted by dieticians and nutritionists; whereas mobile health apps are intended for the general public. In further investigations, a comparison between ratings with the user version of MARS [21] and those from our study could be interesting.

Perspectives

In a recent study [48], clinicians mentioned that nutrition apps may improve patient outcomes when compared to traditional methods of monitoring dietary and physical activity behaviors [48]. Nutrition-related apps are appealing to users, based on the high number of downloads, which supports the fact that diet intake monitoring and recommendations could be managed through these tools [8,59,60]. Thus, the findings of our study could help French users of mobile apps and professionals to select the best nutrition-related apps in terms of quality and to choose the most appropriate health literacy elements. Furthermore, when used as part of an empowerment strategy, the app must adapt to the user's chronic disease.

The implementation of new therapeutic programs that integrate mobile apps associated with follow-up with health professionals could be a key element in changing behavior. On one hand, it is important to remain vigilant with respect to the ethical issues surrounding the use of health data and the development of apps for commercial purposes. On the other hand, the discrepancy between scores obtained for the subjective quality section and for those for the specificity of the apps demonstrated that, although nutrition-related apps could be a key element in modifying the nutritional behavior of patients, for this, it is necessary to integrate the nutrition-related apps in professional practice. It would be interesting to conduct randomized clinical trials or longitudinal studies, using the 15 nutrition-related mobile apps identified in this study, to analyze nutritional behavioral modification from use of the apps and impacts on noncommunicable diseases.

Acknowledgments

We would like to thank the 8 dieticians for their participation (Brigitte Boniteau, Hélène Quelennec, Sarah Achatoui, Anais Chouin, Laure Poindextre, Florian Kwasniewski, Claire Neveu, and NRO), and Stéphane Béchet for statistical analysis expertise.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the raters, the hardware, and the software used.
[PDF File (Adobe PDF File), 61 KB - mhealth_v10i3e35879_app1.pdf]

Multimedia Appendix 2
Descriptive and technical information of the nutrition mobile apps.
[PDF File (Adobe PDF File), 148 KB - mhealth_v10i3e35879_app2.pdf]

Multimedia Appendix 3
Mobile App Rating Scale (MARS) scoring.
[PDF File (Adobe PDF File), 132 KB - mhealth_v10i3e35879_app3.pdf]

References

1. Bennett JE, Stevens GA, Mathers CD, Bonita R, Rehm J, Kruk ME, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards sustainable development goal target 3.4. Lancet 2018 Sep;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5]
2. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
3. Celis-Morales C, Lara J, Mathers JC. Personalising nutritional guidance for more effective behaviour change. Proc Nutr Soc 2015 May;74(2):130-138. [doi: 10.1017/S0029665114001633] [Medline: 25497396]
4. Kraef C, Wood B, von Philipsborn P, Singh S, Peterson SS, Kallestrup P. Primary health care and nutrition. Bull World Health Organ 2020 Dec 01;98(12):886-893 [FREE Full text] [doi: 10.2471/BLT.20.251413] [Medline: 33293749]
5. Ryu S. Book review: mhealth: new horizons for health through mobile technologies: based on the findings of the second global survey on ehealth (global observatory for ehealth series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
6. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of ehealth applications: a scoping review. Int J Med Inform 2019 Dec;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
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. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
9. 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]
10. Jusoh S. A survey on trend, opportunities and challenges of mhealth apps. Int J Interact Mob Technol 2017 Nov 27;11(6):73. [doi: 10.3991/ijim.v11i6.7265]
11. Rollo ME, Haslam RL, Collins CE. Impact on dietary intake of two levels of technology-assisted personalized nutrition: a randomized trial. Nutrients 2020 Oct 29;12(11):3334 [FREE Full text] [doi: 10.3390/nu12113334] [Medline: 33138210]
12. Attwood S, van Sluijs E, Sutton S. Exploring equity in primary-care-based physical activity interventions using PROGRESS-Plus: a systematic review and evidence synthesis. Int J Behav Nutr Phys Act 2016 May 20;13(1):60 [FREE Full text] [doi: 10.1186/s12966-016-0384-8] [Medline: 27206793]
13. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: a systematic literature review. J Telemed Telecare 2015 Feb 12:243-253. [doi: 10.1177/1357633X15572203] [Medline: 25680388]
14. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020;3:110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]

15. Lim SY, Lee KW, Seow W, Mohamed NA, Devaraj NK, Amin-Nordin S. Effectiveness of integrated technology apps for supporting healthy food purchasing and consumption: a systematic review. Foods 2021 Aug 12;10(8):1861 [FREE Full text] [doi: 10.3390/foods10081861] [Medline: 34441638]
16. La langue française dans le monde. Gallimard. URL: https://www.gallimard.fr/Catalogue/GALLIMARD/Gallimard-Education/La-Bibliotheque-Gallimard/Hors-serie/La-langue-francaise-dans-le-monde# [accessed 2021-12-01]
17. 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]
18. Masterson CRM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
19. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
20. 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]
21. 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;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
22. Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, et al. Promoting health via mhealth applications using a french version of the mobile app rating scale: adaptation and validation study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480 [FREE Full text] [doi: 10.2196/30480] [Medline: 34463623]
23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
24. 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]
25. Martinon P, Fraticelli L, Giboreau A, Dussart C, Bourgeois D, Carrouel F. Nutrition as a key modifiable factor for periodontitis and main chronic diseases. J Clin Med 2021 Jan 07;10(2):197 [FREE Full text] [doi: 10.3390/jcm10020197] [Medline: 33430519]
26. Goodman S, Morrongiello B, Randall Simpson J, Meckling K. Vitamin D intake among young Canadian adults: validation of a mobile vitamin D calculator app. J Nutr Educ Behav 2015;47(3):242-247. [doi: 10.1016/j.jneb.2014.11.006] [Medline: 25959447]
27. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
28. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
29. 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 Oct;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
30. Ali ZC, Silvioli R, Rajai A, Aslam TM. Feasibility of use of a mobile application for nutrition assessment pertinent to age-related macular degeneration (MANAGER2). Transl Vis Sci Technol 2017 Jan;6(1):4 [FREE Full text] [doi: 10.1167/tvst.6.1.4] [Medline: 28138414]
31. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
32. Choi J, Chung C, Woo H. Diet-related mobile apps to promote healthy eating and proper nutrition: a content analysis and quality assessment. Int J Environ Res Public Health 2021 Mar 28;18(7):3496 [FREE Full text] [doi: 10.3390/ijerph18073496] [Medline: 33800531]
33. Schumer H, Amadi C, Joshi A. Evaluating the dietary and nutritional apps in the Google Play Store. Healthc Inform Res 2018 Jan;24(1):38-45 [FREE Full text] [doi: 10.4258/hir.2018.24.1.38] [Medline: 29503751]
34. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Hardy D, Smith SA. Smartphone applications for promoting healthy diet and nutrition: a literature review. Jacobs J Food Nutr 2015;2(3):021 [FREE Full text] [Medline: 26819969]
35. Maringer M, Wisse-Voorwinden N, Veer P, Geelen A. Food identification by barcode scanning in the Netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications. Public Health Nutr 2019 May;22(7):1215-1222. [doi: 10.1017/S136898001800157X] [Medline: 29962361]

36. Bzikowska-Jura A, Sobieraj P, Raciborski F. Low comparability of nutrition-related mobile apps against the Polish reference method-a validity study. Nutrients 2021 Aug 20;13(8):2868 [FREE Full text] [doi: 10.3390/nu13082868] [Medline: 34445026]
37. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838 [FREE Full text] [doi: 10.2196/mhealth.9838] [Medline: 30785409]
38. Tosi M, Radice D, Carioni G, Vecchiati T, Fiori F, Parpinel M, et al. Accuracy of applications to monitor food intake: evaluation by comparison with 3-d food diary. Nutrition 2021 Apr;84:111018. [doi: 10.1016/j.nut.2020.111018] [Medline: 33046348]
39. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209. [doi: 10.2196/12209] [Medline: 30816851]
40. Mangieri CW, Johnson RJ, Sweeney LB, Choi YU, Wood JC. Mobile health applications enhance weight loss efficacy following bariatric surgery. Obes Res Clin Pract 2019;13(2):176-179. [doi: 10.1016/j.orcp.2019.01.004] [Medline: 30826256]
41. Patel ML, Brooks TL, Bennett GG. Consistent self-monitoring in a commercial app-based intervention for weight loss: results from a randomized trial. J Behav Med 2020 Jun;43(3):391-401. [doi: 10.1007/s10865-019-00091-8] [Medline: 31396820]
42. Evenepoel C, Clevers E, Deroover L, Van Loo W, Matthys C, Verbeke K. Accuracy of nutrient calculations using the consumer-focused online app MyFitnessPal: validation study. J Med Internet Res 2020 Oct 21;22(10):e18237 [FREE Full text] [doi: 10.2196/18237] [Medline: 33084583]
43. Stehr P, Karnowski V, Rossmann C. The multi-faceted usage patterns of nutrition apps: a survey on the appropriation of nutrition apps among German-speaking users of MyFitnessPal. BMC Med Inform Decis Mak 2020 Oct 28;20(1):279 [FREE Full text] [doi: 10.1186/s12911-020-01294-9] [Medline: 33115444]
44. Downs DS, Savage JS, Rivera DE, Pauley AM, Leonard KS, Hohman EE, et al. Adaptive, behavioral intervention impact on weight gain, physical activity, energy intake, and motivational determinants: results of a feasibility trial in pregnant women with overweight/obesity. J Behav Med 2021 Oct;44(5):605-621. [doi: 10.1007/s10865-021-00227-9] [Medline: 33954853]
45. Bardus M, Borgi C, El-Harakeh M, Gherbal T, Kharroubi S, Fares E. Exploring the use of mobile and wearable technology among university student athletes in Lebanon: a cross-sectional study. Sensors (Basel) 2021 Jun 30;21(13):4472 [FREE Full text] [doi: 10.3390/s21134472] [Medline: 34208798]
46. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. Nutrients 2020 Jul 24;12(8):2214 [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
47. Ward R, Stevens C, Brentnall P, Briddon J. The attitudes of health care staff to information technology: a comprehensive review of the research literature. Health Info Libr J 2008 Jun;25(2):81-97 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00777.x] [Medline: 18494643]
48. Karduck J, Chapman-Novakofski K. Results of the clinician apps survey, how clinicians working with patients with diabetes and obesity use mobile health apps. J Nutr Educ Behav 2018 Jan;50(1):62-69.e1. [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
49. Fakih El Khoury C, Karavetian M, Halfens RJG, Crutzen R, Khoja L, Schols JMGA. The effects of dietary mobile apps on nutritional outcomes in adults with chronic diseases: a systematic review and meta-analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651. [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
50. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484. [doi: 10.1111/obr.12903] [Medline: 31353783]
51. Recio-Rodriguez JI, Agudo Conde C, Calvo-Aponte MJ, Gonzalez-Viejo N, Fernandez-Alonso C, Mendizabal-Gallastegui N, EVIDENT Investigators. The effectiveness of a smartphone application on modifying the intakes of macro and micronutrients in primary care: a randomized controlled trial. the EVIDENT II study. Nutrients 2018 Oct 10;10(10):1473 [FREE Full text] [doi: 10.3390/nu10101473] [Medline: 30309008]
52. Talwar D, Yeh Y, Chen W, Chen L. Characteristics and quality of genetics and genomics mobile apps: a systematic review. Eur J Hum Genet 2019 Jun;27(6):833-840 [FREE Full text] [doi: 10.1038/s41431-019-0360-2] [Medline: 30809045]
53. Chen R, Santo K, Wong G, Sohn W, Spallek H, Chow C, et al. Mobile apps for dental caries prevention: systematic search and quality evaluation. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e19958 [FREE Full text] [doi: 10.2196/19958] [Medline: 33439141]
54. Ciapponi A, Donato M, Gülmezoglu AM, Alconada T, Bardach A. Mobile apps for detecting falsified and substandard drugs: a systematic review. PLoS One 2021;16(2):e0246061 [FREE Full text] [doi: 10.1371/journal.pone.0246061] [Medline: 33539433]
55. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]

56. 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]
57. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the mobile app rating scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
58. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for nutrition app quality evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163 [FREE Full text] [doi: 10.2196/mhealth.7441] [Medline: 29079554]
59. 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]
60. Braz VN, de Moraes Lopes MHB. Evaluation of mobile applications related to nutrition. Public Health Nutr 2019 May;22(7):1209-1214. [doi: 10.1017/S136898001800109X] [Medline: 29734965]

Abbreviations

ICC: intraclass correlation
MARS: Mobile Application Rating Scale
MARS-F: Mobile Application Rating Scale–French

Edited by L Buis; submitted 21.12.21; peer-reviewed by F Denis, H Mehdizadeh; comments to author 29.12.21; revised version received 01.01.22; accepted 27.01.22; published 14.03.22.

Please cite as:
Martinon P, Saliasi I, Bourgeois D, Smentek C, Dussart C, Fraticelli L, Carrouel F
Nutrition-Related Mobile Apps in the French App Stores: Assessment of Functionality and Quality
JMIR Mhealth Uhealth 2022;10(3):e35879
URL: https://mhealth.jmir.org/2022/3/e35879
doi:10.2196/35879
PMID:35285817

Review

Persuasive Technology in an mHealth App Designed for Pelvic Floor Muscle Training Among Women: Systematic Review

Aida Jaffar[1,2], MMed Fam Med; Chai-Eng Tan[1,3], MMed Fam Med; Sherina Mohd-Sidik[1], PhD; Novia Admodisastro[4], PhD; Felicity Goodyear-Smith[5], MBChB, MGP, MD

[1]Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia

[2]Primary Care Unit, Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia

[3]Department of Family Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[4]Software Engineering & Information System Department, Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, Selangor, Malaysia

[5]Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand

Corresponding Author:
Sherina Mohd-Sidik, PhD
Department of Psychiatry
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
Serdang
Selangor, 43400
Malaysia
Phone: 60 03 9769 2541
Fax: 60 03 9769 2706
Email: sherina@upm.edu.my

Abstract

Background: Pelvic floor muscle training (PFMT) is one of the first-line treatments for stress urinary incontinence among pregnant women. Mobile health (mHealth) technology is potentially effective for delivering PFMT to pregnant women. Persuasive technology in the development of such mobile apps may facilitate behavior change by improving adherence to the exercises. The Capability, Opportunity, and Motivation–Behavior (COM-B) model is potentially useful in selecting the appropriate interventions to be incorporated into the apps.

Objective: This review of mHealth apps for PFMT aims to describe the principles of persuasion used for each app and to propose mHealth app design features based on the COM-B model.

Methods: A systematic literature search was conducted to answer three main research questions: what are the available mHealth apps for PFMT in the published literature, what persuasive strategies were used in their studies how were they mapped to the COM-B model, and how effective were the selected persuasive strategies for PFMT adherence? We searched PubMed, CINAHL, Web of Science, Scopus, and local Malaysian databases such as MyCite and MyMedR for articles reporting mHealth apps used for the delivery of PFMT. We included original articles reporting experimental and cross-sectional studies, including pilot or feasibility trials. Systematic and narrative reviews were excluded. Narrative and thematic syntheses were conducted on the eligible articles based on the research questions. The Cochrane risk of bias tool and the Risk of Bias Assessment Tool for Non-randomized Studies were used to assess study bias.

Results: Of the 169 records from the initial search, 10 (5.9%) articles meeting the selection criteria were included in this review. There were 8 mHealth apps designed for the delivery of PFMT. The Tät, which used 3 categories of persuasive system design, improved PFMT adherence and was cost-effective. Only 1 app, the iBall app, used all categories of persuasive system design, by including social support such as "competition" in its design. The Diário Saúde app was the only app developed using operant conditioning. All apps incorporated Tailoring and Expertise as part of their PSD strategies. Only 3 apps, the Diário Saúde, Tät, and Pen Yi Kang demonstrated improved PFMT adherence.

Conclusions: Persuasive technology used in mobile apps may target desired behavior change more effectively. The persuasive system design can be mapped to the COM-B model to explain its effectiveness on behaviour change outcomes.

(JMIR Mhealth Uhealth 2022;10(3):e28751) doi:10.2196/28751

KEYWORDS

urinary incontinence; pelvic floor muscle training; mHealth app; persuasive technology; capability, opportunity, and motivation–behavior model; mobile phone

Introduction

Background

Urinary incontinence (UI) is defined as involuntary urinary leakage or inability to control urine. Various physiological changes during pregnancy, including collagen changes, hormonal changes, and increased uterine and fetal weight, contribute to the weakening of the pelvic floor muscles (PFMs) during pregnancy [1]. Approximately 42% of women experience their first UI during pregnancy, and up to 31% of parous women have UI [2]. Women with persistent UI after delivery may continue to experience UI for another 12 years [3]. Therefore, UI during pregnancy may be an essential risk factor for subsequent UI among women.

UI is troublesome, particularly during pregnancy, and affects women's quality of life physically, emotionally, spiritually, and financially [4-6]. Those who experience it may resort to various methods to deal with the problem, including using pads or incontinence diapers and avoiding social situations because of embarrassment. However, many pregnant women do not seek help despite having UI symptoms because of the perception of UI as a normal pregnancy change or embarrassment to initiate a discussion about UI with their health care provider [7,8]. Some are unaware of treatment availability, such as PFM training (PFMT), whereas others feel they should not disturb their health care provider as UI is a temporary issue [4].

PFMT is a repetitive, voluntary contraction and relaxation of specific PFMs that is recommended for managing and preventing UI during pregnancy and after delivery [9,10]. Focused PFMT is useful for strengthening the PFMs and reducing UI [11]. PFMT is low-cost and noninvasive and has benefits for the prevention and treatment of UI among pregnant women [9].

However, PFMT requires adherence and correct technique to work. Lack of knowledge and skills for PFMT is a barrier to successful management of UI [12,13]. Women report lack of self-confidence, difficulty remembering, and time constraints [7,14-16] as barriers to performing PFMT. This results in the underuse of PFMT by those who experience UI.

Antenatal health care providers need training on how to teach PFMT, and some have difficulties in allocating time for teaching PFMT to pregnant women [7]. In addition, no standard national guidelines for PFMT are available to guide health care providers regarding the appropriate frequency of PFMT and other technical details [4,7]. Therefore, efforts to develop mobile health (mHealth) apps to deliver PFMT have been made to address these challenges.

mHealth is defined as "the use of wireless communication devices to support public health and clinical practice" [17]. mHealth apps are software apps used by health care professionals and patients for conveying health knowledge, research to improve health treatments, and public health [18]. mHealth apps can be classified based on the target users: health care professionals and patients. mHealth apps for patients can be further divided into five subcategories: (1) educational health apps, (2) apps to contact health care professionals, (3) apps to check personal health records, (4) personal care apps, and (5) social networking apps [18].

mHealth apps for PFMT are personal care apps that assist users in training their PFMs, akin to a personal trainer. The apps aim to modify negative attitudes among users to successfully produce behavior change. Developing a positive attitude is essential to enable people to change their behavior successfully [19]. The use of persuasive technology (PT) in mHealth apps may support attitude and behavior change in users to adopt PFMT as part of their lifestyle.

PT in App Design

PT is defined as "interactive computing systems designed to aid and motivate people to alter their attitude and behaviours" [20]. PT can be categorized according to its functional roles, which are tools, media, or social actors [20]. Tools help users perform a target behavior by making the task easier or restructuring the task. Media represents the content that supports users in repeating a behavior or providing emotional support that facilitates the target behavior. Social actors are cues for social responses that promote the target behavior. These functions can be incorporated into the design of an mHealth app for behavior change.

Oinas-Kukkonan [21] defined persuasive systems as "computerized software or information systems designed to reinforce, change or shape attitudes or behaviors or both without using coercion or deception." Adopting a persuasive system design (PSD) may provide effective persuasion [22]. There are three potential successful outcomes for a persuasive system: voluntary reinforcement and change or shaping of attitudes or behaviors. PSD for PFMT apps can be divided into four main categories—primary task support, dialogue support, credibility support, and social support—as seen in Table 1.

App designers need to translate theoretical determinants of behavior into technology design items [23]. Identifying suitable PT elements for the design of PFMT mHealth apps should suit the context of pregnant women who wish to adopt PFMT and the evolving physiological changes of pregnancy.

Table 1. Persuasive system design (PSD) and suggestions for the pelvic floor muscle training (PFMT) apps.

PSD category and subcategories	Suggestions for PFMT app features
Primary task support	
Reduction	The app lists effective intervention for UI[a].
Tailoring	The app provides PFMT information according to the target user characteristics.
Personalization	The app provides personalized content according to the individual user.
Self-monitoring	Users are able to monitor their progress.
Dialogue support	
Rewards	A trophy is given after the user has completed PFMT schedule.
Reminders	This is a crucial feature because PFMT needs to be done daily.
Liking	Likable minimalist design with user's choice of colors
Social role	Use of a virtual physiotherapist
Credibility support	
Trustworthiness	The app provides unbiased information regarding PFMT.
Expertise	The app provides the professional background and expertise of the content developers.
Surface credibility	The app appears professional.
Authority	The app bears the logo of the developer's institution.
Third-party endorsements	The app includes endorsing statements from relevant professionals such as physiotherapists or urogynecologists.
Social support	
Social learning	The app allows users to see the deidentified general performance of all users.
Social comparison	The app also allows users to share their achievements with other users.
Normative influence	The app normalizes the experience of UI and learning PFMT by connecting a user with other similar users.

[a]UI: urinary incontinence.

The Capability, Opportunity, and Motivation–Behavior Model

The Capability, Opportunity, and Motivation–Behavior (COM-B) model has great potential as a theory to guide the selection of persuasive design elements for app development [24]. It is a simple, validated model derived from combining various behavior change theories.

The *Capability* domain in COM-B proposes that a person must acquire correct knowledge, skills, and abilities to allow them to perform a targeted behavior. In the context of pregnant women and PFMT, capability can be classified into two subcategories: (1) physical capability (referring to the woman's physical skills) and (2) psychological capability (referring to the woman's understanding of PFMT and ability to remember to practice PFMT). Both capabilities are required to affect behavior change. Pregnant women in primary care settings lack the necessary knowledge about PFMT, whereas those with good PFMT knowledge may still have poor practices as a result of poor physical capability [12]. Hence, improving PFMT knowledge and skills is a crucial aspect of intervention, which can be addressed through training such as via an educational video or a group intervention session with a physiotherapist.

The *Opportunity* domain in COM-B refers to external factors that influence the targeted behavior. Again, in the context of pregnant women and PFMT, opportunity may consist of (1) physical opportunity (availability of environmental, nonliving elements that support PFMT, such as a private location to learn and perform PFMT) and (2) social opportunity (related to the opportunities afforded by people around the pregnant women, eg, cultural and social norms related to PFMT). The opportunity to learn PFMT in their own comfort and privacy is a form of physical opportunity that can be created by mHealth apps. Having health care providers promote the use of apps to learn PFMT can be a form of social opportunity. Incorporating a function in which users can post questions or communicate with experts regarding UI improves accessibility to knowledge and may also represent the creation of social opportunity for pregnant women [25].

The *Motivation* domain refers to an internal process that influences the decision to change and adopt the targeted behavior. Motivation is divided into (1) reflective motivation (referring to reflective thought processes that improve motivation, eg, identifying solutions to schedule PFMT thrice daily in their daily activities) and (2) automatic motivation (emotional or habitual processes that improve motivation, eg, reinforcing their PFMT routine through reminders and rewards).

The development of effective evidence-based behavioral interventions should incorporate suitable health behavior theories with qualitative and quantitative evidence [26,27]. Previous PFMT interventions have been designed based on various health behavior theories such as the social cognitive theory and the theory of planned behavior [28]. Social cognitive theory is limited by its focus on social influences and self-efficacy, whereas the theory of planned behavior focuses on motivation for PFMT adherence. Therefore, a more comprehensive behavior change theoretical framework that encompasses major theories is desirable [24,29]. The strength of the COM-B model lies in its simplicity in identifying the key components that affect a person's motivation and behavior.

The aim of this review is to (1) list the mHealth apps designed for PFMT, (2) determine the PSD used, and (3) suggest PSD features for new mHealth apps that incorporate the COM-B model.

Methods

Overview

We conducted a systematic literature search for published articles on mHealth apps for PFMT among women in December 2020. For the purpose of this paper, *mHealth app* is defined as an app developed for delivery via a mobile device, including a tablet device, for health care purposes [30]. This definition excludes email, websites, telemedicine, and telehealth.

This literature search aimed to answer three main research questions: (1) What are the available mHealth apps for PFMT in the published literature? (2) How effective were the selected persuasive strategies for PFMT adherence? and (3) What persuasive strategies were used in their studies and how did they map to the COM-B model?

The search was conducted in the PubMed, CINAHL, Web of Science, and Scopus databases, which are reference databases in biomedicine and rehabilitation and comprise the largest general scientific databases [31]. We also searched the Malaysian databases MyCite and MyMedR using the following search keywords: *(pelvic floor muscle training OR pelvic floor muscle exercise OR Kegel exercise) AND (women) AND (digital health OR mhealth OR mobile health OR mobile application OR smartphone OR mobile app OR smartphone app)*. The initial intent of this review was to look at the persuasive strategies used by mobile apps for PFMT in pregnant women. However, because of the extremely limited number of published articles specifically on pregnant women, we expanded the population to include women in general.

Selection Criteria

The selection criteria for this review are listed in Textbox 1.

Titles, abstracts, and full-text articles were further assessed and discussed by 2 reviewers (AJ and CET). Consensus for article inclusion was achieved through discussion with a third reviewer (SMS). Further discussion focusing on persuasion was undertaken with NA, who is the expert in the area. In addition, a manual search was conducted using reference lists of selected articles to identify any additional suitable records that were not included in the database search results.

Textbox 1. Selection criteria for the literature search.

Inclusion criteria

- Population: mobile health (mHealth) apps for pelvic floor muscle training (PFMT) for women
- Intervention: any mHealth app intervention in an educational or training context for patients or a targeted population that enables improvement in knowledge, self-efficacy, and adherence. All types of mHealth apps (eg, synchronous [connected with the health care providers] or asynchronous [stand-alone]) will be included.
- Comparison: usual care
- Outcome: PFMT adherence and usability
- Type of study design: all types of randomized controlled trials, cross-sectional studies, case–control studies, and feasibility studies
- Publication type: published primary or secondary data in a peer-reviewed journal
- Language: English and Malay
- Publication year: only articles published between 2010 and 2020. A preliminary search conducted before this review found that most of the articles regarding PFMT mHealth apps were published after 2010.

Exclusion criteria

- Population: mixed users (men and women) of mHealth apps for urinary incontinence education or other pelvic disorders without PFMT information
- Intervention: reminder system via email, SMS text message, or WhatsApp
- Type of study design: study protocol and validation study
- Publication type: reviews, systematic reviews, conference proceedings, abstract only, book chapter reviews, letters, and editorials
- Language: other than English and Malay

Data Extraction

Both reviewers extracted data from eligible articles and entered them into a Microsoft Excel spreadsheet, including data on study characteristics (year of publication, design, sample size, and duration of follow-up), participant characteristics (sex, age, and type of illness), the intervention (persuasive strategies, behavior change theories, features of the technology, and biofeedback), and results (outcomes of PFMT and adherence). Both reviewers compared and discussed the extracted results to achieve a consensus. The third reviewer (SMS) was called to discuss in the event of any unresolved discrepancies.

Data Synthesis

We conducted a narrative synthesis to answer the first 2 research questions, which aimed to list available apps for PFMT among women and their effectiveness. A full meta-analysis was not feasible owing to the heterogeneity of the studies and insufficient studies with an experimental design. A thematic synthesis was conducted to answer the third research question, which was to determine the persuasive design strategies used by the apps. The findings were then mapped to the COM-B model.

Quality Appraisal

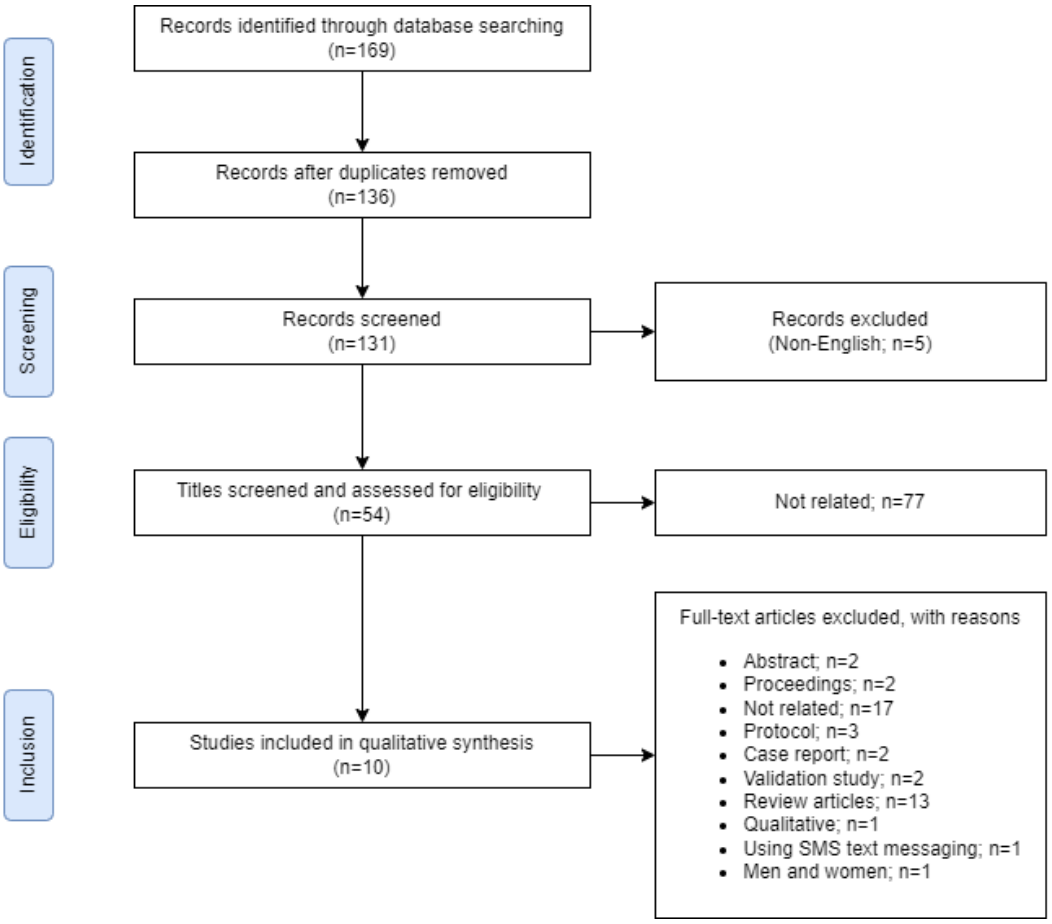
The quality of eligible randomized controlled trials was assessed by AJ and CET using the Cochrane risk of bias tool, concentrating on sequence generation, allocation concealment, blinding, attrition bias, completeness of outcome data, and other sources of bias [32]. For nonrandomized studies, the Risk of Bias Assessment Tool for Non-randomized Studies was used, which has demonstrated its validity to assess the bias specific to observational studies [33]. Any conflicts were resolved through discussion with a third person.

Results

Search Results and Study Characteristics

From the 169 records screened for titles and abstracts, 54 (32%) full texts were reviewed, and 10 (5.9%) articles met our selection criteria and were included in the review (Figure 1) [34-43]. These 10 articles were published between 2015 and 2020. Of the 10 articles, 3 (30%) referred to the same app, Tät [35,39,42].

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



Available mHealth Apps for PFMT

There was a total of 8 mHealth apps developed for PFMT (Table 2). Of the 10 studies, 3 (30%) investigated the effectiveness of stand-alone apps [35,39,42], and 1 (10%) used an audio guidance app [43]. Another study (1/10, 10%) evaluated an app that uses Bluetooth technology to link data from biofeedback [36]. Of the 8 apps, except for 1 (13%) [40], all mHealth apps (n=7, 88%) provided asynchronous PFMT—there was no live communication with a trainer.

Table 2. Details of the primary studies identified and reviewed.

PFMT[a] mHealth[b] app	Country	Year launched	Platform	Study design	Study participants	Participant characteristics	Outcomes
Squeezy App [41]	United Kingdom	2013	iOS	Cross-sectional survey	464—38% pregnant women; mean age unavailable	Mixture of bladder problem and healthy men and women	80% carried out their PFMT at least 3 days a week.
Tät [35]	Sweden	2013	iOS and Android	Randomized controlled trial	123 nonpregnant women; mean age (intervention group): 44.8 (SD 9.7) years; mean age (control group): 44.7 (SD 9.1) years	Stress urinary incontinence	41% (25/61) performed PFMT daily
Tät [39]	Sweden	2013	iOS and Android	Randomized controlled trial (2-year follow-up)	123 nonpregnant women; mean age: 44.2 (SD 10.3) years	Stress urinary incontinence	Use of incontinence protection decreased significantly (P=.04); 31/46 (67.4%; P<.001) could contract their pelvic muscles correctly
Tät [42]	Sweden	2013	iOS and Android	Randomized controlled trial (cost-utility study)	123 nonpregnant women; mean age (intervention group): 44.8 (SD 9.7) years; mean age (control group): 44.7 (SD 9.1) years	Stress urinary incontinence	The extra cost per quality-adjusted life year for the app group ranged from −€2425.70 (US $2718.60) to €14,870.60 (US $16,666.20)
Diário Saúde [34]	Brazil	2016	N/A[c]	Randomized controlled trial	31 nonpregnant women; mean age (intervention group): 47.2 (SD 10.6) years; mean age (control group): 53.3 (SD 13.2) years	Stress urinary incontinence	Adherence was higher in the app group at 1 and 2 months after PFMT (P<.001)
Penyikang app [40] (synchronous app)	China	2017	iOS and Android	Cross-sectional study	1982 postpartum women aged >18 years	Pelvic floor muscle weakness	483 postpartum women had a relatively high degree of participation (15 times)
iBall app [36] (an external device connected via Bluetooth)	Canada	2017	N/A	Pilot randomized controlled feasibility study	23 postpartum women; mean age (intervention group): 31 (SD 2.7) years; mean age (control group): 34 (SD 2.2) years	Not specified	There was no statistically significant difference between the groups for change scores
MyHealtheBladder app [37]	United States	2017	N/A	Pilot single-group, quasi-experimental study	29 nonpregnant women; mean age: 54.4 (SD 10.4) years	Urinary incontinence	97% adherence rate to the daily sessions
Bwom app [38]	United States	2017	N/A	Cross-sectional survey	47 patients and 22 providers (pregnancy status not available); age 20-50 years	Not specified	No adherence outcome

PFMT[a] mHealth[b] app	Country	Year launched	Platform	Study design	Study participants	Participant characteristics	Outcomes
Pen Yi Kang [43] (audio guidance app)	China	2018	iOS and Android	2-arm parallel randomized controlled clinical trial	108 primipara women; mean age (intervention group): 29.2 (SD 2.6) years; mean age (control group): 29.1 (SD 2.9) years	Stress urinary incontinence	Greater self-efficacy with a mean difference of 8.9 points at 6 months after delivery

[a]PFMT: pelvic floor muscle training.

[b]mHealth: mobile health.

[c]N/A: not applicable.

Study Quality Assessment Results

For randomized controlled trials, the detailed quality assessment results are summarized in Figure S1, Multimedia Appendix 1 [34-43]. In general, all trials (4/4, 100%) included in this review showed an acceptable risk of bias. The randomization sequence was adequately generated in 50% (2/4) of the trials, and all trials (4/4, 100%) adequately concealed the allocation of the participants. Blinding of the participants and personnel was not possible in these trials. Therefore, the risk of performance bias in all studies was classified as low. All articles reported an intention-to-treat analysis except for Araujo et al [34], who conducted a per-protocol analysis.

For nonrandomized studies, the detailed results of the quality assessment are summarized in Figure S2, Multimedia Appendix 1. In general, none of the 4 studies reported a strategy for minimizing selection bias and controlling methods for the confounders. More than half of the studies (3/4, 75%) had low response rates (<50%), and only 25% (1/4) of the studies reported the point estimates of the results.

Persuasive Strategies and COM-B Model in the PFMT Apps

Only the iBall app (1/8, 13%) used all 4 categories of PSD (Table 3) [36].

Table 3. Details of the primary studies identified with their persuasive system design (PSD) categories.

PFMT[a] mHealth[b] app	PSD category			
	Primary task	Dialogue support	Credibility support	Social support
Squeezy App [41]	Tailoring, personalization, and self-monitoring	Reminders	Expertise	N/A[c]
Tät [35]	Tailoring, tunneling, and self-monitoring	Reminders	Expertise	N/A
iBall app (an external device connected via Bluetooth) [35]	Tailoring, personalization, and self-monitoring	Rewards	Expertise	Competition
Bwom app [38]	Tailoring, personalization, and self-monitoring	Social role	Trustworthiness, expertise, surface credibility, and third-party endorsements	N/A
Diário Saúde[d] [34]	Tailoring, personalization, and self-monitoring	Reminders	Expertise	N/A
Penyikang app [40]	Tailoring	N/A	Expertise	N/A
MyHealtheBladder app [37]	Tailoring, personalization, and self-monitoring	Rewards and reminders	Expertise	N/A
Pen Yi Kang [43]	Tailoring, personalization, and self-monitoring	Reminders	Expertise	N/A

[a]PFMT: pelvic floor muscle training.

[b]mHealth: mobile health.

[c]N/A: not applicable.

[d]Operant conditioning.

Of the 8 mHealth apps to improve PFMT adherence, only the Diário Saúde app (n=1, 13%) [34] was developed using operant conditioning. Operant conditioning occurs by providing reinforcement to a certain behavior immediately when the behavior is performed [44]. This helps increase the likelihood that the behavior will be performed again by the user. In the intervention involving the Diário Saúde app, the study participants underwent surface electromyography (sEMG) and

PFM examination by a single physiotherapist during PFMT. All the study participants were given immediate feedback on whether they had performed PFMT correctly and were able to observe the sEMG feedback for themselves. Their sEMG graph was then displayed in the app to assist their PFMT at home [34]. This represented an incentive or reinforcement for correctly performing PFMT with the aid of the app.

All the apps (8/8, 100%) incorporated *Tailoring* (under the *Primary Task* category) and *Expertise* (under the *System Credibility* category) as part of their PSD strategies. *Tailoring* is defined as "any of a number of methods for creating communications individualized for their receivers, with the expectation that this individualization will lead to larger intended effects of these communications" [45]. All the apps (8/8, 100%) tailored the intervention by matching the content to the user groups. For *Expertise*, the apps portrayed the credibility and competence of their content by highlighting the expertise of educators in the field of PFMT, such as physiotherapists or physicians.

Although the Bwom app [38] incorporated the most persuasive strategies from the *System Credibility* category, the study did not report the effectiveness of the app. Instead, it was a cross-sectional study aimed at surveying the usefulness of the app for PFMT. The assessment showed that both the understandability and actionability of the app were >90%. These high scores may support the value of the app design among women with incontinence.

All the apps (8/8, 100%) had persuasive strategies that could be mapped to the *Opportunity* and *Motivation* domains of the COM-B model to influence and improve the behavior, which is PFMT (Table 4). Most apps (7/8, 88%) also allowed users to perform self-monitoring. This could be mapped to the *Capability* domain of the COM-B model.

Table 4. Details of the primary studies identified with the Capability, Opportunity, Motivation–Behavior (COM-B) model domains.

PFMT[a] mHealth[b] app	COM-B model					
	Capability		Opportunity		Motivation	
	Psychological	Physical	Physical	Social	Reflexive	Automatic
Squeezy App [41]	Information and links	Visual aid	Overall design of the app (look and feel)	Not available	Snooze	Reminders
Tät [35]	Information on the pelvic floor, stress UI[c], and lifestyle	PFMT skill training	Ability to use the app	Not available	Statistic function and goals	Reminders
iBall app [36] (an external device connected via Bluetooth)	Not available	PFMT skills via gamification	N/A[d]	Gamification—ranking score (web community for original version)	Not available	Rewards
Bwom app [38]	Educational videos on PFMT	Provides list of PFMT exercise plans	Ability to use the app (PEMAT[e])	Culturally relevant	Not available	Not available
Diário Saúde [34]	Not available	The visual component of sEMG[f]	Ability to use the app	Not available	Statistic function and goals	Reminders
Penyikang app [40]	Facilitate PFD[g] information	Self-management	Participation using the app	Ability to consult with the physicians (synchronous app)	Not available	N/A
MyHealtheBladder app [37] (web-based information)	A story - based behavioral program	PFMT strategies	Ability to use the app	N/A	A story-based behavioral program	Rewards and reminders
Pen Yi Kang [43]	N/A	PFMT guidance—audio	Ability to use the app	N/A	N/A	Reminders

[a]PFMT: pelvic floor muscle training.

[b]mHealth: mobile health.

[c]UI: urinary incontinence.

[d]N/A: not applicable.

[e]PEMAT: Patient Education Materials Assessment Tool.

[f]sEMG: surface electromyography.

[g]PFD: pelvic floor dysfunction.

The Effect of the Selected Persuasive Strategies

Diário Saúde, Tät, and Pen Yi Kang [34,35,43] demonstrated improved PFMT adherence. For Tät, 41% of the participants performed PFMT daily, and 42.6% performed PFMT weekly after 12 weeks of using the app [35]. For Pen Yi Kang, self-efficacy for adherence to PFMT was measured as its outcome. The app resulted in a mean difference of 8.9 points for self-efficacy at 6 months after delivery when compared with the control group (conventional home-based PFMT) [43]. This suggests that the audio guidance was beneficial for improving users' self-efficacy and maintaining PFMT adherence. Hence, the incorporation of persuasive strategies in mHealth apps resulted in an improvement in PFMT adherence.

PFMT adherence among users of the MyHealtheBladder [37] and Squeezy apps [41] was good. However, the adherence to these apps was determined using cross-sectional or quasi-experimental study designs. Therefore, the effectiveness of the apps on adherence needs to be interpreted with caution as there was no control group for an objective comparison.

Meanwhile, the iBall app was unable to demonstrate improved adherence owing to technical difficulties and discomfort faced by study participants because of the external device [36]. Although the iBall app had various sophisticated persuasive features such as an interesting user interface, Bluetooth connection with an external device, and gamification, these features were not sufficient to overcome the problems with user acceptance of the external device. Thus, the potential effectiveness of the app itself in supporting adherence was hampered by the low acceptability of the external device.

Discussion

Persuasive Design in mHealth Apps

The search found 8 eligible PFMT mHealth apps with mixed findings. Diário de Saúde, which used operant conditioning, showed a significant improvement in adherence rate [34]. *Expertise* and *Authority* were the most commonly chosen persuasive strategies in mHealth apps for PFMT.

Persuasive strategies are essential for attracting the target users' attention or interest in the app. The background of health care professionals and researchers who recommend the app and its advantages to target users is important in attracting their interest. This approach was explained in detail by Squeezy App as follows: "...has been curated by a specialist women's health physiotherapist, peer-reviewed and endorsed by the National Health Service (NHS)" [41]. This statement was important to convince users that the app was a genuine product developed from trusted sources and compliant with the Data Protection Act [41].

Recommendation by a specialist or expert provides a sense of trust for potential users regarding the importance and safety of the app. This motivates users to continue using the app, improve their adherence, and gradually form habits for the desired target behavior. Squeezy App was designed to be used with or without physiotherapist guidance. Therefore, the credibility of the app is important as users may not have the opportunity to meet the physiotherapist in person for PFMT. Evaluation of the app by physiotherapists and their endorsement of the app increased users' confidence in it [41].

Operant conditioning, which is an important concept in behavior change theories, was used by the Diário de Saúde app [34]. Displaying the biofeedback chart through the app provided users with feedback that reinforced their PFMT exercises [34]. This illustrates a successful use of behavior change theories to improve adherence to PFMT. The Tät app has been recently proven effective in reducing incontinence for urge UI and mixed UI after a 15-week intervention [46] and a 12-week intervention for UI self-management based on a prospective cohort study [47]. However, most studies did not include pregnant women. More studies are needed to determine the effectiveness of mHealth solutions for pregnant women who experience urinary incontinence. A future study has been planned to determine a mobile app's effectiveness in reducing incontinence among pregnant women [48].

Applying the COM-B Model

The COM-B model can be applied to address barriers to PFMT and identify potential behavior change techniques or persuasive strategies to be incorporated into mHealth app design. For example, barriers to PFMT may include being unaware of the need to perform PFMT, inability to understand how to perform PFMT, and lack of self-efficacy [4,7,12,14]. Performing PFMT can be challenging as the PFMs are not visible [13]. Thus, health care providers need to explain clearly and guide pregnant women to visualize the contraction of the correct muscles [7]. This represents the creation of psychological capability in the COM-B model. Therefore, an app that provides training with gradually increasing levels of difficulty would be helpful in improving the psychological capability of users. The reminder function in the app also assists users in remembering when to perform the behavior. The ability to self-monitor their performance, including the provision of rewards for achieving continuous adherence to PFMT, is a persuasive strategy that could provide additional support for the psychological capability component of the COM-B model.

Other barriers to PFMT include limited access to PFMT information and lack of time [4,13]. Providing PFMT information through mHealth apps that can be accessed at one's own convenience is a form of physical opportunity in the COM-B model. Social barriers to PFMT may include a lack of positive role models for performing PFMT, embarrassment, and the social norms among friends and colleagues that hinder a person from performing PFMT [4,13]. Creating a community of app users (eg, an app that includes a Kegel exercise support group) may provide the social opportunity to keep the users motivated to perform PFMT.

Motivating a person to change their behavior is challenging. Reflective motivation based on the COM-B model may be helpful. This may be in the form of guiding app users to evaluate and plan for specific behavior, such as setting the time for PFMT. Consciously reflecting on personal barriers to performing the target behavior and solving them can help motivate a person. This could be incorporated into apps through guided reflections.

This review was limited to mHealth for delivering PFMT to women. There may be other PSD and behavior change theories used in other PFMT apps designed for men. The focus of this review was only on the COM-B model. Other widely used theories, such as the Health Belief Model and Social Cognitive Theory were not discussed in this review.

Conclusions

There have been 8 mHealth apps designed for PFMT among women worldwide in the past decade. Physical capability can be improved by using PFMT skill training and system credibility (expertise) as persuasive strategies. Physical opportunity for PFMT is supported by the app's usability (personalization), whereas the users' motivation can be improved using goals and reminders. The persuasive strategies in these apps were mainly mapped to the *Capability* and *Motivation* components of the COM-B model. Tailoring and self-monitoring were the most commonly used persuasive strategies in the design of PFMT apps.

Acknowledgments

This research was funded by Universiti Putra Malaysia (UPM/800–3/3/1/GPB/2018/9668500).

Authors' Contributions

The concept of this study was designed by AJ. Data were collected and analyzed by AJ and CET. Data interpretation and manuscript preparation were undertaken by AJ, CET, NA, and SMS. Several drafts were reviewed and commented on by AJ, CET, NA, SMS, and FGS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality assessment of the included studies (randomized and nonrandomized controlled trials).
[PDF File (Adobe PDF File), 247 KB - mhealth_v10i3e28751_app1.pdf]

References

1. Sangsawang B, Sangsawang N. Stress urinary incontinence in pregnant women: a review of prevalence, pathophysiology, and treatment. Int Urogynecol J 2013 Jun;24(6):901-912 [FREE Full text] [doi: 10.1007/s00192-013-2061-7] [Medline: 23436035]
2. Moossdorff-Steinhauser HF, Berghmans BC, Spaanderman ME, Bols EM. Prevalence, incidence and bothersomeness of urinary incontinence in pregnancy: a systematic review and meta-analysis. Int Urogynecol J 2021 Jul 22;32(7):1633-1652 [FREE Full text] [doi: 10.1007/s00192-020-04636-3] [Medline: 33439277]
3. Arrue Gabilondo M, Ginto L, Zubikarai M, Galán C, Saro J, Diez-Itza I. Risk factors associated with stress urinary incontinence 12 years after first delivery. Int Urogynecol J 2021 Jan;32(11):3061-3067. [doi: 10.1007/s00192-020-04661-2]
4. Salmon VE, Hay-Smith EJ, Jarvie R, Dean S, Terry R, Frawley H, et al. Implementing pelvic floor muscle training in women's childbearing years: a critical interpretive synthesis of individual, professional, and service issues. Neurourol Urodyn 2020 Feb;39(2):863-870 [FREE Full text] [doi: 10.1002/nau.24256] [Medline: 31845393]
5. Pintos-Díaz MZ, Alonso-Blanco C, Parás-Bravo P, Fernández-de-Las-Peñas C, Paz-Zulueta M, Fradejas-Sastre V, et al. Living with urinary incontinence: potential risks of women's health? A qualitative study on the perspectives of female patients seeking care for the first time in a specialized center. Int J Environ Res Public Health 2019 Oct 08;16(19):3781 [FREE Full text] [doi: 10.3390/ijerph16193781] [Medline: 31597365]
6. Jaffar A, Mohd-Sidik S, Manaf RA, Foo C, Gan Q, Saad H. Quality of life among pregnant women with urinary incontinence: a cross-sectional study in a Malaysian primary care clinic. PLoS One 2021;16(4):e0250714 [FREE Full text] [doi: 10.1371/journal.pone.0250714] [Medline: 33909678]
7. Terry R, Jarvie R, Hay-Smith J, Salmon V, Pearson M, Boddy K, et al. "Are you doing your pelvic floor?" An ethnographic exploration of the interaction between women and midwives about pelvic floor muscle exercises (PFME) during pregnancy. Midwifery 2020 Apr;83(19):102647. [doi: 10.1016/j.midw.2020.102647] [Medline: 32014618]
8. Woodley SJ, Hay-Smith EJ. Narrative review of pelvic floor muscle training for childbearing women-why, when, what, and how. Int Urogynecol J 2021 Jul 05;32(7):1977-1988. [doi: 10.1007/s00192-021-04804-z] [Medline: 33950309]
9. Woodley SJ, Lawrenson P, Boyle R, Cody JD, Mørkved S, Kernohan A, et al. Pelvic floor muscle training for preventing and treating urinary and faecal incontinence in antenatal and postnatal women. Cochrane Database Syst Rev 2020 May 06;5:CD007471 [FREE Full text] [doi: 10.1002/14651858.CD007471.pub4] [Medline: 32378735]
10. National Institute for Health and Care Excellence (NICE). Pelvic Floor Dysfunction: Prevention and Non-surgical Management [NG210]. London: National Institute for Health and Care Excellence (NICE); 2021.

11. Lu J, Zhang H, Liu L, Jin W, Gao J, Min M, et al. Meta-analysis of perinatal pelvic floor muscle training on urinary incontinence. West J Nurs Res 2021 Jun 28;43(6):597-605. [doi: 10.1177/0193945920960779] [Medline: 32988334]
12. Jaffar A, Mohd-Sidik S, Nien FC, Fu GQ, Talib NH. Urinary incontinence and its association with pelvic floor muscle exercise among pregnant women attending a primary care clinic in Selangor, Malaysia. PLoS One 2020 Jul 15;15(7):e0236140 [FREE Full text] [doi: 10.1371/journal.pone.0236140] [Medline: 32667936]
13. Grant A, Currie S. Qualitative exploration of the acceptability of a postnatal pelvic floor muscle training intervention to prevent urinary incontinence. BMC Womens Health 2020 Jan 17;20(1):9 [FREE Full text] [doi: 10.1186/s12905-019-0878-z] [Medline: 31952500]
14. Perera J, Kirthinanda DS, Wijeratne S, Wickramarachchi TK. Descriptive cross sectional study on prevalence, perceptions, predisposing factors and health seeking behaviour of women with stress urinary incontinence. BMC Womens Health 2014 Jul 02;14:78 [FREE Full text] [doi: 10.1186/1472-6874-14-78] [Medline: 24985068]
15. van den Muijsenbergh ME, Lagro-Janssen TA. Urinary incontinence in Moroccan and Turkish women: a qualitative study on impact and preferences for treatment. Br J Gen Pract 2006 Dec;56(533):945-949 [FREE Full text] [Medline: 17132383]
16. Asklund I, Samuelsson E, Hamberg K, Umefjord G, Sjöström M. User experience of an app-based treatment for stress urinary incontinence: qualitative interview study. J Med Internet Res 2019 Mar 14;21(3):e11296 [FREE Full text] [doi: 10.2196/11296] [Medline: 30869644]
17. Barton AJ. The regulation of mobile health applications. BMC Med 2012 May 08;10(1):46 [FREE Full text] [doi: 10.1186/1741-7015-10-46] [Medline: 22569114]
18. Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A research on the classification and applicability of the mobile health applications. J Pers Med 2020 Feb 27;10(1):11 [FREE Full text] [doi: 10.3390/jpm10010011] [Medline: 32120849]
19. Harjumaa M, Oinas-Kukkonen H. Persuasion Theories and IT Design. Berlin, Heidelberg: Springer; 2007:311-314.
20. Fogg BJ. Persuasive Technology: Using Computers to Change What We ThinkDo. San Francisco, CA: Morgan Kaufmann; 2003.
21. Oinas-Kukkonen H, Harjumaa M. A Systematic Framework for Designing and Evaluating Persuasive Systems. Berlin, Heidelberg: Springer; 2008:164-176.
22. Shao X, Oinas-Kukkonen H. Thinking About Persuasive Technology from the Strategic Business Perspective: A Call for Research on Cost-Based Competitive Advantage. Switzerland: Springer; 2018:3-15.
23. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91 [FREE Full text] [doi: 10.1177/1460458216650979] [Medline: 27245673]
24. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. Implement Sci 2009 Jul 16;4(1):40 [FREE Full text] [doi: 10.1186/1748-5908-4-40] [Medline: 19607700]
25. Li T, Chen X, Wang J, Chen L, Cai W. Mobile app-based intervention for pregnant women with stress urinary incontinence: protocol for a hybrid effectiveness-implementation trial. JMIR Res Protoc 2021 Mar 10;10(3):e22771 [FREE Full text] [doi: 10.2196/22771] [Medline: 33688842]
26. Ammerman AS, Lindquist CH, Lohr KN, Hersey J. The efficacy of behavioral interventions to modify dietary fat and fruit and vegetable intake: a review of the evidence. Prev Med 2002 Jul;35(1):25-41. [doi: 10.1006/pmed.2002.1028] [Medline: 12079438]
27. Richardson M, Khouja CL, Sutcliffe K, Thomas J. Using the theoretical domains framework and the behavioural change wheel in an overarching synthesis of systematic reviews. BMJ Open 2019 Jun 22;9(6):e024950 [FREE Full text] [doi: 10.1136/bmjopen-2018-024950] [Medline: 31229999]
28. McClurg D, Frawley H, Hay-Smith J, Dean S, Chen S, Chiarelli P, et al. Scoping review of adherence promotion theories in pelvic floor muscle training - 2011 ICS state-of-the-science seminar research paper i of iv. Neurourol Urodyn 2015 Sep 21;34(7):606-614. [doi: 10.1002/nau.22769] [Medline: 25997975]
29. 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]
30. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
31. Chapman D. Health-related databases. J Can Acad Child Adolesc Psychiatry 2009 May;18(2):148-149 [FREE Full text] [Medline: 19495437]
32. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: John Wiley & Sons; 2019.
33. Kim SY, Park JE, Lee YJ, Seo H, Sheen S, Hahn S, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. J Clin Epidemiol 2013 Apr;66(4):408-414. [doi: 10.1016/j.jclinepi.2012.09.016] [Medline: 23337781]

34. Araujo CC, Marques AD, Juliato CR. The adherence of home pelvic floor muscles training using a mobile device application for women with urinary incontinence: a randomized controlled trial. Female Pelvic Med Reconstr Surg 2020 Nov 8;26(11):697-703. [doi: 10.1097/SPV.0000000000000670] [Medline: 30624250]
35. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Jun 08;36(5):1369-1376 [FREE Full text] [doi: 10.1002/nau.23116] [Medline: 27611958]
36. Dufour S, Fedorkow D, Kun J, Deng SX, Fang Q. Exploring the impact of a mobile health solution for postpartum pelvic floor muscle training: pilot randomized controlled feasibility study. JMIR Mhealth Uhealth 2019 Jul 11;7(7):e12587 [FREE Full text] [doi: 10.2196/12587] [Medline: 31298221]
37. Goode PS, Markland AD, Echt KV, Slay L, Barnacastle S, Hale G, et al. A mobile telehealth program for behavioral treatment of urinary incontinence in women veterans: Development and pilot evaluation of MyHealtheBladder. Neurourol Urodyn 2020 Jan 27;39(1):432-439. [doi: 10.1002/nau.24226] [Medline: 31774200]
38. Han MN, Grisales T, Sridhar A. Evaluation of a mobile application for pelvic floor exercises. Telemed J E Health 2019 Feb;25(2):160-164. [doi: 10.1089/tmj.2017.0316] [Medline: 29792574]
39. Hoffman V, Söderström L, Samuelsson E. Self-management of stress urinary incontinence via a mobile app: two-year follow-up of a randomized controlled trial. Acta Obstet Gynecol Scand 2017 Oct;96(10):1180-1187 [FREE Full text] [doi: 10.1111/aogs.13192] [Medline: 28718223]
40. Li J, Sun X, Wang C, Zhang Z, Xie Z. A mobile application penyikang applied in postpartum pelvic floor dysfunction: a cross-sectional study to analyze the factors influencing postpartum pelvic floor muscle strength and women's participation in treatment. Biomed Res Int 2020;2020:4218371 [FREE Full text] [doi: 10.1155/2020/4218371] [Medline: 32775419]
41. Robson M. The Squeezy pelvic floor muscle exercise app: user satisfaction survey. J Pelvic Obstet Gynaecol Physiother 2017;121:64-68 [FREE Full text]
42. Sjöström M, Lindholm L, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a cost-effectiveness analysis. J Med Internet Res 2017 May 08;19(5):e154 [FREE Full text] [doi: 10.2196/jmir.7383] [Medline: 28483745]
43. Wang X, Xu X, Luo J, Chen Z, Feng S. Effect of app-based audio guidance pelvic floor muscle training on treatment of stress urinary incontinence in primiparas: a randomized controlled trial. Int J Nurs Stud 2020 Apr;104:103527. [doi: 10.1016/j.ijnurstu.2020.103527] [Medline: 32058140]
44. Bunzli S, Gillham D, Esterman A. Physiotherapy-provided operant conditioning in the management of low back pain disability: a systematic review. Physiother Res Int 2011 Mar;16(1):4-19. [doi: 10.1002/pri.465] [Medline: 20310071]
45. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
46. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A mobile app for self-management of urgency and mixed urinary incontinence in women: randomized controlled trial. J Med Internet Res 2021 Apr 05;23(4):e19439 [FREE Full text] [doi: 10.2196/19439] [Medline: 33818395]
47. Rygh P, Asklund I, Samuelsson E. Real-world effectiveness of app-based treatment for urinary incontinence: a cohort study. BMJ Open 2021 Jan 04;11(1):e040819 [FREE Full text] [doi: 10.1136/bmjopen-2020-040819] [Medline: 33397664]
48. Jaffar A, Mohd Sidik S, Foo CN, Muhammad NA, Abdul Manaf R, Fadhilah Ismail SI, et al. Protocol of a single-blind two-arm (Waitlist Control) parallel-group randomised controlled pilot feasibility study for mHealth app among incontinent pregnant women. Int J Environ Res Public Health 2021 Apr 30;18(9):4792 [FREE Full text] [doi: 10.3390/ijerph18094792] [Medline: 33946203]

Abbreviations

COM-B: Capability, Opportunity, and Motivation–Behavior
mHealth: mobile health
PFM: pelvic floor muscle
PFMT: pelvic floor muscle training
PSD: persuasive system design
PT: persuasive technology
sEMG: surface electromyography
UI: urinary incontinence

Edited by L Buis; submitted 13.03.21; peer-reviewed by S Dufour, H Mehdizadeh, GE Iyawa, J Povey, C Khouja; comments to author 20.04.21; revised version received 03.06.21; accepted 02.02.22; published 22.03.22.

Please cite as:
Jaffar A, Tan CE, Mohd-Sidik S, Admodisastro N, Goodyear-Smith F
Persuasive Technology in an mHealth App Designed for Pelvic Floor Muscle Training Among Women: Systematic Review
JMIR Mhealth Uhealth 2022;10(3):e28751
URL: https://mhealth.jmir.org/2022/3/e28751
doi:10.2196/28751
PMID:35315777

Original Paper

Smartphone Apps for Vaping Cessation: Quality Assessment and Content Analysis

Sherald Sanchez[1,2], MSc; Anasua Kundu[2], MScCH; Elizabeth Limanto[3], HBSc; Peter Selby[4,5], MBBS, MHSc; Neill Bruce Baskerville[6], PhD; Michael Chaiton[1,2,7], PhD

[1]Institute of Medical Science, Termerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[2]Ontario Tobacco Research Unit, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Department of Pharmacology & Toxicology, University of Toronto, Toronto, ON, Canada
[4]Nicotine Dependence Service, Centre for Addiction and Mental Health, Toronto, ON, Canada
[5]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[6]Office of Audit and Evaluation, Canadian Institutes of Health Research, Ottawa, ON, Canada
[7]Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Sherald Sanchez, MSc
Institute of Medical Science
Termerty Faculty of Medicine
University of Toronto
1 King's College Circle
Toronto, ON, M5S 1A8
Canada
Phone: 1 416 535 8501 ext 32005
Email: Sherald.Sanchez@utoronto.ca

Abstract

Background: As the prevalence of electronic cigarette (e-cigarette) use, or vaping, continues to grow, particularly among young people, so does the need for research and interventions to address vaping.

Objective: This study examines the quality of free vaping cessation apps, their contents and features, popularity among users, and adherence to evidence-based principles.

Methods: A systematic search of existing apps for vaping cessation was conducted in December 2020. Eligible apps were free, in English, and included features specifically targeting vaping cessation. Each app included in the analysis was used daily for at least seven consecutive days, assessed using the Mobile App Rating Scale, and rated by at least two authors (AK, EL, or SS) based on adherence to evidence-based practices. Intraclass correlation coefficient (ICC) estimates were computed to assess interrater reliability (excellent agreement; ICC 0.92; 95% CI 0.78-0.98).

Results: A total of 8 apps were included in the quality assessment and content analysis: 3 were developed specifically for vaping cessation and 5 focused on smoking cessation while also claiming to address vaping cessation. The mean of app quality total scores was 3.66 out of 5. Existing vaping cessation apps employ similar approaches to smoking cessation apps. However, they are very low in number and have limited features developed specifically for vaping cessation.

Conclusions: Given the lack of vaping cessation interventions at a time when they are urgently needed, smartphone apps are potentially valuable tools. Therefore, it is recommended that these apps apply evidence-based practices and undergo rigorous evaluations that can assess their quality, contents and features, and popularity among users. Through this process, we can improve our understanding of how apps can be effective in helping users quit vaping.

(JMIR Mhealth Uhealth 2022;10(3):e31309) doi:10.2196/31309

KEYWORDS

e-cigarettes; vaping; cessation; mHealth interventions

Introduction

In 2020, the US Surgeon General's *Report on Smoking Cessation* identified the development of interventions to address vaping, or electronic cigarette (e-cigarette) use, particularly among young people, as a research priority [1]. There is "substantial evidence" [2] that vaping can lead to nicotine dependence and growing evidence that dependence symptoms are increasing among young people [3-8]. Several studies have shown that many youth are making attempts to quit, which are often unsuccessful due to the difficulty of quitting unassisted [5-9]. Help-seeking youth need access to tools and services that can aid them in their quitting journey. However, the availability of vaping cessation interventions remains limited in huge part because our understanding of the process of vaping cessation is extremely limited and evidence-based guidelines and interventions have yet to be developed.

Mobile software apps are highly accessible and cost-effective platforms for interventions that can be customized by the user and provide real-time support. The use of apps to provide support for addressing substance use, such as alcohol and tobacco, is well-documented in the literature [10-16]. Among young people who are the top smartphone users [17] globally and a high-risk group for the public health consequences of e-cigarettes [2,18], smartphone apps are a potential way to offer support and promote successful quit attempts among help-seeking e-cigarette users, or vapers. Informal searches revealed a handful of apps for vaping cessation currently available. However, little is known about their quality and the degree to which their contents reflect the current state of evidence [9]. Accordingly, this study examined the quality of free vaping cessation apps, their contents and features, popularity among users, and adherence to evidence-based practices. To our knowledge, this study was the first systematic analysis of free apps for vaping cessation. We focused on apps that were in English and free to download. We also decided to only include apps that were available in both the Apple App Store and Google Play Store at the time of the review. There are 2 key reasons for this. First, at the time, a study in the *Journal of Medical Internet Research* found that 87% of vaping-related apps in the Google Play Store were aimed at assisting users to sustain their vaping and improve their experience of vaping, including informational apps with "recipes" for vapers who want to mix their own vape liquids, retail apps for ordering vaping paraphernalia online, or navigation apps for locating vape stores nearby. The same study found that only 3% had features to help users quit [19]. By contrast, Apple implemented a ban on apps promoting recreational vaping from its online market in response to a recent outbreak of vaping-related lung injuries [20]. Second, compared with Google, Apple has a more rigorous and transparent review process for apps, which is described in a designated developer web page. We decided to use these resources to our advantage in conducting this review.

Methods

Search Strategy

Using the terms *vaping*, *vaping cessation*, *quit vaping*, *stop vaping*, and *no vaping*, we searched for smartphone apps targeting vaping cessation on the Canadian Apple and Google online stores in December 2020. Preliminary searches on the Apple App Store website and the embedded App Store app in iPhones showed drastically different sets of results. Whereas the phone search yielded more than 185 apps, the website search using the same strategy showed between 6 and 10 apps. For consistency with the average user experience, we decided to conduct the formal searches using the phone app instead of the website.

Eligibility Criteria

Apps that were not in English and not free to download were excluded in the preliminary screening. As previously noted, apps that were not available in both the Google Play Store and the Apple App Store were excluded. Based on informal searches conducted prior to the review (AK, EL, and SS), we expected our search terms to yield a large number of apps for smoking cessation. Our approach to these apps was to check the descriptions and profiles of popular smoking cessation apps to confirm whether they also addressed vaping cessation. Those that did not were excluded at this stage. Three authors (AK, EL, and SS) independently reviewed the remaining eligible apps (n=39) through a 2-pronged screening approach: first, we examined their profiles on the online app stores; then, if the information was available, we did an ancestry search of developer websites, profiles, and overall online presence. From this, those that were confirmed to have no association with vaping cessation were excluded at this stage. A total of 9 apps were downloaded as part of the final sample.

Assessment of Quality, Contents and Features, and Popularity Among Users

Each app was used daily for at least seven consecutive days, which informed the assessment of quality, analysis of app contents and features, and classification of apps. To estimate popularity among users, number of downloads was used as an indicator. Given that information on app downloads is only available for the apps in the Google Play Store, user ratings and number of reviews were used as a secondary measure of popularity.

The quality of each app was assessed using the Mobile App Rating Scale (MARS), a multidimensional measure for rating the quality of mobile health apps [21]. The MARS consists of 5 categories—engagement, functionality, aesthetics, information quality, and subjective quality—with 23 collective items rated using a 5-point scale. An option of "N/A" or "not applicable" is available for items that cannot be adequately assessed. Subjective quality is an optional category; thus, it was excluded from our scoring. Mean scores for each of the 4 objective categories were calculated to identify strengths and weaknesses of each app, while the sum of all 4 scores provided an overall quality score for each app. Given the variability found among health-related apps, MARS itself does not prescribe a defined

threshold for the type of scores a high-quality app should obtain on the scale [21]. However, a previous study [15] that assessed the quality and content of smoking cessation apps identified a MARS total mean score of 3 as the cut-off for apps with acceptable quality.

Based on our analysis of app contents and features, each app was classified according to its primary approach to supporting vaping cessation. We followed Abroms et al's [11,12] categories for smoking cessation apps, which included "calculators" tracking money saved and health benefits gained since quitting; "calendars" monitoring number of days before or after the quit date; and "informational content" providing general information on quitting. Informational content on each app was cross-referenced with the current state of evidence on the health consequences of vaping [2].

Adherence to Evidence-Based Practices

To assess clinical quality, previous studies of a similar design evaluated apps based on their level of adherence to clinical practice guidelines [11-15]. In this study, we developed a 14-item Adherence Index based on a modified version of the *Canadian Smoking Cessation Clinical Practice Guideline* [22] with the support of an experienced clinician who led the development of the guideline on smoking cessation (PS). These 14 items are shown in Table 1. The items are written in plain language. Most of the content remained the same, with 1 key change—because there is currently no strong evidence supporting use of nicotine replacement therapy for vaping cessation, endorsement of medication was omitted. The process of modifying the guidelines was further informed by earlier studies in identifying key differences between smoking and vaping based on the experiences of help-seeking youth and young adult vapers [8,23]. Each item on the Adherence Index was coded *present* or *not present*.

Table 1. Characteristics of vaping cessation apps and summary of quality scores.

Characteristics	Kwit	Quit Vaping - For Good	Quit Vaping Addiction Calendar	Escape the Vape	Quit Genius	Smoke Watchers	Aeris	Smokler
App classification	Tracker	Tracker	Tracker	Tracker	Tracker	Other	Tracker	Other
Popularity								
Ratings on Apple App Store[a]	4.5 (1100)	4 (4)	0 (0)	3.4 (5)	4.3 (472)	0 (0)	2.3 (3)	0 (0)
Ratings on Google Play Store[a]	4.5 (3000)	3.7 (248)	0 (0)	3.8 (6)	4.3 (2000)	3.7 (39)	3.9 (84)	4.7 (45)
Number of Google Play Store downloads	100,000+	10,000+	100+	100+	100,000+	5000+	5000+	1000+
MARS[b] scores								
Engagement	4.1	3.8	3.4	3.53	3.1	2.9	2.6	2.3
Functionality	4.9	4.67	4.75	4.67	4.12	3.89	4.5	4.88
Aesthetics	4.83	4	4.5	4	3.83	4.17	3.67	3.5
Information	3.14	3.1	2.86	3.1	3.21	2.64	2.36	2
MARS total mean scores	4.24	3.89	3.88	3.82	3.57	3.4	3.28	3.17
Adherence to Canadian clinical guidelines								
Ask about e-cigarette[c] use status	✓	✓	✓	✓	✓	✓	✓	
Advise user to quit	✓	✓	✓	✓	✓	✓	✓	✓
Assess willingness to quit	✓	✓	✓		✓		✓	
Assess nicotine dependence		✓			✓	✓		
Assist—discuss the benefits of quitting	✓	✓	✓	✓	✓		✓	
Assist—offer tools and resources for quitting	✓	✓	✓	✓	✓	✓	✓	✓
Assist—enhance motivation to quit	✓	✓	✓	✓	✓	✓		
Assist—explore doubts about quitting	✓							
Assist—explore barriers to quitting	✓				✓		✓	
Assist—affirm and encourage decision to quit	✓	✓	✓	✓	✓	✓	✓	
Assist—form a quit plan							✓	
Assist—discuss relapse prevention	✓		✓	✓		✓	✓	
Assist—refer to a quitline							✓	
Arrange for follow-up	✓		✓		✓			✓
Adherence Index score (0-14)	12	8	9	7	10	7	10	3

[a]Number of stars out of 5 (number of reviews).

[b]MARS: Mobile App Rating Scale.

[c]e-cigarette: electronic cigarette.

Data Extraction and Quality Assessment

Prior to assessing the full sample of apps, 3 raters (AK, EL, and SS) tested the MARS and Adherence Index with a randomly selected app. Results from the test app showed substantial agreement with minor differences (≤2 points) that were discussed and resolved. The remaining apps were rated independently by at least two authors. Intraclass correlation coefficient (ICC) estimates were computed to assess interrater reliability for the remaining apps. Results based on a mean rating (k=3), absolute agreement, and 2-way mixed effects model showed excellent agreement (ICC 0.92; 95% CI 0.78-0.98) [24].

Results

Overview

A total of 9 apps for vaping cessation were identified. Once the apps were downloaded, we discovered that 1 app (*Quuit*) was faulty and could not be examined past the sign-up page. Multiple attempts to reach the developers to resolve the issue were unsuccessful. At this point, we decided to exclude *Quuit*. The app selection process is shown in Figure 1. Figure 2 provides a visual overview of the remaining 8 apps included in the analysis. Characteristics and quality scores of each app are presented in Table 1.

Figure 1. Flow diagram of the app selection process.

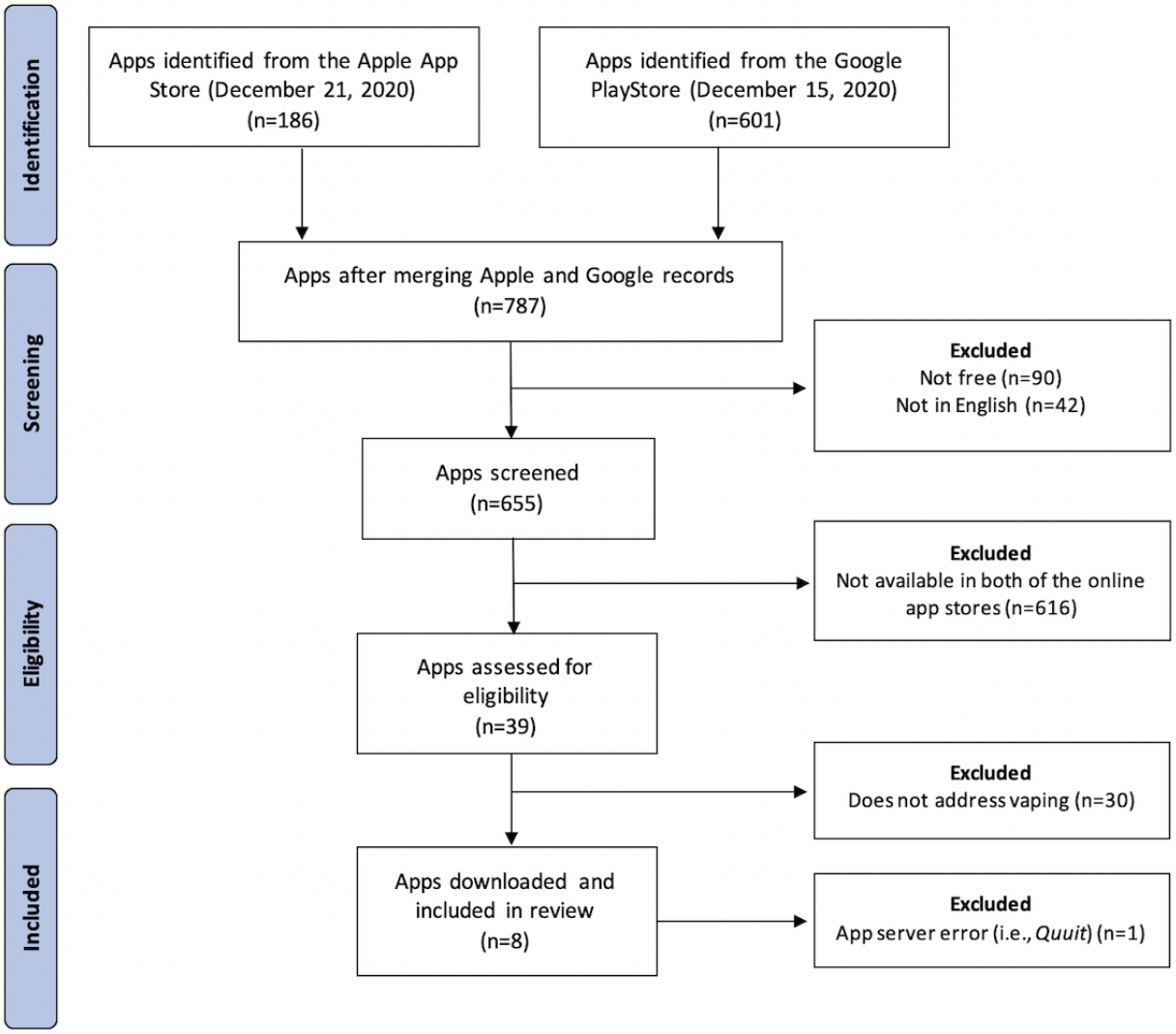
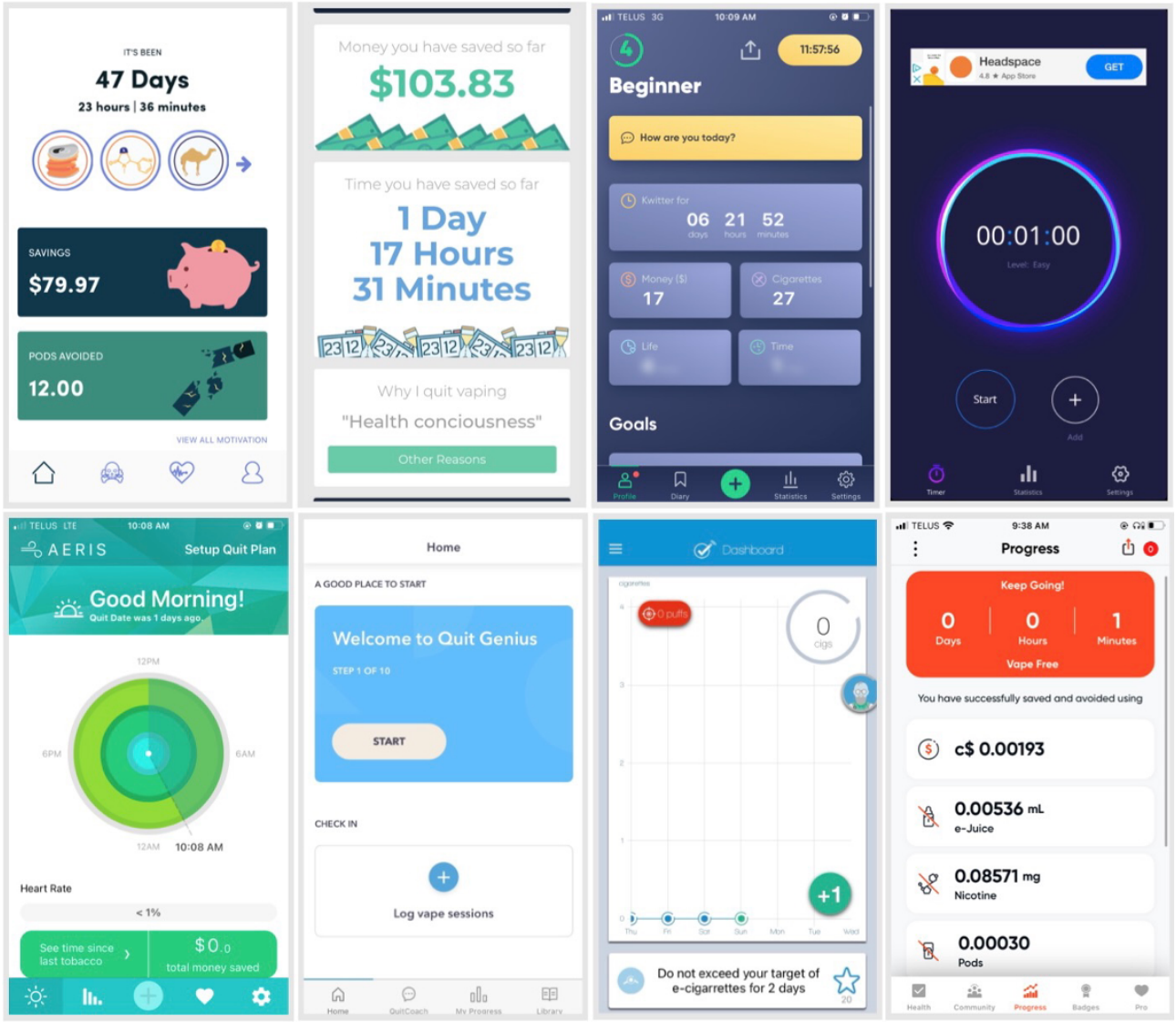


Figure 2. Visual overview of the 8 apps included in the analysis.



Assessment of Quality, Contents and Features, and Popularity Among Users

Among the 8 apps, the overall mean of the MARS total mean scores in Table 1 was 3.66, indicating *acceptable* quality. The mean rating for each of the 4 objective categories among the included apps is as follows: functionality was scored highest (4.54), followed by aesthetics (4.06) and engagement (3.22), while information had the lowest mean score (2.8). Cross-referencing the informational content on the apps with the current state of evidence on the health consequences of vaping [2], we found that some, but not all, informational text was supported by scientific evidence. A few other apps shared links to resources, which often pointed users to news headlines or anecdotal stories.

Of the 8 apps, 3 specifically targeted vaping cessation (*Escape the Vape*, *Quit Vaping Addiction Calendar*, and *Quit Vaping – For Good*) and 5 were focused on smoking cessation while also claiming to address vaping cessation. Although the remaining 5 apps claimed to address vaping, they were primarily developed for smoking cessation and many features were not fully adapted to vaping cessation. This is evident when users are asked to provide information on the number of cigarettes smoked and cost of a pack of cigarettes during the sign-up process. Even among the 3 apps developed primarily for vaping cessation, it was common to see the terms “vaping” and “smoking” used interchangeably. None of the apps asked users to enter information on basic demographics or vape flavor preferences. Some apps included valuable features such as a designated quit plan page that we later learned were available only with paid subscriptions ranging from US $5.99 to US $27.99. In total, 6 apps offered in-app purchases.

Of the included apps, 3 had a built-in community forum where users can anonymously share their experiences; 2 of these had features that allowed users to interact with others’ posts. However, 1 (*Smoke Watchers*) was entirely in French. A fourth app (*Kwit*) offered users the option to join a private support group on Facebook with a membership of 1800. Another app (*Quit Genius*) offered a community feature and access to a quit coach to premium users for US $27.99 per month.

Trackers, characterized as a combination of calculators and calendar apps, were the most common type of vaping cessation apps. As many as 6 of the 8 apps applied a tracker approach to

vaping cessation, documenting money saved over time and number of days since quitting. A total of 4 tracker apps also provided a measure of health gains since quitting, while 2 presented estimations of "nicotine avoided" either in milligrams, number of pods, or both. These valuations, except for health gains, were based on information provided by users during the sign-up process. In addition to money spent on vaping and frequency of vaping, *Quit Vaping – For Good* asked users to indicate the type of vape device used, nicotine concentration preferred, number of pods in a pack, and amount of e-liquid content for each pod. It also tracked "total e-juice avoided." All tracker apps, except *Aeris*, conferred awards corresponding to milestones achieved based on money saved and time vape free, as well as nicotine avoided and community engagement for those apps that had those features. Three of the tracker apps also kept a record of frequency and intensity of cravings that users can enter spontaneously within the app. Of these 3 apps, 2 presented users with a daily check-in: *Quit Genius* asked "Did you vape?" while *Kwit* asked users to provide a scale rating of their emotions and confidence in quitting.

Compared with the tracker apps, *Smoke Watchers* and *Smokler* implemented unique approaches. As the name indicates, *Smoke Watchers* connected vapers with watchers who may be vapers or nonvapers interested in supporting a vaper in their quitting process. Users can invite friends to be their watcher or select 1 from a community of watchers provided by the app. *Smoke Watchers* required vapers to connect Bluetooth-enabled vape devices directly to the app for real-time monitoring without an option for a user to manually enter their own data. Unfortunately, as none of the authors owned a vape device, this was not a feature that was fully explored. By contrast, *Smokler*'s primary approach to vaping cessation was implementing a vaping schedule. Users start a timer, then vape when the timer ends. There are varying levels of difficulty based on how long the timer is set for, and the interval between each vaping occasion increases according to the user's preference.

All apps were developed commercially and most of them typically focused on abstinence with only 2 apps (*Smoke Watchers* and *Smokler*) offering users the option to modify their goals toward reduction or rationing. In terms of substances, nicotine was the primary focus in all 8 apps. *Quit Genius* was the only app linked to an efficacy trial published in the scientific literature [25], which concluded it to be a "superior treatment" compared with "very brief advice." Notably, the authors disclosed that the study was funded by the company that developed *Quit Genius* and 8 of the 10 authors received a salary from or owned equity in the company.

Lastly, popularity among users was estimated using relative frequency of downloads, user ratings, and number of reviews. As shown in Table 1, the 3 most frequently downloaded apps (*Kwit*, *Quit Genius*, and *Quit Vaping – For Good*) were also those that were most highly rated and had the greatest number of user reviews, which suggests strong engagement among users. However, this association was not consistently present among the 5 remaining apps. To illustrate, *Aeris* (3.9 stars) had more user downloads and reviews compared with *Smokler*. However, *Smokler* (4.7 stars) had a higher user rating. In one of our searches using the search term *quit vaping*, *Aeris* was among the top 20 results in the Apple App Store, but *Smokler* was not. Interestingly, there were several apps with no reviews or user ratings that were included in the top 20 results. Overall, we found that a higher user rating was not associated with the order an app shows up in the list of results—a finding that is consistent with previous studies of a similar design [11,12].

Adherence to Evidence-Based Practices

As shown in Table 1, the most commonly represented items on the Adherence Index were advising users to quit and assisting users by offering tools and resources for quitting. By contrast, the least common items were assisting users by exploring doubts about quitting, forming a quit plan, and referring them to a quitline. Each of these were only identified once among the apps included in the analysis. The *Kwit* app had the highest MARS total score, while also having the most Adherence Index items present. It also had the highest user rating based on data from the Apple App Store and the Google Play Store. Meanwhile, *Smokler* had the lowest MARS total score and also had the fewest Adherence Index items present. *Aeris* was the only app that recommended calling a quitline and the only one with a designated feature for setting up a quit plan.

Discussion

Principal Findings

This study identified 8 apps available for vaping cessation. Generally, these existing vaping cessation apps appear to employ similar approaches to smoking cessation apps that have been previously examined [11-15]. However, relative to smoking, apps that aid in quitting vaping are very low in number and are limited in features developed specifically for vaping cessation.

Although we found that the highest- and lowest-rated apps on the MARS tool received parallel scores on the Adherence Index, there was no association between the scores for the remaining apps. Similarly, there was no association between quality scores and app popularity based on user ratings and number of downloads. Previous studies with a comparable study design reported mixed results in this area: for example, Abroms and colleagues [12] concluded that user ratings were positively associated with Adherence Index scores, whereas Ubhi and team [13] found no such association. Consistent with findings in previous studies, we found that (1) the most common strategy employed to promote vaping cessation was tracking and monitoring [11,12,15]; (2) few apps included the behavior change techniques proposed to aid in quitting [13,14]; (3) the apps had little evidence-based informational content [13,14]; and (4) referrals to a quitline were absent [11-15]. As with these previous studies, we also found that the apps in our sample scored better in design and usability components over clinically relevant criteria. MARS items on quality of information and credibility of sources consistently received the lowest marks. These findings suggest that the development team may have been more focused on making the apps user-friendly than incorporating evidence-based content.

Overall, we found that using an app quality assessment tool and Adherence Index was a feasible approach to evaluating vaping cessation apps. Nonetheless, we discourage readers from

interpreting these results to mean that some apps are more *effective* than others in promoting vaping cessation. We learned from this process that some apps are inherently different from others and it is not always appropriate, possible, or desirable to incorporate all of the items on the Adherence Index in any one app. The same issue arises in assessing apps based on the prevalence of behavior change techniques that have been positively associated with higher success rates for quitting smoking in *face-to-face* behavioral interventions [26]. This challenge was one that we encountered repeatedly in discussions about how items on the clinical guideline might be translated within the context of an app. For example, in this study, arranging for follow-up was translated to receiving push notifications from an app. However, several factors will influence the effectiveness of push notifications. Did the user enable notifications? Are there any consequences if a user ignores push notifications? What is the conversion rate between push notifications and interaction with the app? Do users experience desensitization to push notifications over time, thereby diminishing their usefulness? Similar questions were asked for other items on the guidelines and what they might look like in an app, although these were more straightforward compared with the example described here. Perhaps for this reason, an increasing number of studies are exploring the use of SMS text messaging or email services in promoting user engagement within an app [13,14] or as its own form of intervention [27]. Relative to vaping cessation, the best example of this is Truth Initiative's *This is Quitting* program—a free and anonymous SMS text messaging service designed to help young people quit vaping, with preliminary results showing high levels of engagement among the target age group [7,28].

A reviewer of an earlier draft of this paper suggested that a qualitative analysis of user comments could shed more light on what features or characteristics users might find helpful and why. This is a potential way to identify features that future apps could incorporate to help support vaping cessation among users. The same reviewer inquired after possible approaches for addressing concurrent nicotine and cannabis use (or co-use) among users in light of emerging evidence of high rates of co-use and its implications for vaping cessation [29-32]. At the time of our review, we found no evidence of approaches for addressing co-use in our final sample of apps for vaping cessation. Certainly, apps for quitting cannabis use are available in both the Apple App Store and the Google Play Store and we encountered several of them during our searches. However, the degree to which these apps adhere to evidence-based practices warrants investigation. A review conducted in 2015 [33] that focused on the top 20 cannabis apps available online found that "no apps addressed abuse, addiction, or treatment." As in ours, the 2015 review also showed that the majority of freely available apps were geared toward supporting recreational use.

Another possibility worth exploring is an "ecosystem of apps" to promote vaping cessation where users are free to choose the app (or set of apps) most appropriate for them based on their level of motivation, the stage of change they find themselves in, or simply personal preference. Ideally, this would be curated and maintained based on a similar process employed in this study and conducted by a group external to the app development team.

Limitations

This study has several limitations. The evaluation of apps developed for health behavior change is a relatively new practice, and more guidance and resources are needed [34,35]. We acknowledge that app development occurs at a rapid pace and new apps are added regularly that were not available when we conducted our searches, and thus, would not be included here. Paid features offered as in-app purchases within apps that were free to download were also beyond the scope of this review. However, considering that free apps and features have been shown to increase accessibility and user engagement [36], our approach can be justified. Although our searches were conducted in the Canadian online app stores, it is likely that there is significant overlap in the online app market relative to the availability of vaping cessation apps. Finally, we were not able to assess the potential impact of these apps for different cultures and ethnicities.

Strengths

This study also has noteworthy strengths. It is the first systematic analysis of apps for vaping cessation. As previous studies of a similar design have proven to be useful in areas such as alcohol and tobacco use [10-16], we recognize the value in assessing the quality and content of apps for vaping or e-cigarette use in informing research and practice moving forward. Based on our findings, we proposed several recommendations for the improvement of existing and future apps for supporting vaping cessation relative to the current state of evidence. Notably, we propose that researchers pay more attention to the unique qualities of vaping relative to smoking [8,23] and how these can be appropriately incorporated into apps focused on vaping cessation. Industry actors, such as the Apple App Store and the Google Play Store, might also have a role to play in improving accessibility to evidence-informed apps. For example, app stores could provide more transparency with regard to the algorithm behind their search results. One step further to this would be to explore the possibility of including clinical quality as a factor in the algorithm, in addition to the number of downloads and user ratings. The Apple App Store's 2020 ban on recreational vaping apps [19] could be considered as a precedence for this kind of initiative. As vaping research evolves, there are plenty of opportunities to explore these recommendations and their potential for helping improve our understanding of how technology-based interventions, such as apps and SMS text messaging services, can be effective in helping users quit vaping. Finally, a key component to these recommendations is forthcoming research from 2 of the authors in this paper (AK and PS) offering clinical guidance for vaping cessation based on a systematic review. Beyond vaping research, this study also contributes to important and ongoing discussions over the role of mobile health apps in our health system, particularly in the field of substance use and addictions [37-40].

Conclusions

Given the lack of vaping cessation interventions at a time when they are urgently needed, vaping cessation apps are potentially

valuable tools. Therefore, it is recommended that these apps apply evidence-based practices and undergo rigorous evaluations that can assess their quality, contents and features, and popularity among users. Through this process, we can improve our understanding of how apps can be effective in helping users quit vaping.

Acknowledgments

This study was funded by the Substance Use and Addictions Program of Health Canada (Agreement Number 1920-HQ-000073).

Conflicts of Interest

None declared.

References

1. Smoking Cessation: A Report of the Surgeon General. Rockville, MD: US Department of Health and Human Services; 2020.
2. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.
3. Hammond D, Reid JL, Rynard VL, O'Connor RJ, Goniewicz ML, Piper ME, et al. Indicators of dependence and efforts to quit vaping and smoking among youth in Canada, England and the USA. Tob Control 2021 Feb 01:tobaccocontrol-2020-056269. [doi: 10.1136/tobaccocontrol-2020-056269] [Medline: 33526441]
4. Cuccia AF, Patel M, Amato MS, Stephens DK, Yoon SN, Vallone DM. Quitting e-cigarettes: Quit attempts and quit intentions among youth and young adults. Prev Med Rep 2021 Mar;21:101287 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101287] [Medline: 33489721]
5. Amato MS, Bottcher MM, Cha S, Jacobs MA, Pearson JL, Graham AL. “It’s really addictive and I’m trapped:” A qualitative analysis of the reasons for quitting vaping among treatment-seeking young people. Addictive Behaviors 2021 Jan;112:106599. [doi: 10.1016/j.addbeh.2020.106599]
6. Graham AL, Jacobs MA, Amato MS, Cha S, Bottcher MM, Papandonatos GD. Effectiveness of a Quit Vaping Text Message Program in Promoting Abstinence Among Young Adult E-Cigarette Users: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 May 01;9(5):e18327 [FREE Full text] [doi: 10.2196/18327] [Medline: 32356774]
7. Graham A, Jacobs M, Amato M. Engagement and 3-Month Outcomes From a Digital E-Cigarette Cessation Program in a Cohort of 27 000 Teens and Young Adults. Nicotine Tob Res 2020 Apr 21;22(5):859-860 [FREE Full text] [doi: 10.1093/ntr/ntz097] [Medline: 31197320]
8. Sanchez S, Kaufman P, Pelletier H, Baskerville B, Feng P, O'Connor S, et al. Is vaping cessation like smoking cessation? A qualitative study exploring the responses of youth and young adults who vape e-cigarettes. Addict Behav 2021 Feb;113:106687. [doi: 10.1016/j.addbeh.2020.106687] [Medline: 33045643]
9. Berg CJ, Krishnan N, Graham AL, Abroms LC. A synthesis of the literature to inform vaping cessation interventions for young adults. Addict Behav 2021 Aug;119:106898. [doi: 10.1016/j.addbeh.2021.106898] [Medline: 33894483]
10. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone Apps Targeting Alcohol and Illicit Substance Use: Systematic Search in in Commercial App Stores and Critical Content Analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
11. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
12. 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]
13. Ubhi HK, Michie S, Kotz D, van Schayck OCP, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
14. Ubhi HK, Kotz D, Michie S, van Schayck OCP, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
15. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
16. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
17. Wilson K. Mobile cell phone technology puts the future of health care in our hands. CMAJ 2018 Apr 03;190(13):E378-E379 [FREE Full text] [doi: 10.1503/cmaj.180269] [Medline: 29615420]

18. Kwan LY, Eaton DL, Andersen SL, Dow-Edwards D, Levin ED, Talpos J, et al. This is your teen brain on drugs: In search of biological factors unique to dependence toxicity in adolescence. Neurotoxicol Teratol 2020;81:106916. [doi: 10.1016/j.ntt.2020.106916] [Medline: 32698050]
19. Meacham MC, Vogel EA, Thrul J. Vaping-Related Mobile Apps Available in the Google Play Store After the Apple Ban: Content Review. J Med Internet Res 2020 Nov 13;22(11):e20009 [FREE Full text] [doi: 10.2196/20009] [Medline: 33185565]
20. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
21. 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]
22. Canadian Smoking Cessation Clinical Practice Guidelines. In: CAN-ADAPTT. Toronto, Canada: Centre for Addiction and Mental Health; 2011.
23. Sanchez S, Chaiton M. Is vaping dependence like smoking dependence? —A literature review highlighting implications for vaping cessation among youth. J Public Health Emerg 2021 Sep;5:27. [doi: 10.21037/jphe-21-48]
24. 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]
25. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary Outcomes of a Digital Therapeutic Intervention for Smoking Cessation in Adult Smokers: Randomized Controlled Trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
26. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
27. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214 [FREE Full text] [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
28. Graham AL, Amato MS, Cha S, Jacobs MA, Bottcher MM, Papandonatos GD. Effectiveness of a Vaping Cessation Text Message Program Among Young Adult e-Cigarette Users: A Randomized Clinical Trial. JAMA Intern Med 2021 Jul 01;181(7):923-930 [FREE Full text] [doi: 10.1001/jamainternmed.2021.1793] [Medline: 33999133]
29. Hindocha C, McClure EA. Unknown population-level harms of cannabis and tobacco co-use: if you don't measure it, you can't manage it. Addiction 2020 Nov 20;116(7):1622-1630. [doi: 10.1111/add.15290]
30. Smith DM, Miller C, O'Connor RJ, Kozlowski LT, Wadsworth E, Fix BV, et al. Modes of delivery in concurrent nicotine and cannabis use ("co-use") among youth: Findings from the International Tobacco Control (ITC) Survey. Substance Abuse 2020 Jan 17;42(3):339-347. [doi: 10.1080/08897077.2019.1709603]
31. Driezen P, Gravely S, Wadsworth E, Smith D, Loewen R, Hammond D, et al. Increasing Cannabis Use Is Associated With Poorer Cigarette Smoking Cessation Outcomes: Findings From the ITC Four Country Smoking and Vaping Surveys, 2016-2018. Nicotine Tob Res 2022 Jan 01;24(1):53-59. [doi: 10.1093/ntr/ntab122] [Medline: 34111281]
32. Fataar F, Hammond D. The Prevalence of Vaping and Smoking as Modes of Delivery for Nicotine and Cannabis among Youth in Canada, England and the United States. Int J Environ Res Public Health 2019 Oct 25;16(21):4111 [FREE Full text] [doi: 10.3390/ijerph16214111] [Medline: 31731420]
33. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis Mobile Apps: A Content Analysis. JMIR Mhealth Uhealth 2015 Aug 12;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]
34. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces. J Med Internet Res 2019 May 02;21(5):e12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]
35. Larson RS. A Path to Better-Quality mHealth Apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
36. Mobile App Marketing Insights: How Consumers Really Find and Use Your Apps Internet. Google. 2015. URL: https://think.storage.googleapis.com/docs/mobile-app-marketing-insights.pdf [accessed 2021-06-01]
37. Khadjesari Z, Brown T, Naughton F. Regulation and accreditation of addictive behaviour applications-navigating the landscape. Addiction 2021 Dec 26;116(12):3276-3283. [doi: 10.1111/add.15484] [Medline: 33739480]
38. Bricker JB, McAfee T. Smartphone applications for intervening on addictive behaviors: efficacy is just one phase of the marathon. Addiction 2021 Dec 28;116(12):3284-3285. [doi: 10.1111/add.15608] [Medline: 34184335]
39. Perski O. Scientific and ethical challenges to defining what constitutes 'proportionate evidence' for the regulation and accreditation of applications to treat addiction. Addiction 2021 Dec 07;116(12):3285-3287. [doi: 10.1111/add.15619] [Medline: 34231264]
40. Nesvåg S, McKay JR. Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]

Abbreviations

e-cigarette: electronic cigarette
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale

Edited by L Buis; submitted 16.06.21; peer-reviewed by N Krishnan, L Thornton, S Cha, K Yokotani; comments to author 02.08.21; revised version received 08.10.21; accepted 07.01.22; published 28.03.22.

Please cite as:
Sanchez S, Kundu A, Limanto E, Selby P, Baskerville NB, Chaiton M
Smartphone Apps for Vaping Cessation: Quality Assessment and Content Analysis
JMIR Mhealth Uhealth 2022;10(3):e31309
URL: https://mhealth.jmir.org/2022/3/e31309
doi:10.2196/31309
PMID:35343904

Original Paper

Quality of Mobile Apps for Care Partners of People With Alzheimer Disease and Related Dementias: Mobile App Rating Scale Evaluation

Nicole E Werner[1], PhD; Janetta C Brown[2], PhD; Priya Loganathar[1], MSc; Richard J Holden[3], PhD

[1]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States

[2]Indiana University School of Medicine, Indianapolis, IN, United States

[3]Department of Health & Wellness Design, School of Public Health-Bloomington, Indiana University, Bloomington, IN, United States

Corresponding Author:
Nicole E Werner, PhD
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
1513 University Avenue
Madison, WI, 53706
United States
Phone: 1 608 890 2578
Email: nwerner3@wisc.edu

Abstract

Background: Over 11 million care partners in the United States who provide care to people living with Alzheimer disease and related dementias (ADRD) cite persistent and pervasive unmet needs related to their caregiving role. The proliferation of mobile apps for care partners has the potential to meet care partners' needs, but the quality of apps is unknown.

Objective: This study aims to evaluate the quality of publicly available apps for care partners of people living with ADRD and identify design features of low- and high-quality apps to guide future research and user-centered app development.

Methods: We searched the US Apple App and Google Play stores with the criteria that included apps needed to be available in the US Google Play or Apple App stores, accessible to users out of the box, and primarily intended for use by an informal (family or friend) care partner of a person living with ADRD. We classified and tabulated app functionalities. The included apps were then evaluated using the Mobile App Rating Scale (MARS) using 23 items across 5 dimensions: engagement, functionality, aesthetics, information, and subjective quality. We computed descriptive statistics for each rating. To identify recommendations for future research and app development, we categorized rater comments on score-driving factors for each MARS rating item and what the app could have done to improve the item score.

Results: We evaluated 17 apps. We found that, on average, apps are of minimally acceptable quality. Functionalities supported by apps included education (12/17, 71%), interactive training (3/17, 18%), documentation (3/17, 18%), tracking symptoms (2/17, 12%), care partner community (3/17, 18%), interaction with clinical experts (1/17, 6%), care coordination (2/17, 12%), and activities for the person living with ADRD (2/17, 12%). Of the 17 apps, 8 (47%) had only 1 feature, 6 (35%) had 2 features, and 3 (18%) had 3 features. The MARS quality mean score across apps was 3.08 (SD 0.83) on the 5-point rating scale (1=inadequate to 5=excellent), with apps scoring highest on average on functionality (mean 3.37, SD 0.99) and aesthetics (mean 3.24, SD 0.92) and lowest on average on information (mean 2.95, SD 0.95) and engagement (mean 2.76, SD 0.89). The MARS subjective quality mean score across apps was 2.26 (SD 1.02).

Conclusions: We identified apps whose mean scores were more than 1 point below minimally acceptable quality, whereas some were more than 1 point above. Many apps had broken features and were rated as below acceptable for engagement and information. Minimally acceptable quality is likely to be insufficient to meet care partner needs. Future research should establish minimum quality standards across dimensions for care partner mobile apps. Design features of high-quality apps identified in this study can provide the foundation for benchmarking these standards.

(JMIR Mhealth Uhealth 2022;10(3):e33863) doi:10.2196/33863

KEYWORDS

Alzheimer disease and related dementias; mobile app; mHealth; caregivers; dementia caregiving; eHealth; telehealth; mobile phone

Introduction

Background

Over 11 million care partners in the United States who provide care to people living with Alzheimer disease and related dementias (ADRD) are often untrained, underresourced, and unsupported to manage the cognitive, behavioral, and physical changes that characterize ADRD progression [1-3]. Therefore, care partners cite persistent and pervasive unmet needs related to all aspects of their caregiving role, including support for daily care, managing behavioral symptoms of dementia, self-care, resources and support services, health information management, care coordination and communication, and financial and legal planning [4-6]. The ability to address the unmet needs of care partners is a critical health challenge, as these unmet needs are associated with suboptimal psychological and physical outcomes for the care partner and the person living with ADRD [7-10].

National experts call for technologies to be powerful and novel interventions to support care partners [11]. For example, experts from the 2015 Alzheimer Disease Research Summit recommended to "develop new technologies that enhance the delivery of clinical care, care partner support, and in-home monitoring" and "test the use of technology to overcome the workforce limitations in the care of older adults with dementia as well as providing care partner support and education" [11]. The 2018 Research Summit called for "innovative digital data collection platforms" and "pervasive computing assessment methods" [12].

Mobile apps can answer these calls by enabling unique data capture and visualization, multichannel communication, and integration of powerful decision support on increasingly ubiquitous and scalable devices (eg, smartphones). Advancing technological capabilities also increase the potential of mobile apps to provide much-needed individualized, just-in-time support that can adapt to changing needs across the course of the disease [13]. Reviews of mobile apps for care partners report that they are a feasible and acceptable intervention [14] and can reduce ADRD care partner stress and burden [15].

The mere availability of apps is not sufficient to improve health outcomes; these apps must be designed to support and accommodate user needs and abilities, a process called user-centered design (UCD). More formally, UCD is:

> *an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, usability knowledge, and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability, and counteracts possible adverse effects of use on human health, safety and performance.* [16]

UCD provides a scientifically sound, practice-based mechanism for developing mobile apps for care partners of people living with ADRD that are highly feasible and more likely to improve care partner outcomes [17]. Conversely, if apps are not designed using UCD, they are more likely to be of low quality, cause more harm than good, incur avoidable waste of financial and human resources, not provide the needed support, and compound the existing burden on care partners [13,16,18-21].

Despite the potential of mobile apps to meet care partners' needs and improve outcomes using UCD and other industry-standard design practices, the actual quality of mobile apps for care partners—that is, how usable, engaging, valid, acceptable, accessible, aesthetically pleasing, and useful they are to the user—is currently unknown. A recent study by Choi et al [22] used the Mobile App Rating Scale (MARS) to assess app quality across ADRD-related apps focused on self-care management for people living with ADRD. They found that, on average, the evaluated apps met the MARS criteria for minimally acceptable quality, quality scores were higher for those developed by health care–related versus non–health care–related developers, and apps scored lower on average regarding how engaging they were to the user [22]. Although this study included some apps with care partners as the intended primary user, the inclusion and exclusion criteria focused on the person living with ADRD, which limited the inclusion of apps targeted at care partners as the intended end user.

It is critical to evaluate the quality of mobile apps for ADRD care partners for several reasons [17]. First, quality assessment ensures that mobile apps produce benefits and do not have unintended health consequences for care partners or persons living with ADRD; for example, they do not increase care partner stress and burden. Second, quality evaluation can provide insights into whether mobile apps will be used and whether use will withstand the test of time; that is, they will not be abandoned. Third, quality evaluation is important to ensure that research-based mobile apps are sustainable outside academic research settings, meaning they can achieve commercial success among competitors. Fourth, quality evaluation can safeguard against commercial products that may not deliver on their advertised potential.

Objectives

Thus, the aim of this study is to (1) evaluate the quality of publicly available apps for care partners of people living with ADRD and (2) identify the design features of low- and high-quality apps to guide future research and user-centered app development.

Methods

Design

We conducted a multirater evaluation of the quality of mobile apps for caregivers of people living with ADRD available on the US market by applying the MARS [23]. The MARS was

created to be an easy-to-use and objective tool for researchers and developers to evaluate the quality of mobile apps across multiple dimensions. We chose to use the MARS because it is a validated rating scale for mobile app quality, includes a multicomponent evaluation of quality, has clear instructions and a uniform scale, and has been used successfully across multiple health domains, including pain management and ADRD [22,24,25].

Data Collection

App Identification and Selection

We searched the US Apple App and Google Play stores in March 2021 using multiple variations of the terms "caregiver," "carer," "care," "caretaker," "dementia," and "Alzheimer disease." To be included in the analysis, an app needed to be (1) available in US Google Play or Apple App stores; (2) directly accessible to users out of the box (ie, without a separate agreement with an insurer, health care delivery organization, and enrolling in a clinical trial); and (3) primarily intended for use by an informal (family or friend) care partner or care partners of a person with dementia of any severity, stage, or etiology. Four members of the research team independently searched both app stores to identify eligible apps based on the app name and brief description and identified 50 unique apps. Next, 3 members of the research team applied the inclusion criteria to the compiled list of apps by reviewing the full app description and downloading and exploring the app components. One research team member served as the arbiter by reviewing each app for inclusion and documenting the reason for inclusion or exclusion. The arbiter presented their inclusion decisions to the full research team for a consensus. Primary reasons for app exclusion were not having the caregiver as the primary user (eg, apps for the person living with ADRD), needing to sign up for a clinical trial or be part of a specific health system to access the app, and not being specific to ADRD care (eg, targeted for caregivers of people with any condition). We identified 17 unique apps that met our inclusion criteria, 8 (47%) of which were available in both the iOS and Android versions. For apps that were available on both iOS and Android, we randomly selected whether we would evaluate the iOS or Android version. An expert rater also reviewed the version that was not selected to assess quality differences, and no quality differences were identified between platforms for any of the apps.

App Classification

For each included app, we captured descriptive and technical information, such as name, ratings, version history, language, and functionality. We classified the app's purpose and functionality based on the app store description and available functionality within the app.

MARS Evaluation

The MARS includes 23 items across 5 dimensions: engagement, functionality, aesthetics, information, and subjective quality [23]. Each item was scored on a 5-point scale, from inadequate (score=1) to excellent (score=5) or not applicable.

Our MARS evaluation team included 7 research team members: 3 experts in UCD and ADRD caregiving and 4 trainees in these areas. The MARS training process began with the full team independently reviewing the published MARS guide, including instructions, definitions, and rating scales. Next, we conducted 3 team-based training sessions to improve consensus on the MARS ratings. During the training sessions, we evaluated each app as a team, item by item, with a discussion of each item rating to build consensus on how to interpret the items and the criteria for each score within an item. During the team rating sessions, we discussed score anchors and annotated the MARS rating sheet based on consensus anchors. Between team training sessions, team members practiced applying the ratings discussed in the previous session and created additional annotations based on the team consensus discussion, which were then shared with the full team at the subsequent meeting.

Next, each app was rated using the MARS by at least 2 independent raters. To apply the MARS, each trained rater downloaded the app to a testing phone, paid fees, and tested the app to ensure that all components of the app were used. The rater then completed the 23 MARS rating items in order, app by app. In addition to the required MARS rating procedures, raters also documented for each item the score-driving factors for that item and what the app could have done to improve the score. This was done to support our aim of guiding future research and app development.

Two members of the research team reviewed all the scores and identified the items for which the original 2 raters had disagreements in their scores. For items with a disagreement score, an expert rater (JCB, RJH, or NEW) was used as the tiebreaker. The goal of the expert rater as a tiebreaker was to determine which of the scores they agreed with are based on the MARS training and their expertise in evaluating health information technologies. However, if the expert rater disagreed with both scores, the tie-breaking score could be different from the original 2 raters with clear justification. Expert raters were senior members of the research team with doctoral training in UCD, a combined 6 years designing and evaluating dementia caregiving technologies, and a combined 13 years evaluating health information technologies. We calculated the percentage agreement for each rating dyad and the overall agreement rate.

Data Analysis

The ratings were entered into a cloud-based Microsoft Excel spreadsheet, and descriptive statistics were computed for each rating. First, we computed the mean score for each of the quality dimensions (engagement, functionality, aesthetics, information, and subjective quality) for each individual app as the sum of the item scores in each dimension divided by the items in the dimension. Next, we calculated the app quality mean score for each app as the sum of the dimension mean scores divided by the number of dimensions. We calculated the total mean score for each dimension across all apps as the sum of each app's dimension mean score divided by the total number of apps. We computed the overall app subjective mean score as the sum of the mean scores divided by the total number of apps.

To identify recommendations for future research and app development, 2 expert members of the research team categorized rater comments on the score-driving factors for each item and

what the app could have done to improve the score for that item. They then met to discuss the categories and reach a consensus.

Ethics Approval

This study did not involve human subjects.

Results

App Classification

We evaluated 17 apps (n=7, 41%, iOS only; n=2, 12%, Android only; and n=8, 47%, both iOS and Android). Before the expert-based score reconciliation process, across 6 rating dyads, the raters provided the exact same rating on a 1 to 5 scale in 43% of ratings; rater dyads agreed within 1 point in 83% of cases (detailed agreement and disagreement rates of each rating dyad are given in Multimedia Appendix 1). All apps except one were available at no cost for the most basic version, and no apps required an additional cost to upgrade to advanced features or additional content. Apps had affiliations with commercial companies 47% (8/17), universities 24% (4/17), health systems 18% (3/17), governments 12% (2/17), and nongovernmental organizations 6% (1/17). Of the 17 apps evaluated, 14 (82%) were available in English only; 1 (6%) was available in English, Korean, and Spanish; 1 (6%) was available in English and Japanese; and 1 (6%) was available in English and Portuguese. Full descriptions and technical details of the apps are provided in Table 1.

We identified 8 general feature categories supported by the apps (Table 2). These categories included the ability of the app to provide the following: (1) education—the provision of relevant, appropriate content that increases care partner knowledge and self-efficacy to perform their role and make informed decisions (12/17, 71%); (2) interactive training—reciprocal exchange of information for care partner development and learning (3/17, 18%); (3) documentation—storage or recording of information for later retrieval (3/17, 18%); (4) tracking of symptoms (2/17, 12%); (5) care partner community—a platform or feature created for the exchange of social support among care partners (3/17, 18%); (6) interaction with clinical experts (1/17, 6%); (7) care coordination—the organization and distribution of patient care activities among all involved participants (2/17, 12%); and (8) activities for the person living with ADRD (2/17, 12%). Of the 17 apps, 8 (47%) had only 1 feature, 6 (35%) had 2 features, and 3 (18%) had 3 features. Most apps (12/17, 71%) provided information, and for 29% (5/17) of apps, providing information was the only feature. Of the 17 apps, some features such as interaction with clinicians or tracking symptoms were offered by only 1 (6%) or 2 (12%) apps.

Table 1. Apps evaluated in the study and their descriptive and technical details.

App name	Platform	Category	Developer	Year of last update	Country	Language	Purpose	Affiliations
Accessible Alzheimer's and Dementia Care	iOS	Health and fitness	ACHGLOBAL, Inc	N/A[a]	United States	English	Provide valuable information	Commercial
Alzheimer's and Dementia Care	Android	Health and fitness	Accessible home health care	2019	United States	English	Provide information	Commercial
Alzheimer's Daily Companion	iOS[b] Android	Lifestyle	Home Instead Senior Care	2013	United States	English	Build care partner confidence	Commercial
Alzheimer's Manager	iOS	Health and fitness	Point of Care LLC	2021	United States	English	Track and monitor	Commercial
Care4Dementia	iOS	Medical	Univ of New South Wales	N/A	Australia	English	Provide information	University
Clear Dementia Care	iOS Android	Medical	Northern Health and Social Care Trust	2021	United States	English	Provide information and support	Government Health System
CogniCare	iOS[b] Android	Health and fitness	CongniHealth Ltd	2021	United Kingdom	English, Japanese	Improve quality of life	Commercial University
Dementia Advisor	iOS[b] Android	Health and fitness	Sinai Health System	2019	Canada	English	Improve communication	Government Health System
Dementia Caregiver Solutions	iOS[c]	Health	Lorenzo Gentile	2016	Canada	English	Provide expert advice	Commercial
Dementia Guide Expert	iOS[b] Android	Education	University of Illinois	2020	United States	English, Korean, Spanish	Educate and empower	University
Dementia Stages Ability Model	iOS[b] Android	Education	Positive Approach, LLC	2020	United States	English	Help learn characteristics and care of GEMS stages	Commercial
Dementia Talk	iOS	Health and fitness	Sinai Health System- Reitman Centre	2019	United States	English	Track and monitor	Health System
DementiAssist	Android	Medical	Baylor Scott and White Health	2015	United States	English	Provide insights	University
DemKonnect	iOS[b] Android	Medical	Nightingales Medical Trust	2020	United Kingdom	English	Provide care partner connections	NGO[d]
Inspo-Alzheimer's Caregiving	iOS[b] Android	Social networking	Inspo Labs	2020	United States	English	Create safe supportive community	N/A
Remember Me-Caregiver	iOS	Productivity	Daniel Leal	2020	United States	English, Portuguese	Share care responsibility	N/A
Respite Mobile	iOS	Health and fitness	ADC initiatives LLC	2021	United States	English	Provide activities for people with ADRD[e]	Commercial

[a]N/A: not available.

[b]Indicates platform reviewed.

[c]Indicates cost of use.

[d]NGO: nongovernmental organization.

[e]ADRD: Alzheimer disease and related dementias.

Table 2. Feature categories of the evaluated apps (N=17).

Apps	Education (n=12, 71%)	Interactive training (n=3, 18%)	Documentation (n=3, 18%)	Tracking symptoms (n=2, 12%)	Care partner community (n=3, 18%)	Interaction with clinical expert (n=1, 6%)	Care coordination (n=2, 12%)	Activities for person with dementia (n=2, 12%)
Accessible Alzheimer's and Dementia Care	✓							
Alzheimer's and Dementia Care	✓				✓			
Alzheimer's Daily Companion	✓							
Alzheimer's Manager			✓	✓				
Care4Dementia	✓							
Clear Dementia Care	✓	✓	✓					
CogniCare	✓							✓
Dementia Advisor		✓						
Dementia Caregiver Solutions	✓							
Dementia Guide Expert	✓							
Dementia Stages Ability Model	✓	✓						
Dementia Talk			✓	✓			✓	
DementiAssist	✓							
DemKonnect	✓				✓	✓		
Inspo-Alzheimer's Caregiving	✓				✓			
Remember Me-Caregiver							✓	
Respite Mobile								✓

MARS Evaluation

The MARS app quality mean score across all apps was 3.08 (SD 0.83) on the 5-point rating scale (from 1=inadequate to 5=excellent), with apps scoring highest on average on functionality (mean 3.37, SD 0.99) and aesthetics (mean 3.24, SD 0.92) and lowest on average on information (mean 2.95, SD 0.95) and engagement (mean 2.76, SD 0.89; Table 3).

Table 3. Mean scores on the Mobile App Rating Scale (MARS) rating categories with category definitions and subjective evaluation data, including app store number of ratings, app store average ratings, and the MARS subjective quality score.

Apps	Quality mean score, mean	Engagement, mean	Functionality, mean	Aesthetics, mean	Information, mean	Subjective quality score, mean	App store number ratings	App store average rating (out of 5 stars)
Accessible Alzheimer's and Dementia Care	1.26	1.20	1.50	1.00	1.33	1.00	1	5
Alzheimer's and Dementia Care	2.29	2.00	2.00	2.67	2.50	1.00	NR[a]	NR
Alzheimer's Daily Companion	2.50	1.60	3.25	2.67	2.50	1.25	16	4.6
Alzheimer's Manager	2.98	2.60	3.50	3.00	2.83	2.67	1	3
Care4Dementia	3.20	2.20	3.75	3.00	3.83	2.50	1	5
Clear Dementia Care	3.85	3.80	4.25	3.33	4.00	3.25	1	5
CogniCare	3.74	3.80	4.00	4.00	3.17	2.50	NR	NR
Dementia Advisor	4.18	4.20	5.00	3.33	4.17	4.50	6	4.3
Dementia Caregiver Solutions	3.11	2.60	3.50	3.00	3.33	2.00	1	5
Dementia Guide Expert	2.66	2.40	2.75	2.33	3.17	3.25	1	5
Dementia Stages Ability Model	4.26	3.30	4.75	4.67	4.33	3.25	15	5
Dementia Talk	3.29	3.00	3.00	4.00	3.17	1.25	1	1
DementiAssist	2.93	2.80	3.25	3.33	2.33	2.00	7	4.1
DemKonnect-	2.71	3.00	2.00	3.67	2.17	1.75	NR	NR
Inspo-Alzheimer's Caregiving	4.17	4.00	4.00	4.67	4.00	3.25	NR	NR
Remember Me-Caregiver	1.88	1.50	2.50	2.33	1.17	1.00	NR	NR
Respite Mobile	3.35	3.00	4.25	4.00	2.17	2.00	10	5
Overall, mean (SD)	3.08 (0.83)	2.76 (0.89)	3.37 (0.99)	3.24 (0.92)	2.95 (0.95)	2.26 (1.02)	N/A[b]	N/A

[a]NR: not rated.

[b]N/A: not applicable.

The MARS subjective quality mean score across all apps was 2.26 (SD 1.02), with mean scores ranging from 1 to 4.5. The mean score for the question, "Would you recommend the app to people who might benefit from it?" was 2.59 (SD 1.42).

The MARS app quality mean score of 2.94 (SD 0.93) for apps with a commercial affiliation was slightly below the minimally acceptable quality and slightly above the minimally acceptable quality 3.26 (SD 0.57) for apps with noncommercial affiliations (ie, universities, governments, health systems, and nongovernmental organizations). The MARS subjective quality mean score (SD) was below the minimally acceptable quality for apps with both commercial affiliation (mean 1.96, SD 0.83) and noncommercial affiliations (mean 2.64, SD 1.11).

Table 3 provides the mean scores on the MARS quality rating dimensions and subjective evaluation data, including the MARS subjective quality score, app store number of ratings, and app store average ratings for all evaluated apps.

Score-Driving Factors

Table 4 lists the most frequently identified design qualities that led to low or high MARS scores for each MARS dimension. Among factors contributing to low scores, a common one was broken functionality, leading to crashes, error messages, and unresponsiveness, noted in 59% (10/17) of the apps. Among factors contributing to high scores, a common quality included aesthetics where adequate use of multimedia for content presentation, clear and consistent user interface layouts, and high-quality graphics were noted in 53% (9/17) of the apps.

Table 4. Mobile App Rating Scale (MARS) dimensions, categories within those dimensions, and examples of design features that were score drivers for low and high scores.

MARS dimension and categories	Examples of low-score drivers	Examples of high-score drivers
Engagement		
Entertainment	Entertaining content such as games, chat, videos, and forums that do not function; extensive and overwhelming content; and very little content	Use of multimedia (eg, combination of text, video, audio, images, and animations)
Interest	Text only with no images, large blocks of text, frequent system failure, constantly linking to outside website, and no ability to customize experience	Variety of content, features, and color throughout the app
Customization	Limited, inoperable, or missing customization features	Variety of customization options (eg, privacy settings, preference selection, notifications, and favorites)
Interactivity	Interactive content such chat, graphs, and forums does not function; must click what you need every time (the app does not retain information); and community forum but no active users	Feedback systems (eg, confirmations, error messages, and validations) and variety of data visualization with charts, in-app messaging, and features for community building
Target group	Small font, no ability to zoom, provides only general information, and no privacy settings	Content relevance and usefulness of information
Functionality		
Performance	Frequent error messages and crashes, frequently unresponsive or slow, and includes inactive hyperlinks	Responsiveness and efficient transitions throughout the app
Ease of use	Takes a lot of time to figure out how to use, functions difficult to learn to use the complicated app architecture, and no instructions provided	Clarity and intuitiveness of app functions and learnability, operability, and app instructions
Navigation	Menu options change within the app, clicking a link takes you to the incorrect function, consistently sending to an outside website with broken links, no back button provided, and the Menu options do not function	Logic, consistency, and visual cues matched users' expectations; external sources within the app; and minimalist design
Gestural design	Gestures differ from expectation in terms of phone gestures	Logical, consistent, anticipated gestures, links, and buttons
Aesthetics		
Layout	Blocks of text and inconsistent layout across pages	Clear and simple user interface layout
Graphics	Low quality (blurry) and no graphics	Quality, high-resolution graphics
Visual appeal	No color, no graphics, no multimedia, and inconsistent text sizes and colors	Creative, impactful, and thoughtful use of color
Information		
Accuracy of app description	Describes content that does not function or is not available	Features and functions aligned with the app description
Goals	Goals not stated, goals not achievable because app functions are broken or unresponsive, and no ability to measure goal attainment	Goals stated explicitly with measurable or trackable metrics
Quality of information	Sources not cited, sources cited are questionable, links provided to sources are broken, and information disorganized or difficult to locate	Information provided from trusted, cited sources; language used written with end users or target demographic in mind; and information relevant to users
Quantity of information	Extensive and overwhelming amount of information and very little information	Sufficient and comprehensive range of information
Visual information	No visual information available	Logical use of videos, multimedia, and helpful images to provide clarity
Credibility	No sources cited and commercial entity selling other products	Created by a legitimate, verified entity, including hospital, center, government, university, or council
Evidence base	Not tested for effectiveness in improving person living with ADRD[a] or care partner outcomes	N/A[b]

[a]ADRD: Alzheimer disease and related dementias.

[b]N/A: not available.

Discussion

Principal Findings

The objectives of our study were to (1) evaluate the quality of publicly available apps for care partners of people living with ADRD and (2) identify the design features of low- and high-quality apps to guide future research and app development. Our findings show that across all apps, the average MARS quality rating was just above the minimally acceptable cut-off of 3.00 (mean 3.08, SD 0.83; range 1.26-4.26), and the average MARS subjective quality rating of all the apps was less than acceptable (mean 2.26, SD 1.02; range 1.00-4.50). We also identified apps whose individual mean scores were more than 1 point below the minimal acceptable quality, whereas some were more than 1 point above. Furthermore, most of the apps we assessed had broken features and were rated as below acceptable quality for the MARS dimensions of engagement and information quality.

Of the 17 mobile apps, our analysis identified 3 (18%) with a rating of good or higher quality (MARS quality mean score >4). Furthermore, Dementia Advisor scored greater than 4 (ie, indicating good quality) on both the MARS quality mean score and the subjective quality mean score. In contrast to most apps that focus on providing education through text and videos, Dementia Advisor provides interactive training on a wide variety of scenarios with feedback to improve learning. The app was simple and intuitive, without the need for instructions or significant time to learn to use the app features. All features of the app were functional, and the progress through the training scenarios was tracked by the app.

We found that most apps focused on passively delivering educational content. Providing education is important, as care partners report persistent unmet needs related to understanding ADRD as a disease process, including diagnosis, prognosis, and disease progression; long-term care and financial and legal planning; and management of cognitive and behavioral symptoms [4,5,26,27]. However, the extent of the effectiveness of passive learning content (eg, reading an article or watching a video) provided by these apps is unknown and may be limited as opposed to engaged active learning approaches that foster information retention [28-30]. In addition, care partners also reported the need for training, support for coordination across the caregiving network, connection to relevant resources, and social support [4,5,27,31-33]. Some apps did attempt to address care partners' need for social support by offering forums, chats, and community features. However, we found that these features were often not functional or did not have active participation from users, limiting the app's ability to fulfill their promise of social support. Furthermore, the apps were limited in functionality to support coordination across the caregiving network, with only 2 apps supporting coordination with other care partners and only 1 connecting care partners with clinicians. Overall, the limited functionality provided across most apps raises questions about their potential to improve care partner outcomes, as several recent systematic reviews and meta-analyses suggested that effective care partner interventions provide multiple components and social support [34-44].

Overall, the apps scored higher on functionality and aesthetics than on engagement and information quality. The apps, on average, scored just above minimally acceptable for functionality (mean 3.37, SD 0.99), which includes app performance, ease of use, navigation, and gestural design. Functionality is important for care partners because it reflects the potential of the app to meet basic care partner needs in terms of app usability. This score is a point lower than that indicated in the MARS rating reported in a 2020 study by Choi et al [22], which used the MARS to assess the quality of all ADRD-related apps, including those focused on care partners and those focused on the person living with ADRD. It is possible that the higher scores found by Choi et al [22] reflect a higher quality of apps designed for people living with ADRD, as a recent study by Guo et al [45] on rating mobile apps for people living with ADRD reported a similarly high functionality score.

On average, the apps scored as just above minimally acceptable for aesthetics (mean 3.24), including layout, graphics, and visual appeal. This is similar to the aesthetic scores reported by Choi et al [22] and lower than the average aesthetics score reported by Guo et al [45]. Aesthetics is an important dimension of quality that allows apps to stand out in the marketplace. Aesthetics can also facilitate emotionally positive experiences, which can improve user perceptions of the app [46,47].

However, on average, engagement, which included entertainment, customization, interactivity, and fit to the target group, was slightly below acceptable quality (mean 2.76). Similarly, the findings of both Choi et al [22] and Guo et al [45] reported that apps scored lowest on engagement, reporting just-below minimally acceptable quality and above minimally acceptable quality, respectively. These findings further confirm previous research that evaluated 8 commercially available apps for ADRD care partners and found that the majority provided mostly text-based information [48]. Below acceptable engagement scores are concerning, as engagement issues can lead to technology abandonment, reduced acceptance, or failure to use the app to its full potential [49,50]. For care partners, engagement may be critical, as they often experience high levels of demands associated with their caregiver role [31,51,52]. As demonstrated in other populations with chronic health conditions [53,54], engagement is important to sustain care partners' attention when their attention is drawn to the many other demands they experience daily.

Information quality, which included information quantity, visual information, credibility, goals, and app description, also scored, on average, slightly lower than minimally acceptable (mean 2.95, SD 0.95). This is a point lower than the information quality score reported by Choi et al [22]. It is possible that this score difference could be because of information quality differences of the apps designed for people living with ADRD as their target users, which is further supported by a similar high score reported by Guo et al [45]. Information is a critical component to meeting care partners' unmet needs, and low information quality may increase the likelihood of technology abandonment [55]. For example, recent research found that when care partners search for information and cannot meet the information need at that time, they often abandon the information behavior [18]. Furthermore, low-quality information is likely to reduce

perceived usefulness, which has been shown to be a key factor influencing caregivers' intention to adopt mobile health apps [56]. Lower scores on information are also concerning, as this score reflects that apps are often not tested for effectiveness in improving people living with ADRD or care partner outcomes, reducing the ability to safeguard against products that may not deliver on their advertised potential. Specifically, of the 17 apps, 7 (41%) had a mean information quality score that ranged from 1.17 to 2.50 and 11 had a mean subjective quality score that ranged from 1.00 to 2.50. The scoring of both dimensions indicates inadequate quality, which potentially heightens the risk of technology abandonment and loss of the intended impact for target users. Furthermore, apps often state goals without any way to measure or track goal attainment; therefore, there are no clear pathways provided to evaluate whether the stated goals are achievable.

Although most apps met the MARS requirement for minimal acceptability, it may not be sufficient to meet the needs of care partners of people living with dementia. Research on older adults' technology acceptance indicates that they have a higher standard for technology acceptance [57,58]. As many care partners are older adults, raising the bar for acceptable mobile app quality may be critical to sustained care partner use. Furthermore, care partners experience high demands related to their caregiving role and managing complex symptoms and progressive decline and often experience suboptimal health outcomes such as high levels of burden, depression, and anxiety. Therefore, mobile apps may confer some level of risk and need to be held at a high standard so that they do not add burden or increase the risk of suboptimal health outcomes. In addition, an average score at the level of minimal acceptability may mask serious quality violations on one dimension that are counterbalanced by higher-than-average scores on other dimensions. For example, the above average–rated app Respite Mobile (mean MARS quality score 3.35) had a low information quality score (2.27) counterbalanced by particularly high scores on aesthetics (4.0) and functionality (4.25). Thus, minimum standards across dimensions may need to be imposed to avoid harm from counterbalanced weaknesses.

Overall, our ratings of the apps mirror some of those produced from a similar study by Choi et al [22], who also found app engagement scores to be lower than acceptable quality and further highlighted that their scoring differed based on the types of developers (ie, health care–related vs non–health care–related) and intended purpose (ie, awareness, assessment, and disease management). We lacked an appropriate sample to statistically compare differences between developer types. However, we similarly found that for overall mean scores, those developed by commercial entities were just below the minimally acceptable quality, whereas those developed by noncommercial entities were just above the minimally acceptable quality. This comparison further confirms our suggestion to establish higher standardized criteria for health information technology to meet the needs of the care partners of people living with dementia.

Considering the variability in app quality and the failure of many apps to attain acceptable overall and dimension-specific quality ratings, there is a need to adopt quality-focused design and development approaches. One such approach is UCD, introduced earlier and characterized by design driven by a foundational understanding of user needs, direct or indirect input from end users in the design process, and rigorous testing with representative samples of intended end users [16]. In participatory forms of UCD, sometimes called co-design, care partners can also actively contribute to design, leading to a higher likelihood that user needs and abilities are supported and accommodated [59]. UCD approaches can also be used to facilitate engagement through gamification and persuasive design. Furthermore, UCD-based emotional design can increase the quality of aesthetics and functionality [46,47].

Limitations

The results of this study should be considered in light of certain limitations. Not all the raters in our study were experts in technology design. However, we had 3 expert raters who conducted training and acted as arbiters for inclusion decisions and MARS rating. In addition, as per the MARS approach, the raters were not users themselves. To enhance our understanding of the quality of mobile apps for care partners of people living with dementia, future studies should include user testing, such as usability testing and other user tests, alongside expert ratings. Furthermore, we did not rate apps that were available only to study participants. However, the apps we rated are currently available on the market to all users and not limited to the study inclusion and exclusion criteria and participation timelines. Related to this, we were able to rate only what we could access. This means that apps that malfunctioned during log-ins or were only available to customers of a specific health system were not reviewed.

We also identified the limitations of the MARS that should be considered. First, the MARS assumes a typical user and does not address diverse personas, such as users with diverse ages, physical and cognitive abilities, race, ethnicities, and urbanicity or rurality. Second, applying the MARS item definitions is somewhat subjective, and the definitions are not connected to norms, such as a database of prior MARS evaluations. We addressed this limitation through training by reconciling differences in the interpretation of definitions through discussions and consensus building. Third, the MARS does not include certain aspects of design that contribute to app quality, such as security, the design process used, data standards, and accessibility compliance.

Conclusions

In evaluating the quality of publicly available apps for care partners of people living with ADRD, we found that apps, on average, are of minimally acceptable quality. Although we identified apps both above and below the minimally acceptable quality, many apps had broken features and were rated as below acceptable quality for engagement and information quality. Minimally acceptable quality is likely insufficient to meet the needs of care partners without potentially causing harm by increasing burden and stress. Future research should establish minimum quality standards across dimensions for mobile apps for care partners. The design features of high-quality apps identified in this study can provide the foundation for benchmarking these standards.

Acknowledgments

This work was supported by the National Institutes of Health, National Institute on Aging under grants R01 AG056926 and R21 AG062966 (RJH and JCB), 1R21AG072418 (RJH and NEW), and R41AG069607 (NEW and PL). The authors would like to thank Jessica Lee, Manoghna Vuppalapati, and Addison Latterell for their contributions to the MARS rating process.

Conflicts of Interest

RJH provides paid scientific advising to Cook Medical, LLC. None of the apps reviewed are developed or owned by Cook or its subsidiaries.

Multimedia Appendix 1
Agreement and disagreement rates of each dyad before the expert-based score reconciliation process.
[DOCX File , 14 KB - mhealth_v10i3e33863_app1.docx]

References

1. Whitlatch CJ, Orsulic-Jeras S. Meeting the informational, educational, and psychosocial support needs of persons living with dementia and their family caregivers. Gerontologist 2018;58(suppl_1):S58-S73. [doi: 10.1093/geront/gnx162] [Medline: 29361068]
2. Maust D, Leggett A, Kales HC. Predictors of unmet need among caregivers of persons with dementia. Am J Geriatr Psychiatry 2017;25(3):S133-S134. [doi: 10.1016/j.jagp.2017.01.151]
3. Monica MM, Díaz-Santos M, Vossel K. 2021 Alzheimer's disease facts and figures. Alzheimer's Association. 2021. URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf [accessed 2022-03-14]
4. Soong A, Au ST, Kyaw BM, Theng YL, Tudor Car L. Information needs and information seeking behaviour of people with dementia and their non-professional caregivers: a scoping review. BMC Geriatr 2020;20(1):61 [FREE Full text] [doi: 10.1186/s12877-020-1454-y] [Medline: 32059648]
5. McCabe M, You E, Tatangelo G. Hearing their voice: a systematic review of dementia family caregivers' needs. Gerontologist 2016;56(5):e70-e88. [doi: 10.1093/geront/gnw078] [Medline: 27102056]
6. Killen A, Flynn D, De Brún A, O'Brien N, O'Brien J, Thomas AJ, et al. Support and information needs following a diagnosis of dementia with Lewy bodies. Int Psychogeriatr 2016;28(3):495-501. [doi: 10.1017/S1041610215001362] [Medline: 26328546]
7. Yaffe K, Fox P, Newcomer R, Sands L, Lindquist K, Dane K, et al. Patient and caregiver characteristics and nursing home placement in patients with dementia. JAMA 2002;287(16):2090-2097. [doi: 10.1001/jama.287.16.2090] [Medline: 11966383]
8. Coen RF, Swanwick GR, O'Boyle CA, Coakley D. Behaviour disturbance and other predictors of carer burden in Alzheimer's disease. Int J Geriatr Psychiatry 1997;12(3):331-336. [Medline: 9152717]
9. Clyburn LD, Stones MJ, Hadjistavropoulos T, Tuokko H. Predicting caregiver burden and depression in Alzheimer's disease. J Gerontol B Psychol Sci Soc Sci 2000;55(1):S2-13. [doi: 10.1093/geronb/55.1.s2] [Medline: 10728125]
10. Schulz R, O'Brien AT, Bookwala J, Fleissner K. Psychiatric and physical morbidity effects of dementia caregiving: prevalence, correlates, and causes. Gerontologist 1995;35(6):771-791. [doi: 10.1093/geront/35.6.771] [Medline: 8557205]
11. Recommendations from the NIH AD research summit 2015. National Institute on Aging. 2015. URL: https://www.nia.nih.gov/research/recommendations-nih-ad-research-summit-2015 [accessed 2022-03-11]
12. Recommendations from the NIH AD research summit 2018. National Institute of Aging. 2018. URL: https://www.nia.nih.gov/research/administration/recommendations-nih-ad-research-summit-2018 [accessed 2022-03-11]
13. Werner NE, Stanislawski B, Marx KA, Watkins DC, Kobayashi M, Kales H, et al. Getting what they need when they need it. Identifying barriers to information needs of family caregivers to manage dementia-related behavioral symptoms. Appl Clin Inform 2017;8(1):191-205 [FREE Full text] [doi: 10.4338/ACI-2016-07-RA-0122] [Medline: 28224163]
14. Rathnayake S, Moyle W, Jones C, Calleja P. mHealth applications as an educational and supportive resource for family carers of people with dementia: an integrative review. Dementia (London) 2019;18(7-8):3091-3112. [doi: 10.1177/1471301218768903] [Medline: 29631492]
15. Lorca-Cabrera J, Grau C, Martí-Arques R, Raigal-Aran L, Falcó-Pegueroles A, Albacar-Riobóo N. Effectiveness of health web-based and mobile app-based interventions designed to improve informal caregiver's well-being and quality of life: a systematic review. Int J Med Inform 2020;134:104003. [doi: 10.1016/j.ijmedinf.2019.104003] [Medline: 31790857]
16. ISO 9241-210:2019: ergonomics of human-system interaction - Part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2022-03-11]
17. Boustani M, Unützer J, Leykum LK. Design, implement, and diffuse scalable and sustainable solutions for dementia care. J Am Geriatr Soc 2021;69(7):1755-1762. [doi: 10.1111/jgs.17342] [Medline: 34245584]

18. Rutkowski RA, Ponnala S, Younan L, Weiler DT, Bykovskyi AG, Werner NE. A process-based approach to exploring the information behavior of informal caregivers of people living with dementia. Int J Med Inform 2021;145:104341 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104341] [Medline: 33242761]
19. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
20. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
21. Tippey KG, Weinger MB. User-centered design means better patient care. Biomed Instrum Technol 2017;51(3):220-222. [doi: 10.2345/0899-8205-51.3.220] [Medline: 28530864]
22. Choi SK, Yelton B, Ezeanya VK, Kannaley K, Friedman DB. Review of the content and quality of mobile applications about Alzheimer's disease and related dementias. J Appl Gerontol 2020;39(6):601-608 [FREE Full text] [doi: 10.1177/0733464818790187] [Medline: 30049239]
23. 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]
24. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
25. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
26. Peterson K, Hahn H, Lee AJ, Madison CA, Atri A. In the information age, do dementia caregivers get the information they need? Semi-structured interviews to determine informal caregivers' education needs, barriers, and preferences. BMC Geriatr 2016;16(1):164 [FREE Full text] [doi: 10.1186/s12877-016-0338-7] [Medline: 27662829]
27. Esandi N, Nolan M, Alfaro C, Canga-Armayor A. Keeping things in balance: family experiences of living with Alzheimer's disease. Gerontologist 2018;58(2):e56-e67. [doi: 10.1093/geront/gnx084] [Medline: 28633376]
28. Haidet P, Morgan RO, O'Malley K, Moran BJ, Richards BF. A controlled trial of active versus passive learning strategies in a large group setting. Adv Health Sci Educ Theory Pract 2004;9(1):15-27. [doi: 10.1023/B:AHSE.0000012213.62043.45] [Medline: 14739758]
29. Miller JL, Paciga KA, Danby S, Beaudoin-Ryan L, Kaldor T. Looking beyond swiping and tapping: review of design and methodologies for researching young children’s use of digital technologies. Cyberpsychology 2017;11(3):15-27. [doi: 10.5817/CP2017-3-6]
30. Zosh JM, Lytle SR, Golinkoff RM, Hirsh-Pasek K. Putting the education back in educational apps: how content and context interact to promote learning. In: Barr R, Linebarger DN, editors. Media exposure during infancy and early childhood: the effects of content and context on learning and development. Cham, Switzerland: Springer; 2017:259-282.
31. Ponnala S, Block L, Lingg AJ, Kind AJ, Werner NE. Conceptualizing caregiving activities for persons with dementia (PwD) through a patient work lens. Appl Ergon 2020;85:103070 [FREE Full text] [doi: 10.1016/j.apergo.2020.103070] [Medline: 32174358]
32. Neubert L, Gottschalk S, König HH, Brettschneider C. Dementia care-giving from a family network perspective in Germany: a typology. Health Soc Care Community 2022;30(2):579-591. [doi: 10.1111/hsc.13161] [Medline: 32939908]
33. Friedman EM, Kennedy DP. Typologies of dementia caregiver support networks: a pilot study. Gerontologist 2021;61(8):1221-1230 [FREE Full text] [doi: 10.1093/geront/gnab013] [Medline: 33585929]
34. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with Alzheimer disease and related dementia. Can J Aging 2009;28(4):323-336. [doi: 10.1017/S0714980809990158] [Medline: 19925698]
35. Glueckauf RL, Ketterson TU, Loomis JS, Dages P. Online support and education for dementia caregivers: overview, utilization, and initial program evaluation. Telemed J E Health 2004;10(2):223-232. [doi: 10.1089/tmj.2004.10.223] [Medline: 15319052]
36. Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, et al. Categorizing health outcomes and efficacy of mHealth apps for persons with cognitive impairment: a systematic review. J Med Internet Res 2017;19(8):e301 [FREE Full text] [doi: 10.2196/jmir.7814] [Medline: 28855146]
37. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. Am J Alzheimers Dis Other Demen 2013;28(3):216-222 [FREE Full text] [doi: 10.1177/1533317513481091] [Medline: 23528881]
38. Martínez-Alcalá CI, Pliego-Pastrana P, Rosales-Lagarde A, Lopez-Noguerola JS, Molina-Trinidad EM. Information and communication technologies in the care of the elderly: systematic review of applications aimed at patients with dementia and caregivers. JMIR Rehabil Assist Technol 2016;3(1):e6 [FREE Full text] [doi: 10.2196/rehab.5226] [Medline: 28582258]
39. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011;26(1):36-43 [FREE Full text] [doi: 10.1177/1533317510387586] [Medline: 21282276]

40. Lewis ML, Hobday JV, Hepburn KW. Internet-based program for dementia caregivers. Am J Alzheimers Dis Other Demen 2010;25(8):674-679 [FREE Full text] [doi: 10.1177/1533317510385812] [Medline: 21131674]
41. van der Roest HG, Meiland FJ, Jonker C, Dröes RM. User evaluation of the DEMentia-specific Digital Interactive Social Chart (DEM-DISC). A pilot study among informal carers on its impact, user friendliness and, usefulness. Aging Ment Health 2010;14(4):461-470. [doi: 10.1080/13607860903311741] [Medline: 20455122]
42. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
43. Etxeberria I, Salaberria K, Gorostiaga A. Online support for family caregivers of people with dementia: a systematic review and meta-analysis of RCTs and quasi-experimental studies. Aging Ment Health 2021;25(7):1165-1180. [doi: 10.1080/13607863.2020.1758900] [Medline: 32363901]
44. Deeken F, Rezo A, Hinz M, Discher R, Rapp MA. Evaluation of technology-based interventions for informal caregivers of patients with dementia-a meta-analysis of randomized controlled trials. Am J Geriatr Psychiatry 2019;27(4):426-445. [doi: 10.1016/j.jagp.2018.12.003] [Medline: 30642650]
45. Guo Y, Yang F, Hu F, Li W, Ruggiano N, Lee HY. Existing mobile phone apps for self-care management of people with Alzheimer disease and related dementias: systematic analysis. JMIR Aging 2020;3(1):e15290 [FREE Full text] [doi: 10.2196/15290] [Medline: 32012045]
46. Norman DA. Emotional design: why we love (or hate) everyday things. New York, NY: Basic Civitas Books; 2004.
47. Kujala S, Miron-Shatz T. Emotions, experiences and usability in real-life mobile phone use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27-May 2, 2013; Paris, France p. 1061-1070. [doi: 10.1145/2470654.2466135]
48. Wozney L, Freitas de Souza LM, Kervin E, Queluz F, McGrath PJ, Keefe J. Commercially available mobile apps for caregivers of people with Alzheimer disease or other related dementias: systematic search. JMIR Aging 2018;1(2):e12274 [FREE Full text] [doi: 10.2196/12274] [Medline: 31518255]
49. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
50. Zhao Z, Arya A, Whitehead A, Chan G, Etemad SA. Keeping users engaged through feature updates: a long-term study of using wearable-based exergames. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO p. 1053-1064. [doi: 10.1145/3025453.3025982]
51. Waligora KJ, Bahouth MN, Han HR. The self-care needs and behaviors of dementia informal caregivers: a systematic review. Gerontologist 2019;59(5):e565-e583. [doi: 10.1093/geront/gny076] [Medline: 29931147]
52. Ponnala S, Werner NE. Exploring informal caregiver workload using a macroergonomics lens on multiple resources. Proc Hum Factors Ergon Soc Annu Meet 2021;64(1):856-860. [doi: 10.1177/1071181320641198]
53. 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;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
54. Stawarz K, Preist C, Coyle D. Use of smartphone apps, social media, and web-based resources to support mental health and well-being: online survey. JMIR Ment Health 2019;6(7):e12546 [FREE Full text] [doi: 10.2196/12546] [Medline: 31301126]
55. Song T, Deng N, Cui T, Qian S, Liu F, Guan Y, et al. Measuring success of patients' continuous use of mobile health services for self-management of chronic conditions: model development and validation. J Med Internet Res 2021;23(7):e26670 [FREE Full text] [doi: 10.2196/26670] [Medline: 34255685]
56. Mendez KJ, Budhathoki C, Labrique AB, Sadak T, Tanner EK, Han HR. Factors associated with intention to adopt mHealth apps among dementia caregivers with a chronic condition: cross-sectional, correlational study. JMIR Mhealth Uhealth 2021;9(8):e27926 [FREE Full text] [doi: 10.2196/27926] [Medline: 34463637]
57. Mitzner TL, Savla J, Boot WR, Sharit J, Charness N, Czaja SJ, et al. Technology adoption by older adults: findings from the PRISM trial. Gerontologist 2019;59(1):34-44 [FREE Full text] [doi: 10.1093/geront/gny113] [Medline: 30265294]
58. Harris MT, Rogers WA. Understanding acceptance of healthcare technology by older adults: implications for adoption. Innov Aging 2019;3(Suppl 1):S929. [doi: 10.1093/geroni/igz038.3382]
59. Bratteteig T, Wagner I. Disentangling participation: power and decision-making in participatory design. Cham, Switzerland: Springer; 2014.

Abbreviations

ADRD: Alzheimer disease and related dementias
MARS: Mobile App Rating Scale
UCD: user-centered design

Edited by L Buis; submitted 28.09.21; peer-reviewed by K Yin, H Mehdizadeh, SM Ayyoubzadeh, S Rostam Niakan Kalhori ; comments to author 30.12.21; revised version received 16.01.22; accepted 16.02.22; published 29.03.22.

Please cite as:
Werner NE, Brown JC, Loganathar P, Holden RJ
Quality of Mobile Apps for Care Partners of People With Alzheimer Disease and Related Dementias: Mobile App Rating Scale Evaluation
JMIR Mhealth Uhealth 2022;10(3):e33863
URL: https://mhealth.jmir.org/2022/3/e33863
doi:10.2196/33863
PMID:35348467

Original Paper

Comparing Reminders Sent via SMS Text Messaging and Email for Improving Adherence to an Electronic Health Program: Randomized Controlled Trial

Adam Kulhánek[1,2], PhD; Katerina Lukavska[1,2,3], PhD; Roman Gabrhelík[1,2], PhD; Daniel Novák[4], PhD; Václav Burda[2,4], ING; Jindřich Prokop[4], ING; Marianne T S Holter[5,6], PhD; Håvar Brendryen[5,6], PhD

[1]Department of Addictology, General University Hospital in Prague, Prague, Czech Republic
[2]Department of Addictology, First Faculty of Medicine, Charles University, Prague, Czech Republic
[3]Department of Psychology, Faculty of Education, Charles University, Prague, Czech Republic
[4]Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic
[5]The Norwegian Centre for Addiction Research (SERAF), Faculty of Medicine, University of Oslo, Oslo, Norway
[6]Department of Psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway

Corresponding Author:
Katerina Lukavska, PhD
Department of Addictology
General University Hospital in Prague
Apolinářská 4
Prague, 12000
Czech Republic
Phone: 420 224 968 273
Email: katerina.lukavska@lf1.cuni.cz

Abstract

Background: eHealth interventions can help people change behavior (eg, quit smoking). Reminders sent via SMS text messaging or email may improve the adherence to web-based programs and increase the probability of successful behavior change; however, it is unclear whether their efficiency is affected by the modality of the communication channel.

Objective: A 2-armed randomized control trial was conducted to compare the effect of providing reminders via SMS text messaging versus email on the adherence to an eHealth program for smoking cessation and on the probability to initiate a quit attempt.

Methods: Smokers were recruited via an internet-based advertisement. A total of 591 participants who diverted from intended use of the program (ie, failed to log on to a session) were automatically randomized to the experimental (SMS text messaging reminder, n=304) or the active comparator (email reminder, n=287) group.

Results: Unexpectedly, we found that the mode of reminder delivery did not significantly affect either the adherence, namely the number of completed program sessions, with the SMS text messaging reminder group showing a mean of 4.30 (SD 3.24) and the email reminder group showing a mean of 4.36 (SD 3.27) (t_{586}=0.197, *P*=.84, and Cohen *d*=0.016), or the outcome, namely the quit smoking attempt rate (34.2% in the SMS text messaging group vs 31.7% in the email group; χ^2_1=0.4, *P*=.52). Secondary analyses showed that age, gender, and education had significant effects on program adherence and education on the outcome. Moreover, we found a significant interaction effect between the mode of reminder delivery and gender on program adherence, suggesting that the effectiveness of SMS text message reminders might be different for females and males. However, this particular finding should be treated with care as it was based on post hoc subgroup analysis.

Conclusions: This study indicates that the modality of user reminders to log on increased neither the program adherence nor the probability of quitting smoking. This suggests that program developers may save costs using emails instead of SMS text messaging reminders.

Trial Registration: ClinicalTrials.gov NCT03276767; https://clinicaltrials.gov/ct2/show/ NCT03276767

(JMIR Mhealth Uhealth 2022;10(3):e31040) doi:10.2196/31040

KEYWORDS

eHealth; randomized controlled trial; adherence; reminders; SMS text messaging; email; smoking cessation; text message

Introduction

Interventions delivered through the internet may provide people with tailored and real-time suggestions [1] and allow targeting and attracting large populations [2-4]. Web-based interventions can help people change health behavior [5], including quitting tobacco smoking [6,7], which is still one of the leading causes of avoidable mortality and morbidity worldwide [8,9]. Although we know that web-based interventions for smoking cessation may improve quitting rates, there is a need for research into factors that may increase the efficacy of such interventions [10].

The efficacy of web-based interventions is closely associated with users' adherence to them [11], making it pertinent for program designers to find ways of increasing adherence to the programs they design. One way of increasing program adherence is through digital triggers that are external stimuli "designed to make an individual focus on a desired goal by prompting an internal or external reaction at the appropriate time" [12]. Such triggers can be integrated in an otherwise web-based program as notifications when new program content is made available and as reminders for logging on once the user fails to log on as expected. However, developing digital triggers involves a range of design choices, including "who" (ie, sender), "how" (ie, medium), "when" (ie, triggered by what), "how much" (ie, how often), and "what" (ie, content) [12]. Optimizing these design choices in the best manner is an empirical question; however, the evidence on how to design effective triggers is mixed due to insufficient reporting of design choices and heterogeneity in studies [12]. This study seeks to contribute to this knowledge base by specifying the best choice of the delivery mode (the "how") for digital triggers designed to increase adherence of users who fail to log on as expected. Different options exist, but 2 commonly used alternatives are SMS text messaging and emails, with each having different advantages and disadvantages [12]. For example, SMS text messaging (compared to email) is more salient to the receiver [3] and linked to higher open and click rates [12], but there is a higher cost associated with SMS text messaging (ie, the cost of sending as well as that for development and maintenance of the system). A meta-study of web-based interventions found that using additional methods of communicating with participants, like email or SMS text messaging, were associated with larger effects on behavior change, and more specifically, this effect was reported to be large for SMS text messaging and small for email [10]. However, a limitation of this meta-study is that these conclusions were based on comparing the effect sizes in studies with no reminders or email reminders to studies using SMS text messaging reminders. Several direct comparisons of modalities across arms within a randomized trial, and across various contexts, is needed to address this issue. Nevertheless, these studies appear to indicate that regarding reminders to log on to a web-based program, SMS text messaging reminders would be superior to email reminders. Addressing this matter is significant for program designers, as adding an SMS text messaging component will usually entail additional costs to program development and should thus be worth the money.

Therefore, the purpose of this study was to determine whether SMS text messaging would indeed be more effective in reminding the users to log onto the web application once they fail to log on as expected. Our point of departure was a web-based smoking cessation program (described below) that uses email invitations to notify the user of the release of new program modules. If the user does not log on as expected, the program will send out a reminder to prompt the user to log on. We conducted a randomized controlled trial (RCT) in which participants were randomized to receive these reminders either by SMS text messaging or email. We hypothesized that SMS text messaging reminders would have users logging onto the web application frequently (ie, adherence) and possibly lead to an increased likelihood of an initiated attempt to quit smoking (due to increased program use).

Methods

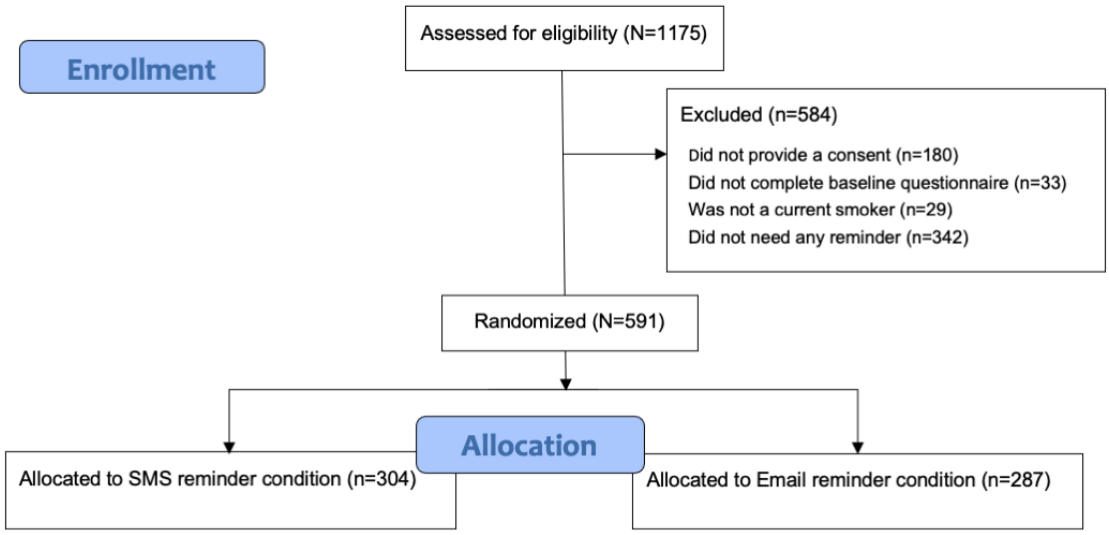
Study Design

A 2-armed RCT was conducted. The 2 arms of the RCT differed only in the modality (SMS text messaging versus email) of the reminders that were issued to remind the smoker about the missed session (module). There were 4 different versions of this message, so that each user would never receive the same message twice in a row. The messages were very similar in content and form, for example, "Hi (name of user) I haven't seen you for a while. You may have been busy? Hope to see you soon! (smart link to the web application) Best Andy (Andy is the English name of the intervention)." The study protocol was registered at ClinicalTrials.gov (trial registration number: NCT03276767).

Participants

The study sample consisted of Czech and Norwegian tobacco smokers using Andy, an eHealth smoking cessation program (described below). Czech participants were recruited through internet-based advertisements (webpages focused on smoking cessation, social media platforms, and internet-based newspapers). Norwegian participants were recruited through internet-based advertisements (Facebook, Google, blogposts, and newspapers) as well as through Healthy Life Centers. Participants needed to be over 18 years old, current tobacco smokers, willing to quit smoking, provide a valid email address and mobile phone number, provide consent to participation, complete the baseline questionnaire, and open the first session of the program. Additionally, only persons who qualified as nonadherers were included; nonadherence was defined as failing to log on to the program by noon on the day after a new session (module) was released. Overall, 584 of the 1175 recruited participants (49.7%) did not meet the inclusion criteria, resulting in a total of 591 participants. Figure 1 shows the participant selection process.

Figure 1. Flow diagram of participant selection.



Data Collection

Baseline characteristics were collected through a web-based questionnaire, whereas program use information was recorded automatically. Program use information included the start and completion of a session, issuing of SMS text messaging/email reminders, and user-reported initiation of a quit attempt.

Intervention and Randomization

The eHealth intervention named "Andy" is a fully automated web-based smoking cessation program that first prepares the user for quitting (preparation phase) and follows up once the user confirms having initiated a quit attempt (follow-up phase); however, this study focuses only on the preparation phase. During the preparation phase, the program releases 1 new session per day. The recommended/default length of the preparation phase is 11 days (which is the designated quit day), but users are free to advance or postpone at their own rate within certain limits. When a new session is released, an email invitation is sent to the user to prompt login (this happens at 5 AM). A reminder is due when a user fails to log onto an assigned session by noon on the next day. In this study, a simple randomization procedure automatically took place when the first reminder was due, following which participants were randomized to either the SMS text messaging or the email condition. More information about the intervention program, how it is used, and its usability can be obtained elsewhere [13-16].

Measures

Demographics

Participants were asked about their age, gender, place of residence (urban or rural area), education, employment, and income.

Baseline Smoking

Participant also reported their average consumption of cigarettes (self-reported number of cigarettes smoked per day). Following a standard procedure [17], we then categorized participants as a mild (<10 cigarettes/day), moderate (11-19 cigarettes/day), or intensive (>20 cigarettes/day) smokers.

Primary Outcome: Initiation of Quit Attempt

The main outcome was whether the participants had initiated a quit attempt. Upon logging on to the program on the quit day, participants are asked whether they have in fact initiated a quit attempt as planned. Those who answer "no" are encouraged to quit the next day and will be asked the same question the day after (and on any subsequent day if they keep logging on to the program). Those who answer "yes" will be transferred to the follow-up phase of the program. Participants who dropped out from the program before they reached their quit day were considered treatment failures (ie, smokers). Participants who continued to log on to the program until their quit day (or beyond) but did not report to have initiated a quit attempt within 6 weeks after accessing the first session of the program were also considered treatment failures.

Secondary Outcome: Number of Completed Sessions

Program adherence was measured as the number of eHealth program sessions that participant had completed. It was possible to complete 0 to 10 sessions during the preparation phase of the eHealth program. A higher number of completed sessions indicated higher adherence to the eHealth program.

Statistical Analysis

Statistical analyses were conducted in multiple steps. First, we aimed to assess randomization by confirming the absence of significant differences in background variables between the experimental group (SMS text messaging reminders) and the active control group (email reminders). Second, we used the Welch 2-tailed *t* test for unequal variance to analyze the effect of SMS text messaging reminders on program adherence, namely the number of completed program sessions (secondary outcome). Next, we used the χ^2 test of association for assessing the effect of SMS text messaging reminders on the initiation of a quit attempt (primary outcome). Moreover, 2 regression analyses were conducted to assess the effects of SMS text messaging reminders when controlling for the effects of all background variables (shown in Table 1) on the primary and

secondary outcomes. Post hoc analyses of the relationship between the background and outcome variables were also conducted. Analyses were conducted with the statistical software R [18,19].

Ethics

The study was approved by the Ethics Committee of the General University Hospital in Prague (no. 7/17GrantGACR-1.LFUK) and by the Norwegian Center for Research Data (no. 52874).

Results

Baseline Characteristics

Background characteristics of each group are reported in Table 1. Participants (program users) had a mean age of 39.5 (SD 12.8) years, and 52% of the 591 participants (n=308) were Czech and 48% (n=288) were Norwegian. Additionally, 61% (n=361) were female, 60% (n=355) were full-time employees, and 56% (n=330) reported high school as their highest completed education. The mean consumption of tobacco among participants at baseline was 18 cigarettes per day, and 43% (n=254) reported consuming more than 20 cigarettes per day. There were no statistically significant differences found between the SMS text messaging and email groups, except for the nationality distribution (Table 1).

Table 1. Baseline characteristics of the experimental (SMS text messaging) group and active control (email) arm of the randomized controlled trial (N=591)[a].

Characteristic	SMS text messaging (n=304, 51.4%)	Email (n=287, 48.6%)	$t/\chi^2(df)$
Age in years (range 18-77), mean (SD)	40.0 (12.9)	38.8 (12.8)	t (587)=1.12
Nationality, n (%)			χ^2 (1)=8.3
Czech	141 (45.8)	167 (54.2)	
Norwegian	163 (57.6)	120 (42.4)	
Gender, n (%)			χ^2 (1)=0.5
Female	190 (52.6)	171 (47.4)	
Male	114 (49.6)	116 (50.4)	
Residence, n (%)			χ^2 (3)=3.8
<1000 inhabitants	47 (46.1)	55 (53.9)	
1000-20,000 inhabitants	94 (57)	71 (43)	
20,000-100,000 inhabitants	92 (52)	85 (48)	
>100,000 inhabitants	71 (48.3)	76 (51.7)	
Education, n (%)			χ^2 (3)=1.7
<HS[b] graduate	26 (47.3)	29 (52.7)	
HS graduate	169 (51.2)	161 (48.8)	
University (BA[c] degree)	73 (55.7)	58 (44.3)	
University (MA[d] degree or higher)	36 (48)	39 (52)	
Employment, n (%)			χ^2 (5)=1.0
Freelancer	35 (53)	31 (47)	
Employed	178 (50.1)	177 (49.9)	
Unemployed	41 (55.4)	33 (44.6)	
Student	25 (50)	25 (50)	
Retired	11 (52.4)	10 (47.6)	
Other	14 (56)	11 (44)	
Income, n (%)			χ^2 (4)=7.4
Very low	41 (56.9)	31 (43.1)	
Low	82 (59.4)	56 (40.6)	
Middle	58 (45)	71 (55)	
High	92 (49.2)	95 (50.8)	
Very high	19 (46.3)	22 (53.7)	
Smoking, n (%)			χ^2 (2)=0.1
<10 cigarettes/day	79 (52.3)	72 (47.7)	
11-19 cigarettes/day	96 (51.6)	90 (48.4)	
>20 cigarettes/day	129 (50.8)	125 (49.2)	
Reminders (range 1-8)	2.58 (1.17)	2.56 (1.09)	t (589)=0.27

[a]*P* values for all variables were not significant except Nationality (*P*<.001).

[b]HS: high school.

[c]BA: Bachelor of Arts.

[d]MA: Master of Arts.

Effectiveness of SMS Text Messaging Versus Email Reminders on Program Adherence – The Number of Completed Sessions (Secondary Outcome)

The number of completed sessions among all participants ranged from 0 to 10 (mean 4.33, SD 3.26). Contrary to our expectations, we did not find any statistically significant difference in the number of completed sessions between participants receiving SMS text messaging reminders (mean 4.30, SD 3.24) and those receiving email reminders (mean 4.36, SD 3.27); Welch t_{586}=0.197, P=.84, and Cohen d=0.02. Given the difference in the Czech/Norwegian participant ratio between the SMS text messaging and email groups, we conducted a regression analysis where we controlled for nationality (and other background variables) with the same result. Receiving SMS text messaging reminders (as compared with receiving email reminders) did not lead to a significantly higher number of completed sessions (P=.98). Table 2 presents the regression analysis results.

Some background characteristics were found to be significant predictors of the number of completed sessions, namely age, nationality, gender, and education. Using separate analyses for each of these predictors, we found that higher age was positively associated with adherence (r=0.17, P<.001) and female participants showed a significantly higher number of completed sessions (mean 4.64, SD 3.32) compared to male participants (mean 3.84, SD 3.09); Welch t_{513}=2.98, P=.003, and Cohen d=0.25. Furthermore, participants with a university degree, namely Bachelor of Arts or Master of Arts, completed more sessions compared to those without high school graduation; the omnibus difference was significant ($F_{3,166}$=3.87, P=.01) with Games-Howell post hoc t tests proving significant for differences between those without high school graduation and those with BA (P=.02) and MA (or higher) degrees (P=.02). The difference in the number of completed sessions between Czech (mean 4.17, SD 3.14) and Norwegian (mean 4.51, SD 3.37) participants was not significant (P=.27).

Although we found no significant main effect of the reminder delivery mode on adherence, post hoc analyses revealed a significant interaction with gender; $F_{1,587}$=4.10 and P=.04. The finding suggests that the effect of replacing email reminders with SMS text messaging reminders is more beneficial for males, relative to females, in terms of improved program adherence. The average number of completed sessions for female users receiving SMS text messaging reminders (n=190) was 4.40 (SD 3.40), whereas for male users receiving SMS text messaging reminders (n=114), it was 4.14 (SD 2.98); further, for female users receiving email reminders (n=171), the average number of completed sessions was 4.91 (SD 3.22), and for male users receiving email reminders (n=116), it was 3.54 (SD 3.19) (Figure 2).

Table 2. Summary of linear regression analysis for variables predicting the number of completed sessions (N=591, R^2=0.0796)[a].

Predictor	Regression analysis variables					
	B	SE *B*	*t*	*P*	β	95% CI
Reminder						
SMS text messaging	0.007	0.2725	0.025	.98	.002	–0.162 to 0.166
Age	0.052	0.0158	3.288	.001	.200	0.081 to 0.320
Nationality						
Czech	0.951	0.4608	2.064	.04	.292	0.014 to 0.569
Gender						
Female	0.771	0.3187	2.419	.02	.236	0.044 to 0.428
Residence						
1000-20,000 inhabitants	0.195	0.4208	0.464	.64	.060	–0.194 to 0.313
20,000-100,000 inhabitants	–0.200	0.4133	–0.485	.63	–.061	–0.310 to 0.187
>100,000 inhabitants	–0.373	0.4455	–0.838	.4	–0.114	–0.383 to 0.154
Education						
HS[b] graduate	0.846	0.5069	1.668	.1	.259	–0.046 to 0.565
University (BA[c] degree)	1.197	0.5763	2.078	.04	.367	0.020 to 0.714
University (MA[d] degree or higher)	1.218	0.6370	1.912	.06	.373	–0.010 to 0.757
Employment						
Employed	0.537	0.4558	1.178	.24	.165	–0.110 to 0.439
Unemployed	–0.231	0.6446	–0.358	.72	–.071	–0.459 to 0.317
Student	–0.458	0.7891	–0.580	.56	–.140	–0.616 to 0.335
Retired	–0.170	0.9076	–0.187	.85	–.052	–0.600 to 0.500
Other	–1.091	0.8291	–1.316	.19	–.334	–0.834 to 0.165
Income						
Low	–0.580	0.5079	–1.142	.25	–.178	–0.484 to 0.128
Middle	–0.286	0.5792	–0.494	.62	–.088	–0.440 to 0.261
High	–0.492	0.6555	–0.751	.45	–.151	–0.546 to 0.244
Very high	–1.031	0.8109	–1.272	.2	–.316	–0.805 to 0.172
Smoking						
11-19 cigarettes/day	0.068	0.3691	0.185	.85	.021	–0.201 to 0.243
>20 cigarettes/day	–0.260	0.3643	–0.713	.48	–.080	–0.300 to 0.140

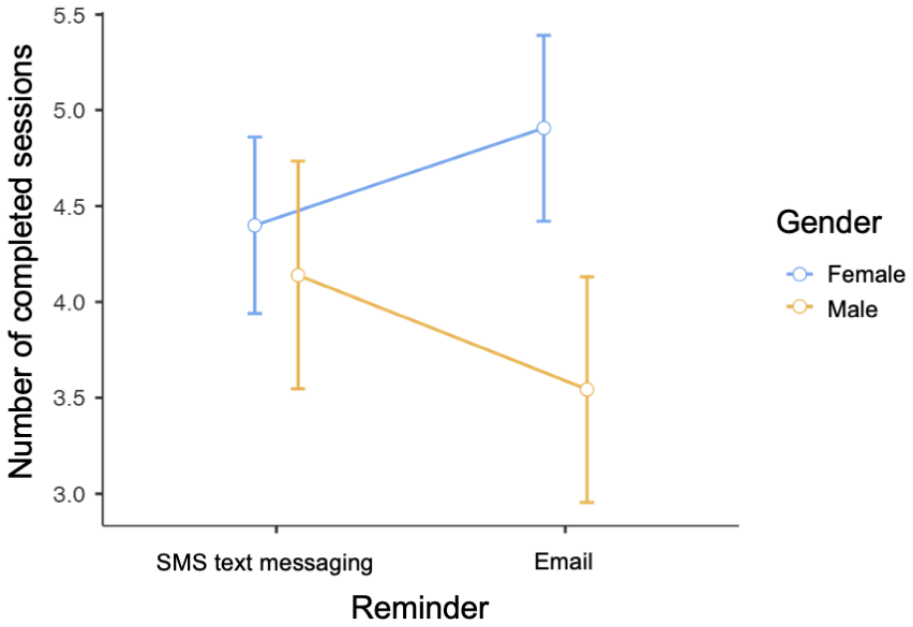
[a]*B* represents the log odds of quit attempt=1 versus quit attempt. represents standardized estimates. “Email” is the reference category for Reminder. “Norwegian” is the reference category for Nationality. “Female” is the reference category for Gender. “<1000 inhabitants” is the reference category for Residence. “<HS graduate” is the reference category for Education. “Freelancer” is the reference category for Employment. “Very low” is the reference category for Income. “<10 cigarettes/day” is the reference category for Smoking.

[b]HS: high school.

[c]BA: Bachelor of Arts.

[d]MA: Master of Arts.

Figure 2. Effects of randomized condition (SMS text messaging versus email reminders) and gender on the electronic health program adherence (the number of completed sessions). Estimated marginal means with 95% CIs are shown.



Effectiveness of the SMS Text Messaging and Email Reminders on the Initiation of Quit Attempt (Primary Outcome)

In the whole sample comprising 591 participants, 195 (33%) participants initiated a quit attempt. Contrary to our hypothesis, there was no significant difference between the 2 randomized groups in initiating a quit attempt (χ^2_1=0.4, *P*=.52). The frequency of quit attempts in the SMS text messaging group was 104 (34.2%) whereas it was 91 (31.7%) in the email group. We did not find any significant interaction between education, receiving SMS text message versus email reminders, and quit attempts. Regression analysis in which background variables were controlled for also showed that receiving SMS text messaging reminders instead of email reminders was not a significant predictor of initiating a quit attempt (Table 3). From all analyzed sociodemographic variables, only education was a significant predictor of initiating a quit attempt (Table 3). Participants with an education lower than high school (ie, elementary or practical education) reported initiating a quit attempt in 9 (16.4%) cases, showing a 2 times lower prevalence than in the whole sample. In comparison, 106 (32.1%) of the high school–graduated participants, 49 (37.4%) of the college-graduated participants with a bachelor's degree, and 31 (41.3%) of the college-graduated participants with a master's degree reported having initiated an attempt to quit smoking (χ^2_3=10.5, *P*=.02).

Table 3. Summary of logistic regression analysis for variables predicting quit attempt in electronic health program users (N=591, R^2=0.034)[a].

Predictor	Regression analysis variables				
	B	SE *B*	Z[b]	*P*	Odds ratio, 95% CI
Reminder					
SMS text messaging	0.1394	0.1852	0.7530	.45	1.150, 0.7997-1.653
Age	0.0163	0.0107	1.5273	.13	1.016, 0.9954-1.038
Nationality					
Czech	0.1477	0.3152	0.4686	.64	1.159, 0.6250-2.150
Gender					
Female	–0.0136	0.2153	–0.0632	.95	0.986, 0.6468-1.504
Residence					
1000-20,000 inhabitants	–0.0909	0.2843	–0.3197	.75	0.913, 0.5231-1.594
20,000-100,000 inhabitants	–0.1642	0.2795	–0.5874	.56	0.849, 0.4906-1.468
>100,000 inhabitants	–0.1370	0.3010	–0.4553	.65	0.872, 0.4834-1.573
Education					
HS[c] graduate	0.8912	0.4158	2.1433	.03	2.438, 1.0792-5.508
University (BA[d] degree)	0.9175	0.4542	2.0202	.04	2.503, 1.0277-6.096
University (MA[e] degree or higher)	1.1533	0.4855	2.3755	.02	3.169, 1.2235-8.206
Employment					
Employed	0.2879	0.3168	0.9087	.36	1.334, 0.7167-2.482
Unemployed	0.0998	0.4419	0.2259	.82	1.105, 0.4647-2.627
Student	0.0595	0.5551	0.1072	.92	1.061, 0.3576-3.150
Retired	–0.5661	0.6345	–0.8923	.37	0.568, 0.1637-1.969
Other	–0.3773	0.6125	–0.6160	.54	0.686, 0.2065-2.278
Income					
Low	–0.1251	0.3522	–0.3552	.72	0.882, 0.4425-1.760
Middle	0.1129	0.3958	0.2851	.78	1.119, 0.5154-2.432
High	–0.0173	0.4484	–0.0386	.97	0.983, 0.4081-2.367
Very high	–0.5921	0.5684	–1.0418	.3	0.553, 0.1816-1.685
Smoking					
11-19 cigarettes/day	0.1315	0.2429	0.5414	.59	1.141, 0.7085-1.836
>20 cigarettes/day	–0.3045	0.2475	–1.2304	.22	0.737, 0.4540-1.198

[a]*B* represents the log odds of quit attempt=1 versus quit attempt. “Email” is the reference category for Reminder. “Norwegian” is the reference category for Nationality. “Female” is the reference category for Gender. “<1000 inhabitants” is the reference category for Residence. “<HS graduate” is the reference category for Education. “Freelancer” is the reference category for Employment. “Very low” is the reference category for Income. “<10 cigarettes/day” is the reference category for Smoking.

[b]Z: regression coefficient divided by the standard error.

[c]HS: high school.

[d]BA: Bachelor of Arts.

[e]MA: Master of Arts.

Discussion

Principal Results

This RCT tested the hypothesis that receiving SMS text messaging reminders (compared to receiving email reminders) increased (1) the adherence to the eHealth program for smoking cessation and (2) the initiation of an attempt to quit smoking. Randomization took place after the first sign of nonadherence to the eHealth program (ie, when a user failed to log on to the program as expected and was due to receive the first reminder). There were no significant differences between the 2 groups in

terms of their background characteristics, except for nationality. Norwegian participants were more often randomized to the SMS text messaging group as compared to Czech participants. The adherence to the eHealth program was measured as the number of completed sessions and the desired outcome was measured as self-reported initiation of a quit attempt. Surprisingly, we did not find any significant differences in the number of completed sessions between participants receiving SMS text messaging reminders (completed 4.30 sessions on average) and those receiving email reminders (completed 4.36 sessions on average), when tested separately (P=.84) or when controlled for all the background variables listed in Table 1 (P=.98). Similarly, we did not find differences in the proportion of reported quit attempts between SMS text messaging (quit attempt reported by 34.2% participants) and email (quit attempt reported by 31.7% participants) groups, either when measured separately (P=.52) or when controlled for all background variables (P=.45).

Additional post hoc analyses revealed significant effects of some sociodemographic variables on program adherence (age, gender, and education) and the initiation of a quit attempt (education). None of these effects interacted significantly with the reminder modality (SMS text messaging versus email), except for gender, which is attributable to the modality interaction effect on program adherence (P=.04). In other words, the effect of replacing email reminders with SMS text messaging reminders on program adherence is heterogenous across genders. This effect suggests that SMS text messaging reminders are more beneficial for men, relative to women, in terms of program adherence (see Figure 2). (Note that we did not find any interaction between the modality and gender on quit attempts, or between the modality and any other background variable.)

In summary, post hoc subgroup analyses revealed that the choice of the optimal modality may depend on gender. SMS text messaging is more beneficial for males relative to females regarding program adherence. We did not find overall improvement in program adherence on receiving SMS text messaging reminders when compared to email reminders. More importantly, we found that the reminder modality did not affect the main outcome, namely smoking cessation.

Comparison With Prior Work

Prior work has shown that external triggers, such as reminders, may improve adherence to eHealth programs and thus the outcomes of these interventions [12]. The evidence of how to design effective triggers is mixed due to insufficient reporting of design choices and heterogeneity in studies [12]. This study focused on the mode of delivery of triggers (reminders), for which SMS text messages and emails are 2 popular options considered by designers. Compared to an email, an SMS text message is more salient to the receiver [3], linked to higher open and click rates [12], and associated with larger effects on eHealth-supported behavioral change [10]. These previous findings suggest that SMS text messaging is superior to email for reminding users to log on to an eHealth web program, thus increasing adherence and the probability of desired outcomes, and is therefore worth the additional cost. However, contrary to our expectations, we did not find evidence supporting this superiority of SMS text messaging over email reminders either with respect to the program adherence or to the outcomes of this specific eHealth program (ie, the initiation of a quit attempt). However, the effect of the reminder delivery mode on program adherence may be affected by gender; in our study, female participants were found to be generally more adherent to the eHealth program (completed more sessions) and the difference was particularly strong in the condition of email reminders. Although there is some evidence that women are more compliant with eHealth interventions in general [20], to our best knowledge, there is no study available that has analyzed the relationships between gender, eHealth adherence, and different modes of delivering reminders. Our results suggested that SMS text messaging reminders (compared to email reminders) might help reduce the gender-based difference in adherence. This would be an interesting area for further research. It should be noted that many other factors might be influencing the efficiency of the reminders, such as content, frequency, time of delivery, type of the intervention program (eg, web-based or mobile app, frequency and number of session releases, etc), or even the phase of the intervention (eg, reminders might affect users differently when received at the beginning as opposed to the later phase).

Strengths and Limitations

This study has several strengths, including a heterogenous sample with participants belonging to a wide spectrum of sociodemographic groups (Table 1) from the Czech Republic and Norway, 2 countries with different levels of smoking prevalence and tobacco use patterns. The other strength is that the 2 groups differed only in the mode of communication for reminders to log on to the program, whereas the content and number of reminders as well as the content of the program in general were the same for both groups. Therefore, we could assess the direct effect of the reminder delivery mode on program adherence and the desired outcome of the intervention. Generally, SMS text messaging reminders are often used within health care but RCTs assessing their effect are lacking [21]. In addition, the use of automatically collected eHealth data reduced selection bias and the risk of recall bias (although the initiation of a quit attempt was self-reported).

One limitation of the study is that the preparation phase of the program was fairly short (maximum 11 days), resulting in a short period for assessing the adherence. Moreover, the study focused on a smoking cessation eHealth program and may not be generalized for other types of eHealth interventions. Findings concerning the interaction between SMS text messaging reminders and gender are based on post hoc subgroup analysis, and as such, it should be treated with care [22]. Further, a question that our study did not address was that individual differences might influence the effect of the reminder modality. Further research might inquire into the potential of tailoring reminder modalities to individual preferences.

Conclusions

In conclusion, and contrary to available literature, our data suggested that when it comes to reminding nonadherent eHealth users to log on to a web-based program, SMS text messaging reminders were not superior to email reminders, neither with respect to increasing program adherence nor in supporting a

desired outcome (ie, the initiation of a quit attempt). However, there may be gender differences affecting the preferred modality (with email reminders being more effective for female users) that may be useful to pursue in further research. The results for both outcomes taken together suggest that there is very little to gain, if anything at all, by choosing SMS text messaging over email reminders for web-based behavior change interventions. Thus, our finding is important for developers and providers of eHealth interventions who may not need to allocate additional costs related to SMS text messaging reminders to enhance program adherence or outcomes, as reminders delivered via email seem to be equally effective.

Acknowledgments

This study was supported by the institutional support program Progres (grant Q06/LF1), the GCDPC (grant AA-01-20), Specific Academic Research (grant 260500), and the Norwegian Research Council (grant 228158/H10). The study was also supported by the Research Center for Informatics (grant CZ.02.1.01/0.0/16_019/0000765) and by Biomedical data acquisition, processing, and visualization (grant SGS19/171/OHK3/3T/13).

Authors' Contributions

AK coordinated the recruitment process, carried out the literature review, and prepared the first draft of the manuscript. HB conceived the idea of the RCT and wrote the protocol. HB and MTSH developed the intervention program, including the content and procedure for reminder messages. Data analysis and the preparation of the second draft of the manuscript were performed by KL. HB, VB, JP, and DN managed internet-based data collection and information technology support of the eHealth program. KL, RG, AK, and HB participated in data interpretation. All authors contributed to the final version of the manuscript and approved the same.

Conflicts of Interest

RG is a shareholder of Adiquit Ltd, which is currently developing apps for treatment of addictions. Nevertheless, no funding was related to this study and the activities had no role in the study design or the collection, analysis, and interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication. DN is a shareholder of Adiquit Ltd developing apps for treatment of addictions, Mindpax Ltd, developing digital tools for psychiatry, and Blindshell Ltd, developing mobile phones for visually impaired users. However, no funding was related to this study and the activities had no role in the study design or the collection, analysis, and interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication. The other authors declare no conflicts of interest.

Multimedia Appendix 1
CONSORT eHealth checklist (V 1.6.1).
[PDF File (Adobe PDF File), 402 KB - mhealth_v10i3e31040_app1.pdf]

References

1. Noar S, Harrington N. eHealth Applications: Promising Strategies for Behavior Change. United Kingdom: Routledge; 2012.
2. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015 Jan;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
3. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
4. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in internet smoking cessation interventions: insights from a cross-sectional survey of "one-hit-wonders". Internet interv 2016 Sep;5:23-29 [FREE Full text] [doi: 10.1016/j.invent.2016.07.001]
5. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. "A meta-analysis of web-delivered tailored health behavior change interventions": corrigendum. J Health Commun 2013 Sep;18(11):1397. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
6. Do HP, Tran BX, Pham QL, Nguyen LH, Tran TT, Latkin CA, et al. Which eHealth interventions are most effective for smoking cessation? A systematic review. Patient Prefer Adherence 2018 Oct;12:2065-2084. [doi: 10.2147/PPA.S169397]
7. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Dec;9:1-129. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
8. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. Br J Cancer 2004;328:1519. [doi: 10.1038/sj.bjc.6602359]
9. Reitsma MB, Kendrick PJ, Ababneh E, Abbafati C, Abbasi-Kangevari M, Abdoli A, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019:

a systematic analysis from the Global Burden of Disease Study 2019. The Lancet 2021 May;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7]
10. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
11. Graham AL, Carpenter KM, Cha S, Cole S, Jacobs MA, Raskob M, et al. Systematic review and meta-analysis of Internet interventions for smoking cessation among adults. Subst Abuse Rehabil 2016 May;7:55-69. [doi: 10.2147/sar.s101660]
12. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]
13. Holter MTS, Johansen A, Brendryen H. How a fully automated eHealth program simulates three therapeutic processes: a case study. J Med Internet Res 2016 Jun;18(6):e176. [doi: 10.2196/jmir.5415]
14. Kulhánek A, Gabrhelík R, Novák D, Burda V, Brendryen H. eHealth intervention for smoking cessation for Czech tobacco smokers: pilot study of user acceptance. Adiktologie 2018:81-85 [FREE Full text]
15. Holter MTS, Johansen AB, Ness O, Brinkmann S, Høybye MT, Brendryen H. Qualitative interview studies of working mechanisms in electronic health: tools to enhance study quality. J Med Internet Res 2019 May;21(5):e10354. [doi: 10.2196/10354]
16. Holter MTS, Ness O, Johansen AB, Brendryen H. Making come-alive and keeping un-alive: how people relate to self-guided web-based health interventions. Qual Health Res 2020 Feb;30(6):927-941. [doi: 10.1177/1049732320902456]
17. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989 Jul;84(7):791-800 [FREE Full text] [doi: 10.1111/j.1360-0443.1989.tb03059.x]
18. Lenth RV. Estimated marginal means, aka least-squares means. R Package. 2018. URL: https://cran.r-project.org/package=emmeans [accessed 2019-01-01]
19. The Comprehensive R Archive Network. 2018. URL: https://cran.r-project.org [accessed 2019-01-01]
20. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
21. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct;16(10):e222. [doi: 10.2196/jmir.3442]
22. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001;5(33):1-56. [doi: 10.3310/hta5330]

Abbreviations

RCT: randomized controlled trial

Edited by L Buis; submitted 07.06.21; peer-reviewed by L Morrison, B Smith; comments to author 30.07.21; revised version received 12.08.21; accepted 20.12.21; published 18.03.22.

Please cite as:
Kulhánek A, Lukavska K, Gabrhelík R, Novák D, Burda V, Prokop J, Holter MTS, Brendryen H
Comparing Reminders Sent via SMS Text Messaging and Email for Improving Adherence to an Electronic Health Program: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(3):e31040
URL: https://mhealth.jmir.org/2022/3/e31040
doi:10.2196/31040
PMID:35302945

Original Paper

Safety of Triage Self-assessment Using a Symptom Assessment App for Walk-in Patients in the Emergency Care Setting: Observational Prospective Cross-sectional Study

Fabienne Cotte[1,2,3], MD; Tobias Mueller[4], MD; Stephen Gilbert[3,5], PhD; Bibiana Blümke[3], MD; Jan Multmeier[3], PhD; Martin Christian Hirsch[3,6], PhD; Paul Wicks[3], PhD; Joseph Wolanski[3], BSc; Darja Tutschkow[7], MSc; Carmen Schade Brittinger[7], BSc; Lars Timmermann[8], PhD; Andreas Jerrentrup[2], MD

[1]Charité Universitätsmedizin Berlin, Berlin, Germany

[2]Department of Emergency Medicine, University Clinic Marburg, Philipps-University, Marburg, Germany

[3]Ada Health GmbH, Berlin, Germany

[4]Center for Unknown and Rare Diseases, UKGM GmbH, University Clinic Marburg, Philipps-University, Marburg, Germany

[5]Else Kröner Fresenius Center for Digital Health, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[6]Institute of Artificial Intelligence, Philipps-University Marburg, Marburg, Germany

[7]Coordinating Center for Clinical Trials, Philipps University Marburg, Marburg, Germany, Marburg, Germany

[8]Department of Neurology, University Hospital of Marburg, Marburg, Germany

Corresponding Author:
Fabienne Cotte, MD
Charité Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 50
Email: faco@gmx.de

Abstract

Background: Increasing use of emergency departments (EDs) by patients with low urgency, combined with limited availability of medical staff, results in extended waiting times and delayed care. Technological approaches could possibly increase efficiency by providing urgency advice and symptom assessments.

Objective: The purpose of this study is to evaluate the safety of urgency advice provided by a symptom assessment app, Ada, in an ED.

Methods: The study was conducted at the interdisciplinary ED of Marburg University Hospital, with data collection performed between August 2019 and March 2020. This study had a single-center cross-sectional prospective observational design and included 378 patients. The app's urgency recommendation was compared with an established triage concept (Manchester Triage System [MTS]), including patients from the lower 3 MTS categories only. For all patients who were undertriaged, an expert physician panel assessed the case to detect potential avoidable hazardous situations (AHSs).

Results: Of 378 participants, 344 (91%) were triaged the same or more conservatively and 34 (8.9%) were undertriaged by the app. Of the 378 patients, 14 (3.7%) had received safe advice determined by the expert panel and 20 (5.3%) were considered to be potential AHS. Therefore, the assessment could be considered safe in 94.7% (358/378) of the patients when compared with the MTS assessment. From the 3 lowest MTS categories, 43.4% (164/378) of patients were not considered as emergency cases by the app, but could have been safely treated by a general practitioner or would not have required a physician consultation at all.

Conclusions: The app provided urgency advice after patient *self-triage* that has a high rate of safety, a *rate of undertriage*, and a *rate of triage with potential to be an AHS*, equivalent to telephone triage by health care professionals while still being more conservative than direct ED triage. A large proportion of patients in the ED were not considered as emergency cases, which could possibly relieve ED burden if used at home. Further research should be conducted in the at-home setting to evaluate this hypothesis.

Trial Registration: German Clinical Trial Registration DRKS00024909; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00024909

(JMIR Mhealth Uhealth 2022;10(3):e32340) doi:10.2196/32340

KEYWORDS

symptom checker; emergency medicine; app; triage; safety; innovative; eHealth; artificial intelligence

Introduction

Background

The need for acute medical care in emergency departments (EDs) and primary care clinics has become increasingly important from medical and health policy perspectives [1-3]. Partially owing to an aging population and difficulty in accessing other care options, an increasing number of patients with chronic conditions and people with general medical illnesses present in EDs [4]. More than 50% of patients attending an ED stated that they considered their level of treatment urgency as low [5], and studies have shown how challenging it is for patients to assess their own medical urgency level [6-9]. Apart from extended waiting times and patient dissatisfaction, crowded EDs are associated with several risks such as delayed care, persisting pain, poor outcomes, and increased mortality [10]. Timely assessment is increasingly a major problem in terms of staffing and organization in both large and small hospitals. An additional digital system could be of much help here. To address this, we explored whether a digital patient triage solution could provide meaningful assistance to the patient in pretriaging their current health problem, guiding patients with urgency to the ED and others to alternative appropriate care providers including urgent care centers, general practitioners (GPs), or even pharmacies or self-care.

Currently, in an international context, there is no established system of remote *pretriage* or urgency advice, which patients can use before visiting an ED, although a number of solutions have been proposed including telephone triage and video-assisted triage through health care apps [11-13].

The symptom assessment class of home-use health care apps (sometimes known as symptom checkers) [14] has the potential to provide useful information for patients on disposition (ie, the urgency of care-seeking and indicating the appropriate type of health care provider to contact) and to increase the efficiency of the medical workflow through hand over of information on symptoms, history, and risk factors. Individual apps within this class differ in their intended purpose; for example, some can only be used for a narrow range of conditions, age groups, or health care settings [15-20]. One of these symptom assessment apps (SAAs) is Ada, an app designed to be used at home, in which patients enter their risk factors and most troubling symptoms. On the basis of this information, an adaptive question flow is generated using a large medical knowledge database and complex Bayesian networks. A report lists the denied and affirmed symptoms, up to 5 suggestions on conditions including their probability, and an overall urgency assessment to provide the user with information about possible causes for their symptoms and the next steps to consider.

Although vignette studies testing SAAs have been conducted [21-24], not many studies have explored them in a prospective ED setting, which is another area of interest for such apps besides the at-home setting. Barriga et al [25] compared ED physicians' diagnoses with those from an app; however, they excluded the patient from their analysis if the physician's diagnosis was not modeled in the app's system. In addition, a retrospective study explored triage and diagnostic accuracy of 5 SAAs for patients presenting in the ED with HIV or hepatitis C [16]. A further study has examined triage acuity of a web-based SAA in a prehospital setting, but without comparing the data with a gold standard [26]. A recently published study compared the National Health Systems 111 telephone triage system with ED triage (for those patients attending the ED) and showed a high proportion of mistriaged cases [27].

Objectives

The aim of this study is to prospectively evaluate the urgency advice provided by an SAA (Ada) to examine its extensibility to the ED waiting room triage.

In an observational approach, the safety of the app's urgency advice in a large German university hospital ED is assessed by comparing the app's urgency advice levels with the assignments by a trained health care professional (HCP) using a validated triage algorithm (Manchester Triage System [MTS]). An expert physician panel evaluated all the cases of the app's undertriaged advice. We investigated the hypothesis that the urgency advice provided by the app to patients in the ED waiting room would be similar to triage by HCPs in terms of safety of advice.

Methods

Study Population, Setting, and Procedure

The study was conducted at the interdisciplinary ED of Marburg University Hospital, which is attended by approximately 48,000 patients per year, with data collection performed between August 2019 and March 2020. The completed Strengthening the Reporting of Observational Studies in Epidemiology checklist is included in (Table S1 in Multimedia Appendix 1). Written informed consent was obtained from all patients before entering the study. Sample size calculation was performed by the Coordinating Centre for Clinical Trials, Marburg.

Patients were triaged by a triage nurse, following the usual workflow and using the MTS implemented through a computerized decision support system. MTS maps the patient's presenting complaint to one of 52 flowchart diagrams. After checking the key discriminators for each of these flowcharts, the MTS groups patients into one of 5 urgency categories [28]. Each category has been assigned a maximum time in which the patient has to be examined by a physician, ranging from red (0 minutes waiting time) to blue (120 minutes waiting time in the German version of MTS). Patients grouped into the two highest triage levels, red and orange (maximum of 10 minutes waiting time), were excluded from the study because in this initial observational study, we did not consider it safe and feasible for these patients, who are not the current target population of the

app, to complete enrollment and conduct a self-assessment in their available waiting time.

All German-speaking walk-in patients aged ≥18 years attending the ED and triaged yellow, green, or blue were eligible to be included in the study. No department was excluded from the study. All patients who met these criteria were enrolled during the working hours of the assistant in charge of the study. Patients who were already called for examination before being approached or who had left the ED before being examined by medical staff were excluded from the study. The recruitment was performed by the study assistant in the waiting room after ED staff triage. After consent was obtained, patients participated in an assessment on a study iPad prepared with an adapted version of the app (study ID was used instead of a name; report was not shown to the patient after use as the study had an observational design, and to be compatible with this, to prevent information from the report from being passed on to the attending physician and potentially influencing the physicians' decision and patient outcomes). The study assistant did not offer any content-related assistance, for example, explanations of terms, but only helped with the technical operation. The assessment report was accessible to study staff only.

The patients entered factors such as sex, age, and specific risk factors (hypertension, diabetes, smoking, and pregnancy), followed by their most troubling symptoms, which were the reasons for their ED visit. The app then proceeded through an adaptive question flow, asking the essential next questions to lead to the optimal condition suggestions and urgency advice (on an 8-level scale; Table 1). Then, the patients proceeded to usual care without seeing the app output.

Table 1. App grading of urgency recommendations.

Urgency assessment level	Short description of advice level	Recommended next steps
1	Call ambulance	May require emergency care; if the patient considers this to be an emergency, calling an ambulance is advised.
2	Emergency care	May require emergency care; if the patient considers this to be an emergency, they should immediately visit an emergency department.
3	Primary care within 4 hours	May require urgent medical care; the patient is advised to see a primary care physician within the next 4 hours.
4	Primary care within same day	May require prompt medical care; the patient is advised to see a primary care physician, ideally on the same day.
5	Primary care within 2-3 days	No urgent medical care is required; the patient is advised to see a primary care physician, ideally in the next couple of days.
6	Primary care within 2-3 weeks	No urgent medical care is required; the patient is advised to see a primary care physician in a routine appointment.
7	Self-care or pharmacy	No medical consultation is needed; the patient can probably manage symptoms safely at home, and possibly, it could be helpful to consult a pharmacist.
8	Self-care	No medical consultation is needed; the patient can probably manage symptoms safely at home.

Ethics Approval

This study was approved by the Philipps University Marburg Ethics Committee for the Department of Medicine (133/18).

Study Design

This study had a single-center cross-sectional prospective observational design to evaluate the safety of urgency advice given by the app. To assess safety and identify all cases where a less conservative advice level could have the potential to harm the patient's health, we used the approach reported by Meer et al [9] via a physician panel who adjudicated on potential avoidable hazardous situations (AHSs). An AHS is defined as a health-damaging situation that is preventable through timely medical intervention.

The MTS score given by the triage nurse in the ED was compared with the advice level given by the SAA. This comparison was conducted using a predefined mapping (Table S2 in Multimedia Appendix 1), which has three categories: (1) exact match of the recommendations, (2) higher triage recommendation than MTS, and (3) lower urgency than MTS. Patients whose advice level from the SAA was higher than or matching with the MTS score were considered to have been safely triaged. For other patients whose advice level was lower than the MTS score, all case information was collected and reviewed by a panel of physicians. The panel members had no connection to the study center or study team and had a minimum of 9 years of clinical experience and different specialties: a GP and active emergency physician, a specialist in internal medicine, and the chief physician of the ED at a large hospital. The panel considered all the clinical information included in the physicians' reports and collected by the SAA. Each panel physician individually checked all the cases, assessed the urgency, and, from his point of view, selected the most appropriate advice level without being made aware of the MTS score or the app's advice level. This was later compared with the actual urgency advice provided by the app for an additional blinded comparison.

Then, each panel physician saw the MTS score and the app's advice level to adjudicate whether the app's advice would have been health-damaging if the patient had used the app at home

and followed the provided advice (categories: *unlikely, rather unlikely, rather likely,* and *likely health-damaging*). The panel members were asked to record a brief justification for each of their decisions (Table S3 in Multimedia Appendix 1). In a videoconference, the panel members discussed all the cases in which at least one of them chose the categories *likely* or *rather likely* health-damaging. A panel decision for these cases was reached by majority voting (ie, potential to be an AHS).

Data Collection and Analysis

All data entered by the participants were stored electronically and on paper and entered manually by a study staff member into a database created and managed by the Coordinating Centre for Clinical Trials in Marburg. Clinical data were obtained from the hospital information system (Dedalus ORBIS).

To determine the correspondence of the urgency assessment of the app and the triage nurse, Cohen κ coefficient [29,30], including the weighted, prevalence, and bias-adjusted Cohen κ coefficient [31], was calculated using R (version 3.6.1; R Foundation for Statistical Computing). This is a classical matching measure but designed for quadratic contingency tables. Therefore, the app's categories were combined into 3 categories according to content (Table S2 in Multimedia Appendix 1) and assigned to the 3 MTS categories, respectively. In this analysis, the MTS triage was regarded as the reference standard. To assess the advice's safety, we estimated the probability that the urgency assessment of the app was less conservative than the reference standard. We calculated Fleiss κ to assess the interrater agreement.

In addition, the chi-square test was performed to determine whether the distribution of the assessments of urgency differs between the ED medical departments [32]. Missing data were described per analysis, and participants were not excluded for missing data.

Results

Patient Selection

In this single-center cross-sectional prospective observational study, patients were enrolled between August 2019 and March 2020. A total of 544 patients were estimated to be enrolled using Cohen κ with a 2-sided 95% CI and a width of 0.1 units (p0=0.80; Cohen κ=0.70). Owing to the COVID-19 pandemic, the study had to be terminated prematurely before reaching the calculated 554 patients. Early in the pandemic, additional staff, including clinical study personnel, were not permitted to work in the ED to prevent further endangerment of patients and clinical staff.

During the study period, 1640 patients who met the eligibility criteria used the facility. Owing to staff availability, only 24.21% (397/1640) of them could be approached, and all of them agreed to participate. Of the 397 patients, 4 (1%) were excluded because informed consent was not complete and 8 (2%) were excluded because of technical problems (report was not sent from the app to the study email address). Therefore, 96.9% (385/397) of the patients were included in the study, of whom 98.2% (378/385) provided enough information to analyze the primary endpoint. For the 1.8% (7/385) excluded patients, either the MTS or the app's triage level was missing. See Figure 1 for the flowchart of patient recruitment and Table 2 for patient characteristics. The raw data are available in Multimedia Appendix 2.

Figure 1. Flowchart of patient recruitment. AHS: avoidable hazardous situation; MTS: Manchester Triage System.

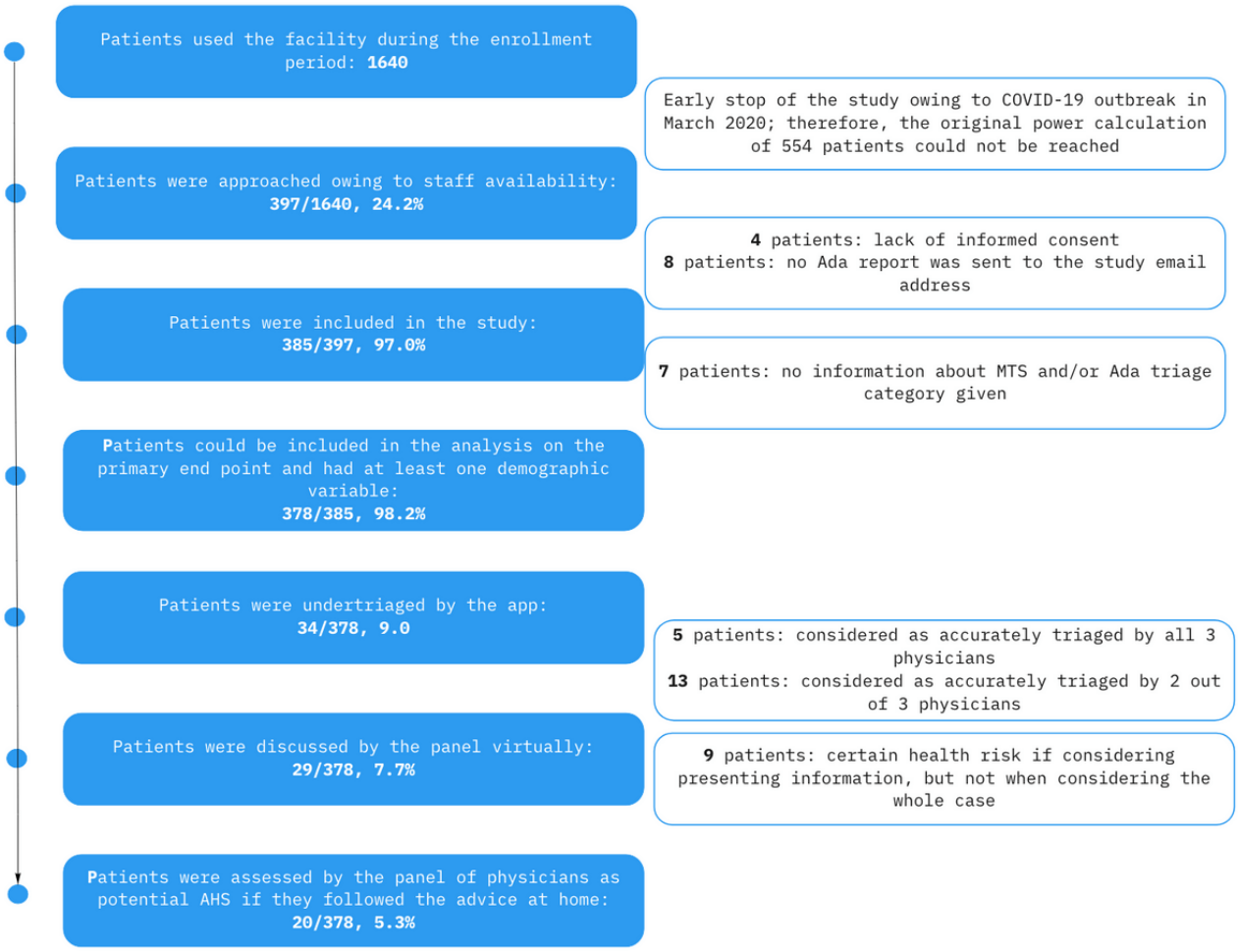


Table 2. Patient characteristics.

Characteristic	Value, n (%)
Age (years; n=377)	
18-29	93 (24.7)
30-39	59 (15.6)
40-49	58 (15.4)
50-59	77 (20.4)
60-69	55 (14.6)
70-79	28 (7.4)
80-89	6 (1.6)
90-99	1 (0.3)
Sex (n=377)	
Men	215 (57)
Women	162 (43)
Location of presenting symptom (N=378)	
Infection or feeling generally unwell	6 (1.6)
Pathological laboratory results	9 (2.4)
Paresthesia	25 (6.6)
Digestive	47 (12.4)
Chest, heart, or lungs	30 (7.9)
Face: eye, ear, nose, throat, or teeth problem	40 (10.6)
Head	44 (11.6)
Upper extremity	59 (15.6)
Lower extremity	77 (20.4)
Genitourinary problems	6 (1.6)
Neck or back	18 (4.8)
Skin	5 (1.3)
Other	5 (1.3)
Missing	7 (1.9)
Departments (N=378)	
Orthopedics and trauma	164 (43.4)
Internal medicine	102 (26.9)
Neurology	72 (19)
Other	40 (10.6)

Patient Characteristics

The mean age of participants was 46 (SD 17.54; median 46) years. In all, 43.4% (164/378) of the patients were aged ≥50 years and 57% (215/377) of the patients were men. The most common presenting symptom was extremity pain (136/378, 35.9%) followed by gastrointestinal symptoms such as abdominal pain, nausea, or change of bowel movement (47/378, 12.4%). Of the 378 participants, 44 (11.6%) participants presented at the ED with headache or vertigo. Of all participants, 43.4% (164/378) were allocated to the orthopedics and trauma department, 26.9% (102/378) to internal medicine department, and 19% (72/378) to neurology department. Totally, 10.6% (40/378) of included patients were examined at and treated by other departments. When comparing the data from this study with data from a study focusing on patient characteristics in an ED of a German university hospital over the period of a year in 2019, we could see that the mean age of their patient population was 47 (SD 24; median 47, range 0-106) years, which was similar to that reported in this study (mean 46, SD 17.54 years; range 18-94 years) [33]. This study reported fewer female patients (162/377, 43%) than the previous study (48%). When only considering the lower 3 MTS categories, the study reported 39.8% of patients classified as MTS 3, 41.4% of patients as MTS 4, and 4% of patients as MTS 5. Although that study

showed a higher proportion of patients in MTS 3 than that in this study, the low proportion of patients in MTS 5 was similar.

Results of Ada and MTS

All patients were recruited from the ED waiting room, and the app provided advice to 56.3% (213/378) of cases to seek emergency treatment, as seen in Table 3. The app advised 39.4% (149/378) of the patients to see a GP and 4.2% (16/378) of the patients to make no physician appointment at all. The triage nurse assigned 19.8% (75/378) of patients as *urgent* (MTS 3; to be examined within 30 minutes), 75.9% (287/378) of the patients as *standard* (MTS 4; up to 90 minutes waiting time), and 4.2% (16/378) of patients as *nonurgent* (MTS 5; up to 120 minutes waiting time). To determine the safety of the app's urgency assessment, the 2 systems were compared in Tables 3 and 4. Totally 91% (344/378) of patients were triaged the same or more conservatively when compared with the stand-alone MTS assessment, whereas 8.9% (34/378) of the patients were undertriaged. The chi-square test showed that the app's urgency assessments' distribution was equal in all examined departments (Cohen d=4.97; P=.05). Cohen κ calculated based on the merged comparison table showed low agreement between MTS and the app's advice level (Cohen κ=0.033, 95% CI –0.023 to 0.089; weighted Cohen κ=0.035, 95% CI –0.630 to 0.700; prevalence-adjusted and bias-adjusted Cohen κ=–0.002, 95% CI –0.056 to 0.053; Table S4 in Multimedia Appendix 1).

Of 8.9% (34/378) of the undertriaged cases, 15% (5/34) were considered to be accurately triaged by all 3 panel physicians. The panel judged that for 26% (9/34) of the participants, the app's urgency assessment could have posed a particular risk to the patient's health when only considering the information the patient presented with, but when considering the whole case retrospectively, there was no risk. Of the 9 patients, 4 (44%) were considered to have received accurate advice by at least one physician.

Table 3. Overview of the urgency assessments by the two systems (rater 1: MTS[a]; rater 2: Ada; N=378) grouped in categories.

	MTS 3 (yellow), n (%)	MTS 4 (green), n (%)	MTS 5 (blue), n (%)	Total, n (%)
Call ambulance	23 (6.1)[b]	68 (17.9)[b]	5 (1.3)[b]	96 (25.4)
Emergency care	22 (5.8)[c]	91 (24.1)[b]	4 (1.1)[b]	117 (30.9)
Primary care within 4 hours	10 (2.6)[c]	20 (5.3)[b]	2 (0.5)[b]	32 (8.5)
Primary care within same day	10 (2.6)[d]	60 (15.9)[c]	3 (0.8)[b]	73 (19.3)
Primary care within 2 to 3 days	6 (1.6)[d]	34 (8.9)[c]	2 (0.5)[c]	42 (11.1)
Primary care within 2 to 3 weeks	2 (0.5)[d]	0 (0)[d]	0 (0)[c]	2 (0.5)
Self-care or pharmacy	2 (0.5)[d]	11 (2.9)[d]	0 (0)[c]	13 (3.4)
Self-care	0 (0)[d]	3 (0.8)[d]	0 (0)[c]	3 (0.8)
Total	75 (19.8)	287 (75.9)	16 (4.2)	378 (100)

[a]MTS: Manchester Triage System.

[b]Overtriage.

[c]Match.

[d]Undertriage.

Table 4. Urgency assessment results (N=378).

Description	Value, n (%)
App's urgency assessments that matched with MTS[a]	128 (33.9)
App's urgency assessments that were overtriaged in comparison with MTS	216 (57.1)
App's urgency assessments that were undertriaged in comparison with MTS	34 (8.9)
App's urgency assessments that were undertriaged in comparison with MTS but considered accurate by all panel physicians	5 (1.3)
App's urgency assessments that were retrospectively not considered as an AHS[b]	9 (2.4)
Of the app's urgency assessments that were retrospectively not considered as an AHS, the advices considered accurate by at least one physician	4 (1.1)
App's urgency assessments that were considered as a potential AHS	20 (5.3)
Advice considered safe (all patients who were not considered to be in a potential AHS)	358 (94.7)

[a]MTS: Manchester Triage System.

[b]AHS: avoidable hazardous situation.

Describing Potential AHS

In 5.3% (20/378) of the cases, at least one physician considered the app's advice as potentially health-damaging if followed by the patient after considering all the case information. Of the 20 patients, 5 (25%) patients were admitted as inpatients, 10 (50%) patients were treated in the ED and subsequently discharged for further outpatient treatment, and 5 (25%) patients were discharged without treatment. The most common reason for attending the ED for these patients were wounds that needed to be stitched (5/20, 25%), followed by fractures (3/20, 15%), and infections (3/20, 15%). The list of potential AHS characteristics is presented in Table S5 in Multimedia Appendix 1**.** Fleiss κ, interpanel physician reliability of agreement when evaluating the likelihood of health risk, was Fleiss κ=0.0533 (95% CI –0.2267 to 0.3333), indicating slight agreement, following the interpretive guidelines reported by Fleiss et al [30].

Discussion

Principal Findings

Compared with usual hospital triage, 91% (344/378) of the participants were triaged identically or more conservatively by the app, and there was a *total undertriage* of 8.9% (34/378) of the participants, of which 59% (20/34) *were potential AHSs*. The app provided safe advice for 94.7% (358/378) of the patients when compared with the stand-alone MTS assessment, which served as the gold standard in this study. This includes identical or more conservative advice (344/378, 91%) and cases defined as safe by the physician panel (14/378, 3.7% no potential AHS). Of the 378 participants, 164 (43.4%) were not considered as emergency cases by the app.

Degree of App Undertriage

The app's rate of *undertriage* and rate of *leading to a potential AHS* are similar to those reported for telephone triage by HCPs. Placing this in context with the literature on triage, Meer et al [11] reported 4.6% (7/153) *potential AHS triage* (95% CI 1.85%-9.20%), Morreel et al [34] reported 17.01% (175/1029) undertriage for computer-assisted telephone triage, and Røtvеit et al [35] reported 10.8% (26/240) undertriage. In addition, Graversen et al [36] reported 17.7% (75/423) undertriage and specified 7.3% (31/423) *clinically relevant undertriage.* The urgency advice safety of the app is similar to or better than that reported in vignette studies of HCPs in a GP clinic setting, with 19.6% (69/352 vignette assessments) [37] and 17.1% (166/973 vignette assessments) undertriage for GP assistants [38]. In a recent vignettes study, GPs were compared with 8 SAAs, leading to a rate of undertriage of 13.74% (169/1230 vignette assessments) for GPs, with 2.92% (36/1230) of advice considered potentially unsafe [22]. This was compared with SAAs, including the Ada app, which reported a rate of 15% (30/200 vignette assessments) undertriage, with a 1.5% (3/200) rate of potentially unsafe advice (range for all SAAs 2.2%-20%).

Of the few studies reporting app-based self-assessment triage, a study reported 11.1% (14/126) undertriage [13] and another reported 5.2% (8/154) [39]. The latter was performed in a student health care center, exploring a different population with likely different presenting problems.

Examination of all collected information in this study enabled the identification of the reason for undertriage. Most commonly, the relevant condition was not modeled by the app; for example, 25% (5/20) of potential AHSs were related to non–life-threatening skin wounds, which must be examined and treated. Adapting the app to provide these scenarios would be a simple improvement. Potential AHSs also resulted from limitations in gathering information on previous injuries, a common reason for ED visits, as mild pain symptoms reported for the pre-existing injury site were not accompanied with descriptions of the injury itself, and therefore, received lower triage than appropriate after an accident. This can be resolved through an initial question about accidents. Multimorbidity also led to triage inaccuracy, as patients intermixed old and new symptoms and conditions.

Degree of App Overtriage

The number of the apps *total overtriage* compared with nurse MTS triage was 57.1% (216/378), which compares with 42.1% (101/240) for intuitive triage of patients in ED by GPs [27], 19.3% (188/973 contacts) of patients in GP clinic by triage nurses studied through vignettes [38], 23.5% (101/430) for computerized triage decision support assisted nurses [36], and 55.8% (86/154) of patients for a prototype self-assessment triage system [39]. In the binary approach (call a physician or do not call a physician) of Verzantvorrt et al [13], a rate of 11.1% (14/126 home user) overtriage was reported, which is not comparable with an 8-point classification used in this study, in which the *total overtriage* includes even the most minor overtriage. The patient populations in the studies listed above partly differ from that in this study owing to a pretriage setting, making a direct comparison moderately difficult.

The *total overtriage* in this study was relatively high, as the app advises appropriately for the home setting. Although acknowledging that very conservative advice is undesirable, the approach of app manufacturers has been to reflect a *safety-first* approach [22,24,40]. This approach can be seen in a recently published comparison study of urgency assessments of 15 SAAs with those of laypersons, stating that SAAs classified a high number of low-urgency cases as emergencies (43/174, 24.7% vignettes), whereas true emergencies were detected in 80.6% (SD 17.9%) of cases [41]. The calculated ratio of overtriage to undertriage errors for SAAs was 3.5:1, showing the strong risk aversion of those apps with a number of overtriaged vignettes of 34.2% (182/532) and a range of accurate triage from 9% to 32%. This study falls in line with the studies mentioned above, showing the difficulty of web-based triage, and although a degree of overcautiousness is appropriate for safety, a balance should be maintained.

Study Limitations and Strengths

A significant strength of this study is the variety of medical specialties included, as this provides a good representation of the average patient population in ED, in contrast to previous ED triage studies [42,43]. In addition, in contrast to vignette studies, the evaluation was directed prospectively, and patients performed the assessment by entering their data independently on their own. Moreover, unlike in previous studies [25], patients were included in the analysis of this study irrespective of

whether the app's medical knowledge modeled their diagnosed conditions.

To ensure that there was no delay in the treatment of patients with life-threatening conditions, the 2 highest categories of MTS were excluded. This also partially applies to patients who were triaged as *yellow* by MTS, who were often called by a nurse or student physician before recruitment, resulting in smaller proportion of these patients in the study.

A recognized challenge in comparison with triage methodologies is defining a reference standard [11]. Comparison of an innovative system against an established system is the most apparent validation approach; however, there are challenges. The MTS was created for patients in need of emergency treatment, who should all be examined by a physician within the same day, within a maximum waiting time of 120 minutes (German MTS) or 240 minutes (international MTS), even though not all patients who are presenting can be considered as patients with emergency [44,45]. However, the app was created for at-home use and has a broad spectrum of urgency advice gradations from *call ambulance* to *self-care*. Therefore, when creating the matching table between the MTS and the app, several app categories were equated with those of the MTS. Therefore, analysis using the widely used Cohen κ statistic could only be applied after merging the app's urgency advice categories, where a low agreement between the raters was observed partly owing to the reasons listed above. For future studies, another statistical approach to measure the agreement between 2 raters (ie, triage approaches) should be developed.

In addition, the MTS, which was the system used as a gold standard in this study, has been shown in previous studies to have deficiencies affecting its overall safety and performance [46,47]. Specifically, it has been shown that the MTS has a high tendency to undertriage (range 11%-25%) and a low sensitivity for patients with high urgency; these were data that led to questioning the safety of the system. The rates for overtriage in the systemic review ranged from 7.6% to 54%, indicating potentially unnecessary resource use [46]. As only the undertriaged cases were individually assessed by the physician panel, all other results can only be considered as safe as the MTS itself. This underlines the importance of a good assessment and matching of urgency advice in future studies.

In the interpretation of these results, it should be considered that some of the authors of this paper were also affiliated with the company that developed the app.

In addition, the reported *total overtriage* is relatively high owing to a limitation in the study methodology. This could be resolved by having the appropriateness of all the app advices assessed by a panel instead of assessing only the patients who were undertriaged. Every patient in the ED, irrespective of whether he or she required ED treatment, received an MTS level requiring ED consultation. Studies have shown that the condition of 32%-37% of patients in the ED waiting room cannot be considered as urgent [37,38]. Therefore, we specified the matching criteria in the study planning phase, such that only MTS categories 1 to 3 were considered appropriate for patients in the ED. This led to the limitation that all patients categorized as MTS 4, who were advised by the app to see a GP within 4 hours or to go to the ED, were already considered as overtriaged according to the analysis plan. This included 47.4% (179/378) of patients, many of whom were likely appropriately triaged by the app and only classified as not matching owing to a limitation of the matching design in the analysis plan. However, it is recognized that owing to study design, as per definition, all patients who were classified as patients with emergency by the app could not be undertriaged as the MTS categories 1 and 2 were excluded from the study for reasons of safety and feasibility.

Implications for Clinicians and Policy Makers and Future Research

It has been proposed that self-assessment triage apps could reduce unnecessary ED visits [18]. Of those patients who were assigned urgency advice in the lowest 3 MTS categories, only slightly more than half were considered as patients with emergency by the app, with 39.4% (149/378) of the patients being referred to a GP and 4.2% (16/378) of the patients being advised to not see a physician at all. This is a substantial number of patients who, by their own assessment, considered themselves as patients with emergency, but from what they stated in the app, possibly would have been comfortable with outpatient care. If these 43.4% (164/378) of patients could be redirected before their visit to the ED, this could lead to a substantial decrease in the number of patients in EDs.

The relatively high number of overtriaged cases in this study was, to a large degree, a result of the limitations of the study matching design, in which patients already waiting to seeing a GP in the next 4 hours were counted as overtriage. However, overtriage error of the app was detected in the context of this study. Pretriage has high potential to reduce the burden on EDs and to support the patient's decision-making regarding where and when to best seek medical care before they visit the ED. It is important that developers of SAAs address the degree of overtriage by systems, while still ensuring that their system provides safe advice.

Further research in this area is needed to measure not only appropriate and safe advice in the at-home setting but also the willingness of patients to follow this. Studies should also address in more detail the appropriateness of advice of SAAs for users who require emergency treatment. We have identified several usability optimizations and recommendations for additions and optimizations of some ED-relevant presentations. These were incorporated into Ada's product development process after reporting the study; for example, revising the advice level for ED cases that showed a high rate of overtriage.

Conclusions

This observational study addressed an underresearched SAA triage topic in the ED [18]. We showed that the app provides urgency advice after patient *self-triage* that can be considered safe in 94.7% (358/378) of assessments when compared with the stand-alone MTS assessment, has a *rate of undertriage* and a *rate of triage with potential to be an AHS* equivalent to those of telephone triage by HCPs, and still, is a more conservative approach than direct ED triage by HCPs. In all, 43.4% (164/378) of patients who considered themselves as emergency cases were

not considered so, indicating a possible relieve on EDs if the app was used at home. Continuous app optimization, followed by future research, should be conducted specifically in the at-home setting to investigate this hypothesis.

Acknowledgments

The case review panel was funded by the Ada Health GmbH, Berlin, Germany. Data analysis performed by the Coordinating Center for Clinical Trials, Marburg, Germany, was funded by Rhön Klinikum AG, Bad Neustadt, Germany. All data used for the analysis described in this manuscript have been made openly available in the linked JMIR data files.

Authors' Contributions

FC, SG, TM, JW, BB, AJ, MCH, DT, and CSB were responsible for the study conception and design. FC, SG, TM, and BB reviewed the literature; FC, SG, TM, AJ, and JW contributed to data collection; FC, SG, TM, DT, and JM analyzed and interpreted the data; and FC, SG, and TM drafted the manuscript. All authors critically reviewed the manuscript and approved the final version.

Conflicts of Interest

FC, SG, PW, JM, and MCH are employees, consultants, or company directors of Ada Health GmbH, and some of them hold stock options in the company. BB and JW were former employees of Ada Health GmbH. The Ada Health GmbH research team has received research grant funding from Fondation Botnar, Bill & Melinda Gates Foundation, and Rockefeller Foundation. PW is employed by and owns shares in Wicks Digital Health Ltd, which has received funding from Ada Health, AstraZeneca, Baillie Gifford, Bold Health, Camoni, Compass Pathways, Coronna, EIT, Happify, HealthUnlocked, Inbeeo, Kheiron Medical, Sano Genetics, Self Care Catalysts, The Learning Corp, The Wellcome Trust, VeraSci, and Woebot. Between 2018 and 2022, LT received occasional payments as a consultant for Boston Scientific and received honoraria as a speaker on symposia sponsored by UCB, Desitin, Boston Scientific, AbbVIE, Novartis, GlaxoSmithKline, Neuraxpharm, the Movement Disorders Society, und DIAPLAN. The institution at which LT works (not LT personally) received funding from Boston Scientific, the German Research Foundation, the German Ministry of Education and Research, the Otto-Loewi-Foundation, and the Deutsche Parkinson Vereinigung.

Multimedia Appendix 1
Additional study information and calculations.
[DOCX File , 111 KB - mhealth_v10i3e32340_app1.docx]

Multimedia Appendix 2
Study data.
[XLS File (Microsoft Excel File), 378 KB - mhealth_v10i3e32340_app2.xls]

References

1. Moskop JC, Sklar DP, Geiderman JM, Schears RM, Bookman KJ. Emergency department crowding, part 1 - concept, causes, and moral consequences. Ann Emerg Med 2009 May;53(5):605-611. [doi: 10.1016/j.annemergmed.2008.09.019] [Medline: 19027193]
2. Moskop JC, Sklar DP, Geiderman JM, Schears RM, Bookman KJ. Emergency department crowding, part 2 - barriers to reform and strategies to overcome them. Ann Emerg Med 2009 May;53(5):612-617. [doi: 10.1016/j.annemergmed.2008.09.024] [Medline: 19027194]
3. Pines JM, Hilton JA, Weber EJ, Alkemade AJ, Al Shabanah H, Anderson PD, et al. International perspectives on emergency department crowding. Acad Emerg Med 2011 Dec;18(12):1358-1370 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01235.x] [Medline: 22168200]
4. Colwill JM, Cultice JM, Kruse RL. Will generalist physician supply meet demands of an increasing and aging population? Health Aff (Millwood) 2008;27(3):232-241. [doi: 10.1377/hlthaff.27.3.w232] [Medline: 18445642]
5. Scherer M, Lühmann D, Kazek A, Hansen H, Schäfer I. Patients attending emergency departments. Dtsch Arztebl Int 2017 Sep 29;114(39):645-652 [FREE Full text] [doi: 10.3238/arztebl.2017.0645] [Medline: 29034865]
6. Niemann S, Meer A, Simonin C, Abel T. Medical telephone triage and patient behaviour: how do they compare? Swiss Med Wkly 2004 Mar 06;134(9-10):126-131. [doi: 10.4414/smw.2004.10276] [Medline: 15106022]
7. Meer A, Siminon C, Trapp A, Niemann S, Abel T. Einfluss der medizinischen computerassistierten Telefontriage auf das Patientenverhalten: erste Erfahrungen in der Schweiz. Schweiz Ärzteztg 2003 Oct 08;84(41):2160-2165. [doi: 10.4414/saez.2003.10001]
8. Mills B, Hill M, Buck J, Walter E, Howard K, Raisinger A, et al. What constitutes an emergency ambulance call? Austral J Paramed 2019 Mar 22;16:1-9. [doi: 10.33151/ajp.16.626]

9. Cooper AA, Humphreys KR. The uncertainty is killing me: self-triage decision making and information availability. E-J Appl Psychol 2008 Oct 14;4(1):1-6. [doi: 10.7790/ejap.v4i1.124]
10. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. J Nurs Scholarsh 2014 Mar;46(2):106-115 [FREE Full text] [doi: 10.1111/jnu.12055] [Medline: 24354886]
11. Meer A, Gwerder T, Duembgen L, Zumbrunnen N, Zimmermann H. Is computer-assisted telephone triage safe? A prospective surveillance study in walk-in patients with non-life-threatening medical conditions. Emerg Med J 2012 Feb;29(2):124-128. [doi: 10.1136/emj.2009.080614] [Medline: 20961939]
12. Cady RG, Finkelstein SM. Mixed-methods approach for measuring the impact of video telehealth on outpatient clinic triage nurse workflow. Comput Inform Nurs 2013 Sep;31(9):439-449 [FREE Full text] [doi: 10.1097/01.NCN.0000432126.99644.6c] [Medline: 24080753]
13. Verzantvoort NC, Teunis T, Verheij TJ, van der Velden AW. Self-triage for acute primary care via a smartphone application: practical, safe and efficient? PLoS One 2018;13(6):e0199284 [FREE Full text] [doi: 10.1371/journal.pone.0199284] [Medline: 29944708]
14. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
15. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a popular online symptom checker for ophthalmic diagnoses. JAMA Ophthalmol 2019 Jun 01;137(6):690-692 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]
16. Berry AC, Cash BD, Wang B, Mulekar MS, Van Haneghan AB, Yuquimpo K, et al. Online symptom checker diagnostic and triage accuracy for HIV and hepatitis C. Epidemiol Infect 2019 Jan;147:e104 [FREE Full text] [doi: 10.1017/S0950268819000268] [Medline: 30869052]
17. Berry AC, Berry NA, Wang B, Mulekar MS, Melvin A, Battiola RJ, et al. Symptom checkers versus doctors: a prospective, head-to-head comparison for cough. Clin Respir J 2020 Apr;14(4):413-415. [doi: 10.1111/crj.13135] [Medline: 31860762]
18. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
19. Yu SW, Ma A, Tsang VH, Chung LS, Leung S, Leung L. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. Hong Kong J Emerg Med 2019 Apr 16;27(4):217-222. [doi: 10.1177/1024907919842486]
20. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
21. Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, Sangar D, et al. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. Front Artif Intell 2020 Jun 27;3:543405 [FREE Full text] [doi: 10.48550/arXiv.1806.10698]
22. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
23. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
24. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
25. Barriga EM, Ferrer IP, Sánchez M, Baranera MM, Utset JM. [A new artificial intelligence tool for assessing symptoms in patients seeking emergency department care: the Mediktor application]. Emergencias 2017;29(6):391-396 [FREE Full text] [Medline: 29188913]
26. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
27. Lewis J, Stone T, Simpson R, Jacques R, O'Keeffe C, Croft S, et al. Patient compliance with NHS 111 advice: analysis of adult call and ED attendance data 2013-2017. PLoS One 2021 May 10;16(5):e0251362 [FREE Full text] [doi: 10.1371/journal.pone.0251362] [Medline: 33970946]
28. Costa JP, Nicolaidis R, Gonçalves AV, Souza EN, Blatt CR. The accuracy of the Manchester Triage System in an emergency service. Rev Gaucha Enferm 2020;41:e20190327 [FREE Full text] [doi: 10.1590/1983-1447.2020.20190327] [Medline: 33111760]
29. Cohen J. A coefficient of agreement for nominal scales. Edu Psychol Measur 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
30. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
31. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]

32. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149 [FREE Full text] [doi: 10.11613/bm.2013.018] [Medline: 23894860]
33. Michael M, Al Agha S, Böhm L, Bosse HM, Pohle AN, Schürmann J, et al. Alters- und geschlechtsbezogene Verteilung von Zuführung, Ersteinschätzung, Entlassart und Verweildauer in der zentralen Notaufnahme. Notfall Rettungsmed 2021 May 18:1-10. [doi: 10.1007/s10049-021-00895-9]
34. Morreel S, Colliers A, Remmen R, Verhoeven V, Philips H. How accurate is telephone triage in out-of-hours care? An observational trial in real patients. Acta Clin Belg 2020 Oct 30:1-6. [doi: 10.1080/17843286.2020.1839719] [Medline: 33124524]
35. Rørtveit S, Meland E, Hunskaar S. Changes of triage by GPs during the course of prehospital emergency situations in a Norwegian rural community. Scand J Trauma Resusc Emerg Med 2013 Dec 19;21(1):1-6. [doi: 10.1186/1757-7241-21-89] [Medline: 24354953]
36. Graversen DS, Christensen MB, Pedersen AF, Carlsen AH, Bro F, Christensen HC, et al. Safety, efficiency and health-related quality of telephone triage conducted by general practitioners, nurses, or physicians in out-of-hours primary care: a quasi-experimental study using the Assessment of Quality in Telephone Triage (AQTT) to assess audio-recorded telephone calls. BMC Fam Pract 2020 May 09;21(1):1-12 [FREE Full text] [doi: 10.1186/s12875-020-01122-z] [Medline: 32386511]
37. Giesen P, Ferwerda R, Tijssen R, Mokkink H, Drijver R, van den Bosch W, et al. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? Qual Saf Health Care 2007 Jun 01;16(3):181-184 [FREE Full text] [doi: 10.1136/qshc.2006.018846] [Medline: 17545343]
38. Smits M, Hanssen S, Huibers L, Giesen P. Telephone triage in general practices: a written case scenario study in the Netherlands. Scand J Prim Health Care 2016;34(1):28-36 [FREE Full text] [doi: 10.3109/02813432.2016.1144431] [Medline: 26893132]
39. Poote AE, French DP, Dale J, Powell J. A study of automated self-assessment in a primary care student health centre setting. J Telemed Telecare 2014 Apr;20(3):123-127. [doi: 10.1177/1357633X14529246] [Medline: 24643948]
40. Dunn AG. Will online symptom checkers improve health care in Australia? Med J Aust 2020 Jun;212(11):512-513. [doi: 10.5694/mja2.50621] [Medline: 32441062]
41. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]
42. Kellermann AL, Isakov AP, Parker R, Handrigan MT, Foldy S. Web-based self-triage of influenza-like illness during the 2009 H1N1 influenza pandemic. Ann Emerg Med 2010 Sep;56(3):288-294 [FREE Full text] [doi: 10.1016/j.annemergmed.2010.04.005] [Medline: 20605260]
43. Price RA, Fagbuyi D, Harris R, Hanfling D, Place F, Taylor TB, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: a cautionary tale. JAMA Pediatr 2013 Feb;167(2):112-118. [doi: 10.1001/jamapediatrics.2013.1573] [Medline: 23254373]
44. Durand A, Gentile S, Devictor B, Palazzolo S, Vignally P, Gerbeaux P, et al. ED patients: how nonurgent are they? Systematic review of the emergency medicine literature. Am J Emerg Med 2011 Mar;29(3):333-345. [doi: 10.1016/j.ajem.2010.01.003] [Medline: 20825838]
45. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care 2013 Jan;19(1):47-59 [FREE Full text] [Medline: 23379744]
46. Parenti N, Reggiani ML, Iannone P, Percudani D, Dowding D. A systematic review on the validity and reliability of an emergency department triage scale, the Manchester Triage System. Int J Nurs Stud 2014 Jul;51(7):1062-1069. [doi: 10.1016/j.ijnurstu.2014.01.013] [Medline: 24613653]
47. Slagman A, Greiner F, Searle J, Harriss L, Thompson F, Frick J, et al. Suitability of the German version of the Manchester Triage System to redirect emergency department patients to general practitioner care: a prospective cohort study. BMJ Open 2019 May 06;9(5):e024896 [FREE Full text] [doi: 10.1136/bmjopen-2018-024896] [Medline: 31064804]

Abbreviations

AHS: avoidable hazardous situation
ED: emergency department
GP: general practitioner
HCP: health care professional
MTS: Manchester Triage System
SAA: symptom assessment app

Edited by L Buis; submitted 23.07.21; peer-reviewed by F Magrabi, J Knitza, M Schmieding, J Ropero; comments to author 27.10.21; revised version received 17.12.21; accepted 18.02.22; published 28.03.22.

Please cite as:
Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, Wicks P, Wolanski J, Tutschkow D, Schade Brittinger C, Timmermann L, Jerrentrup A
Safety of Triage Self-assessment Using a Symptom Assessment App for Walk-in Patients in the Emergency Care Setting: Observational Prospective Cross-sectional Study
JMIR Mhealth Uhealth 2022;10(3):e32340
URL: https://mhealth.jmir.org/2022/3/e32340
doi:10.2196/32340
PMID:35343909

Original Paper

An Educational and Exercise Mobile Phone–Based Intervention to Elicit Electrophysiological Changes and to Improve Psychological Functioning in Adults With Nonspecific Chronic Low Back Pain (BackFit App): Nonrandomized Clinical Trial

Carolina Sitges[1], PhD; Juan L Terrasa[1], PhD; Nuria García-Dopico[2], MSc; Joan Segur-Ferrer[3], BSc; Olga Velasco-Roldán[2], PhD; Jaume Crespí-Palmer[3], BSc; Ana María González-Roldán[1], PhD; Pedro Montoya[1], PhD

[1]Research Institute of Health Sciences (IUNICS) and Balearic Islands Health Research Institute (IdISBa), Department of Psychology, University of the Balearic Islands (UIB), Palma, Spain

[2]Research Institute of Health Sciences (IUNICS) and Balearic Islands Health Research Institute (IdISBa), Department of Nursing and Physiotherapy, University of the Balearic Islands (UIB), Palma, Spain

[3]Research Institute of Health Sciences (IUNICS) and Balearic Islands Health Research Institute (IdISBa), University of the Balearic Islands (UIB), Palma, Spain

Corresponding Author:
Carolina Sitges, PhD
Research Institute of Health Sciences (IUNICS) and Balearic Islands Health Research Institute (IdISBa)
Department of Psychology
University of the Balearic Islands (UIB)
Cra. de Valldemossa km 7.5
Palma, 07122
Spain
Phone: 34 971259885
Fax: 34 971259501
Email: carol.sitges@uib.es

Abstract

Background: Concomitant psychological and cognitive impairments modulate nociceptive processing and contribute to chronic low back pain (CLBP) maintenance, poorly correlated with radiological findings. Clinical practice guidelines recommend self-management and multidisciplinary educational and exercise-based interventions. However, these recommendations are based on self-reported measurements, which lack evidence of related electrophysiological changes. Furthermore, current mobile health (mHealth) tools for self-management are of low quality and scarce evidence. Thus, it is necessary to increase knowledge on mHealth and electrophysiological changes elicited by current evidence-based interventions.

Objective: The aim of this study is to investigate changes elicited by a self-managed educational and exercise-based 4-week mHealth intervention (*BackFit app*) in electroencephalographic and electrocardiographic activity, pressure pain thresholds (PPTs), pain, disability, and psychological and cognitive functioning in CLBP versus the same intervention in a face-to-face modality.

Methods: A 2-arm parallel nonrandomized clinical trial was conducted at the University of the Balearic Islands (Palma, Spain). A total of 50 patients with nonspecific CLBP were assigned to a self-managed group (23/50, 46%; mean age 45.00, SD 9.13 years; 10/23, 43% men) or a face-to-face group (27/50, 54%; mean age 48.63, SD 7.54 years; 7/27, 26% men). The primary outcomes were electroencephalographic activity (at rest and during a modified version of the Eriksen flanker task) and heart rate variability (at rest), PPTs, and pressure pain intensity ratings. The secondary outcomes were pain, disability, psychological functioning (mood, anxiety, kinesiophobia, pain catastrophizing, and fear-avoidance beliefs), and cognitive performance (percentage of hits and reaction times).

Results: After the intervention, frequency analysis of electroencephalographic resting-state data showed increased beta-2 (16-23 Hz; 0.0020 vs 0.0024; *P*=.02) and beta-3 (23-30 Hz; 0.0013 vs 0.0018; *P*=.03) activity. In addition, source analyses revealed higher power density of beta (16-30 Hz) at the anterior cingulate cortex and alpha (8-12 Hz) at the postcentral gyrus and lower power density of delta (2-4 Hz) at the cuneus and precuneus. Both groups also improved depression (7.74 vs 5.15; *P*=.01), kinesiophobia (22.91 vs 20.87; *P*=.002), activity avoidance (14.49 vs 12.86; *P*<.001), helplessness (6.38 vs 4.74; *P*=.02), fear-avoidance beliefs (35 vs 29.11; *P*=.03), and avoidance of physical activity (12.07 vs 9.28; *P*=.01) scores, but there was an

increase in the disability score (6.08 vs 7.5; *P*=.01). No significant differences between the groups or sessions were found in heart rate variability resting-state data, electroencephalographic data from the Eriksen flanker task, PPTs, subjective ratings, or cognitive performance.

Conclusions: Both intervention modalities increased mainly beta activity at rest and improved psychological functioning. Given the limitations of our study, conclusions must be drawn carefully and further research will be needed. Nevertheless, to the best of our knowledge, this is the first study reporting electroencephalographic changes in patients with CLBP after an mHealth intervention.

Trial Registration: ClinicalTrials.gov NCT04576611; https://clinicaltrials.gov/ct2/show/NCT04576611

(JMIR Mhealth Uhealth 2022;10(3):e29171) doi:10.2196/29171

KEYWORDS

low back pain; chronic pain; mobile apps; education; exercise; brain; cognition; depression; pain threshold; mHealth; mobile phone

Introduction

Background

Low back pain (LBP) is a highly experienced symptom in the general population and the main cause of disability in industrialized countries [1]. Although its origin is usually unknown and multicausal, numerous factors such as age, sedentary lifestyle and excess weight, psychosocial factors [2], and brain changes related to pain processing [3] favor its maintenance. In addition, symptoms, pathology, and radiological findings are poorly correlated [1], and, consequently, approximately 90%-95% have a nonspecific origin [4]. Moreover, 24%-87% will be recurrent and 50%-70% will be considered chronic LBP (CLBP; symptoms experienced for >12 weeks) [1]. Therefore, effective treatments to prevent and reduce public health expenditure in care and labor concepts [1] and alleviate the symptoms of patients are needed.

Evidence-based clinical guidelines consider physical exercise a key component among the nonpharmacological interventions for patients with LBP, and education has traditionally been used as an integral part of the multidisciplinary treatment, with its importance highlighted in recent decades [5-8]. Specifically, the combination of pain neurophysiology education and therapeutic exercise has shown improvements in pain and functioning in patients with nonspecific CLBP [5-7]. Education must be adapted to individual needs to provide skills to self-manage pain coping [9], include information on the origin and nature of the impairment, and encourage patients to continue with daily life activities [10]. It could be done in person or through brochures, webpages, and mobile apps [11]. Accordingly, the so-called mobile health (mHealth) tools are presented as a cost-effective option for continuously recording type, quantity, and quality of patients' daily activities using discrete wireless sensors, providing rapid feedback to users and clinicians, supporting telerehabilitation efforts, and decreasing clinic visits [12]. Studies to date using mHealth apps have also shown moderate-quality evidence of reductions in pain and disability in patients with CLBP [13,14].

Regarding physical exercise, a systematic review showed that stretching and strengthening exercises delivered with supervision may improve pain and function, respectively, in patients with CLBP [15]. However, stability exercises seem to be more effective than general exercise and as effective as manual therapy in reducing pain and improving functionality in patients with LBP [16]. Motor control exercises further reduce pain and improve mobility compared with general exercises [17]. Moreover, the performance of hip exercises by patients with CLBP and lumbar instability is more effective than conventional therapy at reducing LBP and levels of disability [18]. Consequently, it seems that trunk stability and resistance exercises are known to be effective interventions to improve the stabilization of the spine [19], but most studies are focused on changes in pain and disability. However, other exercise - induced changes such as psychological factors (eg, reduced fear, anxiety, and catastrophizing, as well as increased pain self - efficacy), exercise - induced analgesia, and functional and structural brain adaptations need to be explored [20].

Therefore, current interventions are inadequate because they are often based on a biomedical model, sidelining the well-documented impairments in central nociceptive processing mechanisms [21]. Evidence of enhanced central sensitization to external painful stimuli is reported in patients with CLBP, manifested by increased subjective pain sensitivity and pain-related structural, functional, and metabolic brain changes, even at rest [22]. A recent review stated that chronic pain mostly changes theta and beta oscillations, particularly in the frontal brain areas [23]. These electrophysiological changes have been recently used as markers for therapeutic efficacy, showing a significant association between pain decrease and a peak theta-alpha frequency increase [24]. Likewise, heart rate variability (HRV) is also postulated as an index of how strongly top-down appraisals, mediated by brain areas (eg, amygdala and medial prefrontal cortex) shape brainstem activity that regulate the heart, providing information about the capacity of an organism to function effectively in a complex environment [25]. A meta-analysis evidenced lower parasympathetic activation in chronic pain, especially in fibromyalgia, compared with healthy controls [26]. Moreover, some studies showed a negative correlation between low-frequency beta rhythms (13-20 Hz), as an index of activity of the somatomotor cortex, and the low-frequency component (0.04-0.15 Hz) of the HRV spectrum, as an index of sympathetic activity [27].

Therefore, it is necessary to clarify the usefulness of these physiological measures in patients with CLBP and the relationship of these measures to concomitant psychological

(eg, pain beliefs, catastrophizing, and depression) and cognitive (eg, processing speed, memory, and executive function) alterations that may contribute to the mechanisms of central sensitization [28,29]. Some studies showed that structural brain abnormalities and the cognitive impact of CLBP could be reversed by effective treatments (eg, cognitive behavioral therapy and multidisciplinary pain therapy) [30,31]. However, a recent study revealed significant clinical improvements after pain neuroscience education combined with cognition-targeted motor control training in pain, disability, pressure pain thresholds (PPTs), and physical and mental health without substantial changes in brain gray matter morphologic features [6]. Accordingly, a recent systematic review stated that the effect of exercise therapy on pain and pain modulatory substances (eg, serotonin, norepinephrine, and opioids) or their effects on altering pain-related brain activity areas in patients with musculoskeletal pain remains unclear [32].

Goal of This Study

The goal of this study is to investigate whether a self-managed program based on education and exercise using a mobile app (*BackFit app*), compared with the same program in a supervised face-to-face modality, produces changes in brain activity, HRV, and pain sensitivity (as primary outcomes) and in self-reported measures of clinical pain, disability, and psychological and cognitive functioning (as secondary outcomes) among patients with nonspecific CLBP. We also explore the relationship between electrophysiological changes in pain sensibility, clinical pain, and disability data and psychological and cognitive functioning.

Methods

Study Design

This 2-arm parallel design nonrandomized clinical trial was submitted to ClinicalTrials.gov (NCT04576611). This study is also reported according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) statement [33] (Multimedia Appendix 1).

Participants and Procedure

Recruitment

A total of 59 patients with nonspecific CLBP initially participated in this study. First, participants were contacted through email or telephone using a database from a previous study [34]. In addition, information about the study was spread by institutional emailing as well as social media, posters, and leaflets at the University of the Balearic Islands and Sant Joan de Déu Hospital (Palma, Balearic Islands). Potential participants were informed about the aim and development of the study, and if they agreed to participate, they were asked about possible contraindications and exclusion criteria. If participants met the inclusion criteria, they were interviewed at the Research Institute of Health Sciences to collect preintervention data (see *Outcomes* section). Before data collection, participants were given an information sheet, and they signed the informed consent paper form to indicate agreement to participate.

Inclusion and Exclusion Criteria

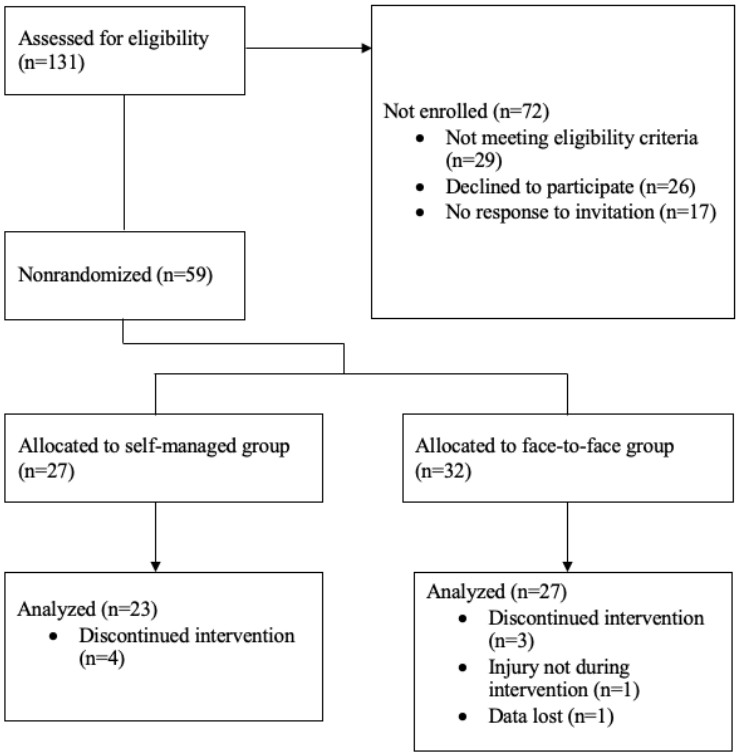
The inclusion criteria were as follows: participants aged 18-59 years with nonspecific CLBP lasting for >12 weeks, of which they have experienced at least three episodes of LBP (lasting for >1 week) [35] during the year before the study, and with access to a smartphone with internet access. The exclusion criteria [36] were as follows: high functional impairment compromising activities such as walking, sitting, or getting up from a chair; pain exacerbated by movement; presence of irradiated pain (sciatic type) or referred pain (pain perceived at a location remote from the site of origin) at lower extremities comprising sensitive or motor alterations; history of spine surgery or spinal or pelvic fracture; hospitalization for serious trauma or injuries due to traffic accidents; history of osteoarthritis in the lower extremities; and history of any systemic diseases with involvement of the locomotor system.

Sample Size, Randomization, and Blinding

The sample size was calculated using GRANMO-IMIM [37]. Accepting an α risk of .05 and a β risk of .20, assuming an estimated common SD of 2.5, and anticipating a dropout rate of 10% in a 2-sided test, 28 participants were needed in each group to recognize as statistically significant a minimum difference of 2 units (in pain intensity measured using a numerical rating scale as an indicator of the therapeutic outcome [38]) between groups, assuming that 2 groups exist.

After compliance to the treatment sessions was checked through the BackFit app, of the 59 participants, we excluded 6 (10%) from the analysis for having undergone fewer than 7 sessions, 1 (2%) because intensity of use was <10 minutes per session in more than one session, 1 (2%) because data were lost (server error), and 1 (2%) for a nonreported previous traffic injury; the remaining 50 (85%) participants were nonrandomly distributed (ie, considering their preferences to promote treatment adherence) into two groups of a 4-week educational and exercise program (total of 8 sessions of approximately 50 minutes' duration; Figure 1): (1) face-to-face group, supervised by a trained professional (with a degree in physiotherapy and science in physical activity and sport) in small groups (maximum of 4 participants) or individually (as an exception), or (2) self-managed at home using the BackFit app (version 1.0.7 for iOS and version 1.1.5 for Android). The researchers tasked with analyzing data were not involved in the intervention protocol administration, and they were also blinded to treatment allocation.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the progress of enrollment, intervention allocation, and data analysis.

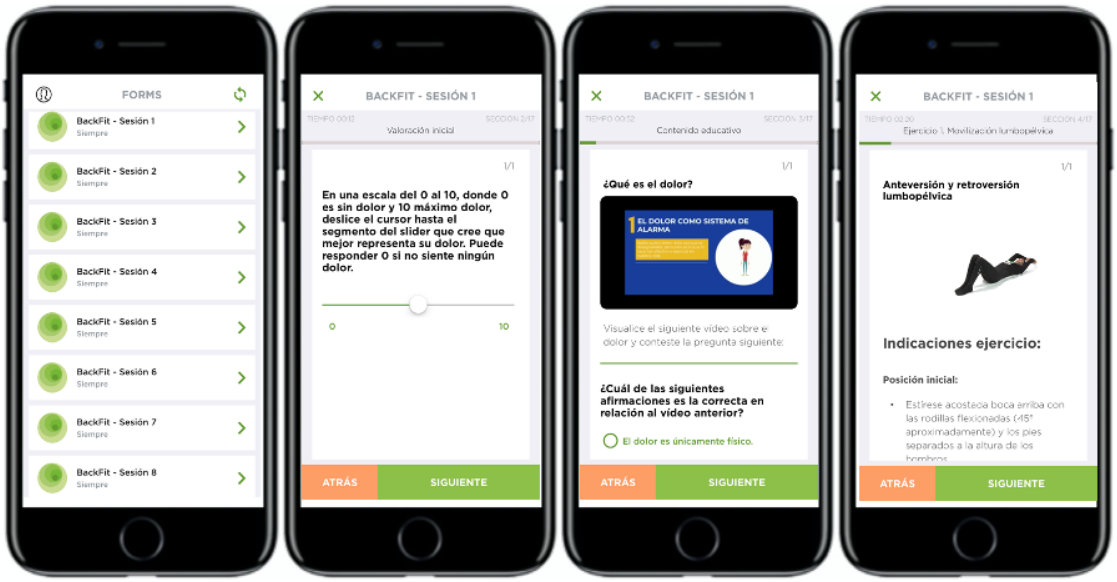


Intervention

Protocol

All participants had to perform the same intervention protocol twice a week for 4 consecutive weeks, completing up to 8 sessions. Each session consisted of the following: (1) viewing a pain education video <4 minutes in duration [39] (which included information about the neurophysiology of pain; the relationship among pain, exercise, and emotions; causes; risk factors; treatments for LBP [physical activity, cognitive behavioral therapy, and self-massage]; health habits; and self-management for chronic pain); (2) answering a question about the video to ensure that participants have watched it; and (3) performing an approximately 50-minute exercise session based on the recommendations of the American College of Sports Medicine [40], the European guidelines for the management of nonspecific CLBP [41], and some previous studies [42,43], which consisted of muscle strength exercises, motor control, relaxation routines, flexibility, and self-massage, guided by the supervisor or supported by a video showing the exercises and a detailed written description of how to perform them correctly (Figure 2). All participants also rated their actual clinical pain using a slider (0-10) before and after each session and their perceived exertion using a Borg Rating of Perceived Exertion Scale (0-10) after each exercise. The researchers provided a user account (email) and a password to all participants and helped them to configure the BackFit app on their own mobile device. All participants were informed in advance about the weekday on which the session was scheduled (and they were also reminded through notifications from the app or WhatsApp messages delivered to their mobile phone). If a participant could not perform the session on the assigned day, it was rescheduled for another day, always keeping in mind a rest period of 1-4 days between sessions. If participants were in the face-to-face intervention group, they met with the supervisor twice a week at the University of the Balearic Islands (in a room equipped for physical exercise). If participants were in the self-managed intervention group, they received the material (a rubber massage ball [60 kg/cm^2 in density and 6 cm in diameter] and a foam roller (60 kg/cm^2 in density, 90 cm in length, and 10 cm in diameter]) for use when performing the exercises at home. After the 4-week intervention, participants returned the material and met with the researchers again to enable collection of the outcome measures described in the *Sociodemographic and Clinical Data* section (after the intervention). The BackFit app had been previously tested among the researchers and by a regulatory agency (the Andalusian Agency for Healthcare Quality) [44], to acknowledge its quality and safety.

Figure 2. Screenshots from the BackFit app showing examples of the intervention protocol (ie, session, pain rating scale, educational video, and exercise).



Sociodemographic and Clinical Data

Sociodemographic and clinical data (using a semistructured interview, height and weight measuring scales, and a digital tensiometer [OMRON M3; OMRON Healthcare]), as well as clinical pain intensity ratings (using a digital slider integrated into the BackFit app) were collected.

Outcomes

All outcome measures, whether primary or secondary, were collected before and after the intervention.

Primary Outcomes

Electrophysiological Data Acquisition, Preprocessing, and Analysis

Electroencephalographic signals were continuously recorded for 5 minutes in the eyes-open resting state and during the performance of a cognitive task in an acoustically attenuated room using a QuickAmp amplifier (Brain Products GmbH) at 1000 Hz sampling rate from 29 silver or silver chloride scalp electrodes placed according to the 10-20 System of Electrode Placement. Active electrodes were recorded against an average reference. A ground electrode was located at the AFz position. An electro-oculogram channel was obtained by placing an electrode above the left eye and another below the same eye. An electrocardiogram (ECG) channel was also obtained by placing an electrode at both wrists. Electrode impedances were kept below 10 kΩ.

During electroencephalography (EEG) data preprocessing performed with BrainVision Analyzer software (version 1.05; Brain Products GmbH), signals were segmented in epochs of 1000 ms (for resting-state data) or in epochs of 600 ms (−100 to 500 ms, relative to the stimulus onset for cognitive task data) and digitally filtered (high-pass filter at 0.10 Hz, low-pass filter at 30 Hz, and notch filter at 50 Hz). We corrected eye movement artifacts using the Gratton and Coles algorithm [45]. Next, an artifact rejection protocol with the following criteria was applied: maximal allowed voltage step per sampling point=100 mV, minimal allowed amplitude=−100 mV, maximal allowed amplitude=100 mV, and maximal allowed absolute difference in the epoch=100 mV.

Regarding EEG resting-state data, frequency power densities at delta (2-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta-1 (12-16 Hz), beta-2 (16-23 Hz), and beta-3 (23-30 Hz) were computed by using the fast Fourier transformation obtained from each artifact-free EEG epoch. A source localization of the frequency bands was also performed by using low-resolution electromagnetic tomography analysis [46]. Electrode coordinates were based on an extended 10-20 system template and expressed as Talairach space coordinates. Subsequently, current source densities of all frequency bands during the resting state were estimated. To reduce interparticipant variability, spectra values were normalized at each voxel. Furthermore, a statistical nonparametric mapping randomization test was used to correct critical probability threshold values for multiple comparisons. A total of 5000 permutations were used to determine the significance of each randomization test. Subsequently, the standardized low-resolution electromagnetic tomography analysis images at each frequency band were generated by comparing the current density after the intervention with that before the intervention for all participants (paired sample 2-tailed *t* tests) and by comparing the current density in the face-to-face group with that in the self-managed group for each session separately (independent sample 2-tailed *t* tests). Voxels with significant session or group differences ($P<.05$) were located using the Montreal Neurological Institute and Hospital coordinates and Brodmann areas (BAs).

Regarding the ECG data, resting-state raw signals were offline filtered (bandpass filter 0.5-30 Hz) and hand corrected for artifacts such as missed, erroneous, or ectopic beats by using QRSTool software [47]. Next, interbeat interval values were extracted and several HRV metrics of the time and frequency domain were computed using Kubios HRV Standard software (version 3.3.1) [48]. In the time domain, mean heart rate (HR), SD of the normal-to-normal (R-R) intervals (SDNN), and the

root mean square of the successive differences (RMSSD) were calculated. In the frequency domain, the power in ms^2 of the very low frequency (VLF; 0-0.04 Hz), the low frequency (LF; 0.04-0.15 Hz), and the high frequency (HF; 0.15-0.4 Hz) were calculated. All these metrics, except for the mean HR, were transformed using a Napierian logarithm scale before the statistical analyses.

Regarding EEG registration during the cognitive task data, a nonparametric cluster-based permutation test (CBPT), which allows for testing group differences in high-dimensional neural data while it deals with the multiple-comparison problem [49], was performed by using FieldTrip toolbox [50] running in MATLAB R2018b. We used the data recorded by the 29 scalp electrodes and a time window from 0 to 500 ms after the congruent and incongruent stimulus presentations for CBPT. For every sample (electrode×millisecond), the face-to-face and self-managed groups were compared in each condition and session by means of an independent sample *t* test (2-tailed). In addition, pre- and postintervention data were compared separately for each group and in each condition by means of a dependent sample *t* test (2-tailed). Samples with *t* values higher than the critical level ($P<.05$) were selected and clustered by temporal and spatial adjacency. Next, *t* values within each cluster were summed to calculate the cluster-level statistics. These observed cluster-level statistics were evaluated through a nonparametric permutation test. The permutations were created by randomly assigning labels and running the test 1000 times, retrieving the maximum cluster statistic every time. Only if the observed cluster-level statistics from the real data were >95% of the maximum cluster statistics in the permutation distribution (Monte Carlo significance probability) were they considered significant.

Pain Sensitivity

To assess PPTs, we used a digital algometer (FPIX 50; Wagner Instruments) at an individual unilateral low back location (spinal erector muscle, 2 cm from the spine at the most painful point) and at the forefinger (control) 3 consecutive times in counterbalanced order (maximum pressure of 5 kg/cm^2). Subjective pressure pain intensity ratings were measured using a visual analog scale (0-10). The average of 3 measurements of both variables was used for the statistical analysis. Algometry was always conducted by the same researcher (OVR).

Secondary Outcomes

Self-reported Data

Handedness, physical disability, mood, anxiety, fear of movement, pain catastrophizing, and fear-avoidance beliefs were self-assessed on paper using the Spanish versions of the Edinburgh Handedness Inventory [51], the Oswestry Disability Index (ODI) [52], the Profile of Mood States (POMS) [53], the State–Trait Anxiety Inventory [54], the Tampa Scale for Kinesiophobia (TSK-11) [55], the Pain Catastrophizing Scale (PCS) [56], and the Fear-Avoidance Beliefs Questionnaire (FABQ) [57], respectively.

Cognitive Performance

A modified computerized version of the Eriksen flanker task [58], frequently and successfully used as a measure of interference control, was used. It included 288 trials, presented in 6 blocks of 48 stimuli (ie, 5 arrows) with an intertrial interval of 600-800 ms. Half of the trials were congruent (ie, the middle arrow points in the same direction as the flankers) and the other half were incongruent (ie, the middle arrow points in a direction opposite to that of the flankers). At each trial, participants were asked to indicate the direction of the middle arrow as quickly and as accurately as possible by pressing the left or right button on a 2-key device. We analyzed cognitive performance as accuracy (percentage of hits) and reaction times (RTs; in ms).

Statistical Analysis

Effects of the Intervention

To investigate the effects of the intervention and the group differences, 2-way analyses of variance with repeated measures were performed using *group* (face-to-face group and self-managed group) as the between-participant factor and *session* (before and after the intervention) as the within-participant factor in sociodemographic, clinical, and self-reported data; in PPTs and pressure pain ratings in both body locations (spinal erector muscle and forefinger); and in each HRV metric in the time (HR, SDNN, and RMSSD) and frequency domain (VLF, LF, and HF), with *condition* (congruent and incongruent) as the within-participant factor in cognitive performance (percentage of hits and RTs) and *channels* (29 electrodes) in each frequency band (delta, theta, alpha, beta-1, beta-2, and beta-3). The chi-square test was used for testing the groups' gender distribution.

We also calculated pre–post differences in each group and ran a bivariate Pearson correlation analysis only among the variables that showed significant differences in the previous analysis.

All significant results are presented with the original df, the *P* values, and the partial eta squared (η_p^2) parameters. Except for the CBPT and source localization analysis, all statistical analyses were performed using SPSS for Mac (version 25.0; IBM Corp).

Data Exclusion

In all, 10 and 13 outlier values (>3 times the IQR) were excluded from the self-reported (ODI, POMS, PCS, and TSK-11) data analysis and accuracy data analysis, respectively.

Ethics Approval

This study was conducted according to the Declaration of Helsinki and approved by the research ethics committee of the Balearic Islands (IB 3186/16 PI).

Results

Sociodemographic and Clinical Data

As shown in Table 1, both groups were comparable in terms of gender, age, anthropometrics (BMI, waist-to-height ratio, and waist-to-hip ratio), blood pressure (systolic and diastolic), pain duration, handedness, and anxiety (state and trait). Both groups were also comparable in all preintervention measures.

Table 1. Sociodemographic, clinical, and self-reported data of participants (N=50).

Characteristics	Before the intervention		After the intervention		*P* value
	Face-to-face group (n=27)	Self-managed group (n=23)	Face-to-face group (n=27)	Self-managed group (n=23)	
Sex (male), n (%)	7 (26)	10 (43)	N/A[a]	N/A	.19[b]
Age (years), mean (SD)	48.63 (7.54)	45.00 (9.13)	N/A	N/A	.13[c]
BMI, mean (SD)	0.43 (0.09)	0.41 (0.07)	N/A	N/A	.62[c]
WHtR[d], mean (SD)	0.55 (0.08)	0.53 (0.06)	N/A	N/A	.43[c]
WHR[e], mean (SD)	1.12 (0.12)	1.14 (0.11)	N/A	N/A	.60[c]
Pain duration (years), mean (SD)	11.81 (7.47)	8.06 (8.74)	N/A	N/A	.16[c]
EHI[f] (10-50), mean (SD)	18.05 (5.06)	18.22 (3.83)	N/A	N/A	.91[c]
Systolic BP[g], mean (SD)	112.67 (12.58)	115.30 (14.79)	N/A	N/A	.12[c]
Diastolic BP, mean (SD)	77.61 (8.62)	76.16 (9.74)	N/A	N/A	.92[c]
Pain intensity (0-10), mean (SD)	2.87 (2.27)	3.57 (2.50)	2.67 (2.36)	3.83 (2.20)	.93
ODI[h] (0-100, %), mean (SD)	6.15 (5.35)	6.01 (3.92)	7.85 (6.22)	7.15 (5.66)	.01[i]
POMS[j], mean (SD)					
Tension or anxiety (0-36)	9.96 (7.47)	8.83 (5.94)	7.46 (3.67)	8.00 (5.89)	.06
Anger or hostility (0-48)	11.46 (8.48)	9.11 (7.32)	7.96 (4.66)	8.78 (5.29)	.10
Vigor or activity (0-32)	15.87 (4.66)	14.28 (5.13)	16.04 (4.75)	16.33 (4.52)	.09
Fatigue or inertia (0-28)	9.38 (7.93)	10.15 (7.14)	8.03 (5.91)	8.94 (5.71)	.12
Depression or dejection (0-60)	9.26 (11.10)	6.22 (6.65)	5.12 (5.95)	5.17 (5.68)	.01[i]
Confusion or bewilderment (0-28)	6.28 (5.18)	5.00 (4.51)	5.04 (4.39)	5.06 (3.57)	.23
STAI[k], mean (SD)					
State (0-30)	15.51 (8.76)	13.75 (7.29)	15.49 (8.07)	14.96 (10.12)	.53
Trait (0-30)	19.68 (8.02)	20.15 (8.83)	N/A	N/A	.84[c]
TSK-11[l] (11-44), mean (SD)	23.14 (3.61)	22.68 (3.59)	20.41 (3.45)	21.32 (3.50)	.002[i]
Activity avoidance (7-28)	15.09 (2.27)	13.89 (2.31)	12.67 (2.21)	13.05 (2.39)	<.001[i]
Harm (4-16)	8.06 (1.69)	8.79 (1.84)	7.74 (1.68)	8.26 (1.52)	.21
PCS[m] (0-52), mean (SD)	12.21 (8.96)	15.35 (8.53)	12.29 (10.33)	11.05 (6.74)	.19
Rumination (0-16)	3.75 (3.77)	4.57 (3.20)	4.29 (3.69)	4.05 (3.39)	.98
Helplessness (0-24)	5.83 (4.36)	6.92 (4.14)	4.92 (4.60)	4.55 (3.10)	.02[i]
Magnification (0-18)	2.62 (1.64)	3.55 (2.33)	3.08 (2.60)	2.45 (1.64)	.39
FABQ[n] (0-96), mean (SD)	34.75 (24.08)	35.25 (21.33)	31.96 (19.72)	26.25 (14.54)	.03[i]
Avoidance of physical activity (0-24)	12.04 (4.93)	12.10 (6.36)	11.71 (6.52)	6.85 (4.44)	.01[i]
Avoidance of work (0-42)	15.79 (12.25)	16.90 (11.72)	13.96 (10.77)	13.65 (9.25)	.10

[a]N/A: not applicable.

[b]Chi-square test.

[c]Both groups were comparable in terms of gender, age, anthropometrics (BMI, waist-to-height ratio, and waist-to-hip ratio), systolic and diastolic blood pressure, pain duration, handedness, and anxiety trait.

[d]WHtR: waist-to-height ratio.

[e]WHR: waist-to-hip ratio.

[f]EHI: Edinburgh Handedness Inventory.

[g]BP: blood pressure.

[h]ODI: Oswestry Disability Index.

[i]Both groups showed decreased depression, kinesiophobia (and activity avoidance), helplessness, and fear-avoidance beliefs (and avoidance of physical activity), as well as increased disability after the intervention. No significant differences between the groups were found in any of these data.

[j]POMS: Profile of Mood States.

[k]STAI: State–Trait Anxiety Inventory.

[l]TSK-11: Tampa Scale for Kinesiophobia.

[m]PCS: Pain Catastrophizing Scale.

[n]FABQ: Fear-Avoidance Beliefs Questionnaire.

Primary Outcomes

EEG and ECG Resting-State Data

Regarding the frequency power density of the EEG resting-state data analysis, no differences between the groups were found at delta, theta, alpha, beta-1, beta-2, or beta-3 (results not shown). We only found main effect of *session* at beta-2 ($F_{1,47}$=5.178; P=.02; η_p^2=0.099) and at beta-3 ($F_{1,47}$=4.701; P=.03; η_p^2=0.091), showing increased beta-2 (mean 0.0020, SD 0.0013 μV^2/Hz vs mean 0.0024, SD 0.0017 μV^2/Hz) and beta-3 (mean 0.0013, SD 0.0010 μV^2/Hz vs mean 0.0018, SD 0.0019 μV^2/Hz) after the intervention in comparison with before the intervention.

Differences between before and after the intervention on statistical maps of source analyses in all participants are displayed in Table 2 and Figure 3. These analyses (paired sample 2-tailed *t* tests) revealed a significant lower current density of delta activity after the intervention compared with before the intervention in the occipital lobe areas at the cuneus (BA30 and BA18) and the middle occipital gyrus (BA18), as well as in the parietal lobe areas at the precuneus (BA7). Moreover, a significant higher current density of alpha activity after the intervention compared with before the intervention was found at the postcentral gyrus (BA2, BA3, BA5, and BA7). Finally, a significant higher current density of beta-2 and beta-3 activity after the intervention compared with before the intervention was found at the anterior cingulate cortex (ACC; BA32 and BA24) and at the medial frontal gyrus (BA10, BA9, BA8, and BA6). No significant differences between the groups before and after the intervention (independent sample 2-tailed *t* tests) were found.

Regarding ECG data, no differences between the groups or sessions were found in HR, SDNN, or RMSSD. No differences between the groups or sessions were found in VLF, LF, or HF (Multimedia Appendix 1).

Table 2. Summary of significant results[a] from whole-brain standardized low-resolution electromagnetic tomography analysis comparisons between before the intervention and after the intervention for delta, alpha, beta-2, and beta-3 frequency bands in all participants.

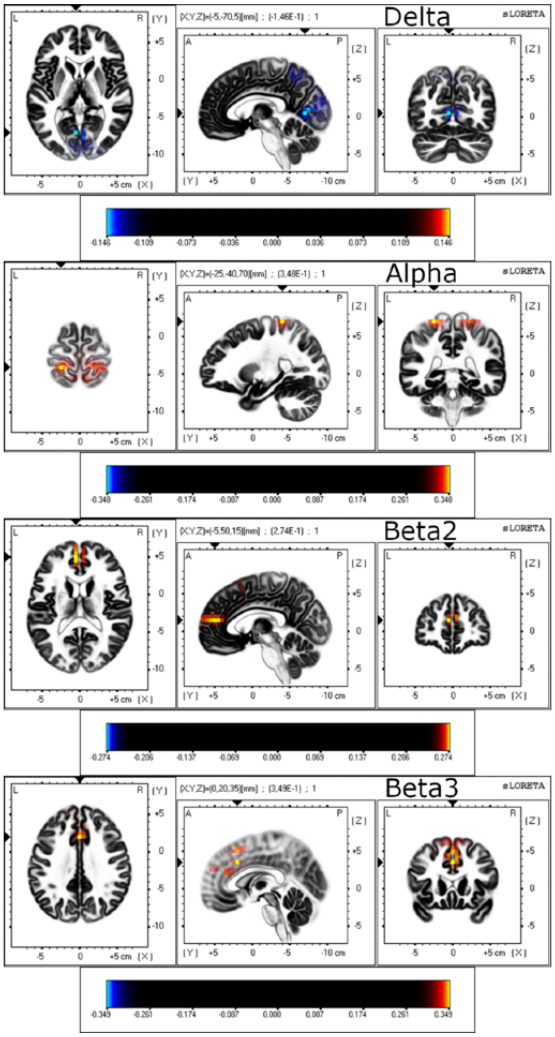
Lobe and region	BA[b]	X[c]	Y[c]	Z[c]
Delta (after the intervention<before the intervention)				
Occipital				
Cuneus	30	–5	–70	5
Cuneus	18	0	–75	10
Middle occipital gyrus	18	25	–90	15
Parietal				
Precuneus	7	0	–60	55
Alpha (after the intervention > before the intervention)				
Parietal				
Postcentral gyrus	2	–25	–40	70
Postcentral gyrus	3	–20	–40	70
Postcentral gyrus	5	–25	–45	70
Postcentral gyrus	7	5	–65	65
Beta-2 (after the intervention > before the intervention)				
Limbic				
Anterior cingulate	32	0	35	20
Frontal				
Medial frontal gyrus	10	–5	50	15
Medial frontal gyrus	9	5	50	20
Beta-3 (after the intervention > before the intervention)				
Limbic				
Anterior cingulate	32	0	20	35
Anterior cingulate	24	0	30	25
Frontal				
Medial frontal gyrus	8	0	20	50
Medial frontal gyrus	6	–5	15	50

[a]Significant (P<.05) regions are indicated with the name of Brodmann area and Montreal Neurological Institute and Hospital coordinates of the higher statistical 2-tailed threshold voxel.

[b]BA: Brodmann area.

[c]Montreal Neurological Institute and Hospital coordinates.

Figure 3. Standardized low-resolution electromagnetic tomography analysis (sLORETA) results for 3 orthogonal brain slices (horizontal, sagittal, and coronal) of delta, alpha, beta-2, and beta-3 frequency bands in all participants. Yellow-red voxels represent increased (P<.05) current density after the session compared with before the session. Blue voxels represent decreased (P<.05) current density after the session compared with before the session.



EEG Flanker Task Data

The CBPT revealed no differences between the groups or sessions in the EEG response to the congruent or incongruent conditions (results not shown).

Pain Sensitivity

No significant differences between the groups or sessions were found in PPTs either in pressure pain intensity ratings at the spinal erector muscle or the forefinger (Multimedia Appendix 1).

Secondary Outcomes

Self-reported Data

Both groups showed decreased depression, kinesiophobia (and activity avoidance), helplessness, and fear-avoidance beliefs (and physical activity avoidance), as well as increased disability after the intervention (Table 1). However, most of our participants showed a minimal disability at baseline (48/50, 96% showed a score of 0-20 measured using the ODI) and did not show a minimal clinical difference after the intervention (45/50, 91% showed a score difference between before the intervention and after the intervention of <10 points).

Cognitive Performance

The Eriksen flanker task involved a low level of difficulty (mean overall hit rate 98.93%, SD 0.14%), and no significant main effects in the percentage of hits among the groups, sessions, conditions, or interaction effects were found (results not shown). No significant main effects in RTs among the groups, sessions, or interaction effects were found (results not shown). We only found an expected significant main effect of *condition* ($F_{1,47}$=47.255; P<.001; η_p^2=0.501), showing slower RTs in the incongruent trials than in the congruent trials (mean 495.28, SD 76.11 ms vs mean 481.08, SD 84.67 ms).

Correlational Data

We only computed a bivariate Pearson correlation analysis of the pre–post differences in psychological outcomes (ODI, POMS depression and dejection scale, TSK-11 total, TSK-11 activity avoidance scale, PCS helplessness scale, FABQ total, and FABQ avoidance of physical activity scale) and EEG resting-state data (delta at the cuneus [BA30], alpha at the postcentral gyrus [BA2], and beta-2 and beta-3 at the ACC [BA32]) in all participants. After applying multiple comparison corrections, no significant correlations were found among these variables.

Discussion

Principal Findings

Both groups showed an increase in beta-2 and beta-3 in EEG resting-state data after the intervention. Source localization data analysis also showed a significant higher current density of beta-2 and beta-3 mainly located at the ACC after the intervention, as well as a higher current density of alpha mainly located at the postcentral gyrus and a significant lower current density of delta frequency located at the cuneus and precuneus. Several studies demonstrate that alpha and beta oscillations are related to feedback (top-down) brain signaling or contextual (ie, cognitive, emotional, or motivational) processing of pain [23]. Moreover, changes in brain activation and connectivity during rest in patients with chronic pain are often circumscribed to brain regions related to pain perception [22,59], which would involve the brain regions showing changes in our study (ie, postcentral gyrus and ACC). In this regard, beta band activity in somatosensory areas is increased during motor planning or during maintenance of steady posture, reflecting top-down control of behavior [60]. In contrast, the amplitude of alpha oscillations (before a phasic painful stimulation) over the sensorimotor cortex is negatively correlated with pain perception [23]. Furthermore, patients with chronic pain are characterized by a general trend toward increased power at lower EEG frequencies [61]. Indeed, delta oscillations seem to increase in states of motivational urges triggered by biological rewards and danger (eg, sustained pain) [62], and our intervention has succeeded in reducing the current density of this frequency band. Therefore, our study suggests that both intervention modalities, based on education and exercise, were able to induce neurophysiological changes, mainly in beta-2 and beta-3 frequency bands located at the ACC in patients with CLBP. However, these results should be interpreted carefully because the absence of a control group does not allow establishing a cause-and-effect relationship because some confounding variables (eg, regression to the mean) may be influencing these postintervention effects.

Nevertheless, no significant differences between the groups or sessions in HRV resting-state measures were found. Previous research stated that self-reported pain and RMSSD were inversely associated in healthy individuals but not in chronic pain, concluding that this vagal tone measure is disturbed [63]. Another study conducted on patients with CLBP showed a negative correlation between HRV and physical disability but not with pain [64]. A previous study conducted on patients with CLBP showed that a 3-month yoga intervention decreased self-reported worst pain in the past 2 weeks, LF-HRV, and rate of respiration and increased HF-HRV and PNN50 (indicating parasympathetic activity; PNN50 is the proportion of NN50 divided by the total number of normal-to-normal [R-R] intervals, and NN50 is the number of times successive heartbeat intervals exceed 50 ms) compared with standard medical care [65]. Thus, perhaps the duration of the intervention program or the intensity of the exercises was not sufficient to elicit significant changes in HRV resting-state data.

In addition, the self-managed intervention was as effective as the face-to-face intervention in improving depression, kinesiophobia (plus activity avoidance), helplessness, and fear-avoidance beliefs (plus physical activity avoidance). However, both modalities failed to reduce pain and disability and increase PPTs. In this regard, a previous study found improvements not only in health-related quality of life (mental and physical well-being), kinesiophobia, and hypervigilance, but also in pain sensitivity and disability in patients with CLBP after a 12-week intervention combining pain neuroscience education and cognition-targeted motor control training [6]. However, pain-reducing effect sizes were small to medium (ie, an increase in PPTs of >15% and a decrease in pain scores measured using a numerical rating scale) and failed to find brain morphologic changes. In contrast, we found psychological improvements accompanied by changes in EEG resting-state data, but we failed to find enhancements in pain and disability self-reported scores. Notably, our intervention was of shorter duration and the baseline pain and disability scores of our participants were clearly lower than those reported in previous studies, hindering the possibility of finding significant changes after the intervention and compromising the external validity of our study. Similarly, a 4-week program with 8 sessions, using a self-managed website (including cognitive behavioral therapy as well as motivational and wellness activity advice), evidenced clinically significant decreases in depression, anxiety, and stress, as well as greater use of positive coping strategies but no improvements in self-efficacy, self-reported pain, or physical functioning versus the control group [66]. Thus, it is possible that the duration of the intervention was not sufficient to elicit self-reported changes in pain and disability scores. Nevertheless, the presence of significant electrophysiological changes without improvements in self-reported pain and disability scores challenges the clinical relevance of our results. In addition, the inconsistences found between our research and previously reported studies highlight the need for further research in this field.

Regarding the modality of the intervention, a recent meta-analysis concluded that mHealth-based self-managed programs revealed better immediate effects on pain and disability than web-health–based programs, with better immediate effects on pain but not on disability for programs with durations of ≤8 weeks [13]. High-quality clinical practice guidelines for the noninvasive management of CLBP recommend pain education and physical exercise, considering patient preferences, with a maximum frequency and duration of 8 sessions over 12 weeks [67]. Updated evidence on rehabilitation for chronic pain showed that all exercise modalities seem to be effective compared with minimal, passive, or conservative exercise modalities or no intervention; therefore, there is no evidence indicating which duration, intensity, and training parameters are the most effective [68]. Although we tried to accommodate patient preferences by allowing them to choose the intervention modality (face-to-face group vs self-managed group with the BackFit app), the duration and intensity of our intervention seem insufficient to produce significant changes.

Furthermore, no significant differences between the groups or sessions in performance in terms of EEG activity during the Eriksen flanker task were found. Regarding cognitive performance, we found expected slower RTs in the incongruent trials, which confirmed the validity of this task to measure interference control, with greater cognitive resources needed to process stimuli in the incongruent condition. As the mean overall hit rate of this task was 98.93% (SD 0.14%), perhaps it was too effortless and not sensitive enough to observe changes produced by our educational and exercise-based intervention. Although current evidence backs cognitive improvements after aerobic exercise, we focused on a nonaerobic exercise–based intervention (including muscle strength exercises, motor control, relaxation, flexibility, and self-massage) to add novel evidence. In this regard, a previous study showed that a single session of aerobic exercise had no effect either on RTs or on brain activation in the Eriksen flanker task, but an explorative analysis revealed that RTs improved in both conditions after high-intensity exercise [69]. There is robust evidence in the literature of aerobic exercise being associated with structural and functional neuroplastic changes, partly mediated by epigenetic mechanisms, and improvements in cognitive functions and well-being [70]. It seems that in physical or metabolic training (eg, aerobic and strength), it is the intensity of training that enhances neuroplasticity (eg, reducing task-related activation of the superior and middle frontal cortex) and consequently improves cognition in a more global manner. Otherwise, in motor or neuromuscular training (eg, balance and coordination), it is the motor complexity that produces neuroplastic changes (eg, increasing activation in the inferior frontal gyrus and the superior parietal cortex, as well as in subcortical structures such as the thalamus and caudate body) and specific cognition improvements (eg, improving perceptual speed). Thus, according to current evidence, both intensity and motor complexity are important parameters to consider in the design of exercise interventions, which might have influenced our results.

Limitations

Although the results are novel and interesting, there are several limitations in the design of this study that should be considered. The main limitation was not having a passive control group to compare both interventions. As mentioned previously, clinical practice guidelines recommend accommodating patient preferences in the design of such interventions. Therefore, to promote treatment adherence, participants were not randomly distributed; as a result, a risk of selection bias must be assumed. Because of the nature of the intervention, blinding of both researchers and participants was practically unattainable; however, this is also a bias that could compromise the internal validity of the study. Because of the exclusion of the data of 15% (9/59) of the participants, the study did not reach the planned sample size to achieve an adequate statistical power. Finally, we did not control for the use of caffeine before data collection and we did not restrict the use of medications, but there were no differences between the groups (Multimedia Appendix 2).

Conclusions

Both intervention modalities (face-to-face group and self-managed group with the BackFit app) were equally effective at increasing beta activity at rest and located at the ACC, as well as at improving psychological functioning among patients with nonspecific CLBP. However, these results should be interpreted carefully because of the aforementioned limitations, which could compromise both internal and external validity of our study. The baseline pain and disability scores of our participants were clearly lower than those reported in previous studies; thus, they cannot be a representative sample of the population being studied. These limitations notwithstanding, to the best of our knowledge, this is the first study reporting brain changes in patients with CLBP after an mHealth intervention. Double-blinded randomized controlled studies with larger sample sizes are needed to increase the evidence for the efficacy of mHealth interventions in clinical practice for CLBP care. Furthermore, there is still conflicting evidence regarding the most adequate parameters for exercise prescription in chronic pain management, which must be considered in the design of novel exercise-based programs.

Acknowledgments

This research was funded by grants from the Spanish Ministry of Science and Innovation (PSI2015-66295-R, PSI2017-88388-C4-1-R–AEI/FEDER, and PID2019-110096GB-I00–AEI/10.13039/501100011033) and the Office for Transfer of Research Results of the University of the Balearic Islands. The authors acknowledge Dr Francisco García-Ginard and Dr José A Morales from Sant Joan de Déu Hospital (Palma, Balearic Islands) for their support during patient recruitment. The authors also acknowledge Pauline Bécognée, Maria Morro, Alicia Florit, and Miguel Ruiz de la Torre for their assistance during data acquisition.

Authors' Contributions

All the authors have read and approved the article and the procedures. CS, NGD, JSF, OVR, and JCP discussed the original design of the experiment and acquired the data. CS and JLT analyzed the data and drafted the original manuscript. All the authors revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nonsignificant results regarding electrocardiography data and pain sensitivity.
[DOCX File , 17 KB - mhealth_v10i3e29171_app1.docx]

Multimedia Appendix 2
Percentage of causes of onset of pain, diagnosis, and surgery or invasive treatment received, as well as medication use.
[DOCX File , 28 KB - mhealth_v10i3e29171_app2.docx]

References

1. van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;15 Suppl 2(Suppl 2):S169-S191 [FREE Full text] [doi: 10.1007/s00586-006-1071-2] [Medline: 16550447]
2. Kent P, Kjaer P. The efficacy of targeted interventions for modifiable psychosocial risk factors of persistent nonspecific low back pain - a systematic review. Man Ther 2012;17(5):385-401. [doi: 10.1016/j.math.2012.02.008] [Medline: 22421188]
3. Giesecke T, Gracely RH, Grant MA, Nachemson A, Petzke F, Williams DA, et al. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum 2004;50(2):613-623 [FREE Full text] [doi: 10.1002/art.20063] [Medline: 14872506]
4. Bardin LD, King P, Maher CG. Diagnostic triage for low back pain: a practical approach for primary care. Med J Aust 2017;206(6):268-273. [doi: 10.5694/mja16.00828] [Medline: 28359011]
5. Bodes Pardo G, Lluch Girbés E, Roussel NA, Gallego Izquierdo T, Jiménez Penick V, Pecos Martín D. Pain neurophysiology education and therapeutic exercise for patients with chronic low back pain: a single-blind randomized controlled trial. Arch Phys Med Rehabil 2018;99(2):338-347. [doi: 10.1016/j.apmr.2017.10.016] [Medline: 29138049]
6. Malfliet A, Kregel J, Coppieters I, De Pauw R, Meeus M, Roussel N, et al. Effect of pain neuroscience education combined with cognition-targeted motor control training on chronic spinal pain: a randomized clinical trial. JAMA Neurol 2018;75(7):808-817 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0492] [Medline: 29710099]
7. Galán-Martín MA, Montero-Cuadrado F, Lluch-Girbes E, Coca-López MC, Mayo-Iscar A, Cuesta-Vargas A. Pain neuroscience education and physical exercise for patients with chronic spinal pain in primary healthcare: a randomised trial protocol. BMC Musculoskelet Disord 2019;20(1):505 [FREE Full text] [doi: 10.1186/s12891-019-2889-1] [Medline: 31679512]
8. Engers A, Jellema P, Wensing M, van der Windt DA, Grol R, van Tulder MW. Individual patient education for low back pain. Cochrane Database Syst Rev 2008;2008(1):CD004057 [FREE Full text] [doi: 10.1002/14651858.CD004057.pub3] [Medline: 18254037]
9. Managing chronic pain: a booklet for patients, their families and carers. Scottish Intercollegiate Guidelines Network (SIGN). 2013. URL: https://www.sign.ac.uk/patient-and-public-involvement/patient-publications/managing-chronic-pain/ [accessed 2022-03-09]
10. Managing low back pain and sciatica - NICE Pathways. National Institute for Health and Care Excellence (NICE). 2018. URL: https://pathways.nice.org.uk/pathways/low-back-pain-and-sciatica/managing-low-back-pain-and-sciatica [accessed 2022-03-09]
11. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
12. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
13. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
14. Pfeifer AC, Uddin R, Schröder-Pfeifer P, Holl F, Swoboda W, Schiltenwolf M. Mobile application-based interventions for chronic pain patients: a systematic review and meta-analysis of effectiveness. J Clin Med 2020;9(11):3557 [FREE Full text] [doi: 10.3390/jcm9113557] [Medline: 33167300]
15. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med 2005;142(9):776-785. [doi: 10.7326/0003-4819-142-9-200505030-00014] [Medline: 15867410]
16. Gomes-Neto M, Lopes JM, Conceição CS, Araujo A, Brasileiro A, Sousa C, et al. Stabilization exercise compared to general exercises or manual therapy for the management of low back pain: a systematic review and meta-analysis. Phys Ther Sport 2017;23:136-142. [doi: 10.1016/j.ptsp.2016.08.004] [Medline: 27707631]

17. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American college of physicians clinical practice guideline. Ann Intern Med 2017;166(7):493-505 [FREE Full text] [doi: 10.7326/M16-2459] [Medline: 28192793]
18. Lee SW, Kim SY. Effects of hip exercises for chronic low-back pain patients with lumbar instability. J Phys Ther Sci 2015;27(2):345-348 [FREE Full text] [doi: 10.1589/jpts.27.345] [Medline: 25729164]
19. Hwangbo G, Lee CW, Kim SG, Kim HS. The effects of trunk stability exercise and a combined exercise program on pain, flexibility, and static balance in chronic low back pain patients. J Phys Ther Sci 2015;27(4):1153-1155 [FREE Full text] [doi: 10.1589/jpts.27.1153] [Medline: 25995577]
20. Booth J, Moseley GL, Schiltenwolf M, Cashin A, Davies M, Hübscher M. Exercise for chronic musculoskeletal pain: a biopsychosocial approach. Musculoskeletal Care 2017;15(4):413-421. [doi: 10.1002/msc.1191] [Medline: 28371175]
21. Nijs J, Clark J, Malfliet A, Ickmans K, Voogt L, Don S, et al. In the spine or in the brain? Recent advances in pain neuroscience applied in the intervention for low back pain. Clin Exp Rheumatol 2017;35 Suppl 107(5):S108-S115 [FREE Full text] [Medline: 28967357]
22. Zhang L, Zhou L, Ren Q, Mokhtari T, Wan L, Zhou X, et al. Evaluating cortical alterations in patients with chronic back pain using neuroimaging techniques: recent advances and perspectives. Front Psychol 2019;10:2527 [FREE Full text] [doi: 10.3389/fpsyg.2019.02527] [Medline: 31798496]
23. Ploner M, Sorg C, Gross J. Brain rhythms of pain. Trends Cogn Sci 2017;21(2):100-110 [FREE Full text] [doi: 10.1016/j.tics.2016.12.001] [Medline: 28025007]
24. Ngernyam N, Jensen MP, Arayawichanon P, Auvichayapat N, Tiamkao S, Janjarasjitt S, et al. The effects of transcranial direct current stimulation in patients with neuropathic pain from spinal cord injury. Clin Neurophysiol 2015;126(2):382-390. [doi: 10.1016/j.clinph.2014.05.034] [Medline: 25027640]
25. Thayer JF, Ahs F, Fredrikson M, Sollers 3rd JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
26. Tracy LM, Ioannou L, Baker KS, Gibson SJ, Georgiou-Karistianis N, Giummarra MJ. Meta-analytic evidence for decreased heart rate variability in chronic pain implicating parasympathetic nervous system dysregulation. Pain 2016;157(1):7-29. [doi: 10.1097/j.pain.0000000000000360] [Medline: 26431423]
27. Triggiani AI, Valenzano A, Del Percio C, Marzano N, Soricelli A, Petito A, et al. Resting state Rolandic mu rhythms are related to activity of sympathetic component of autonomic nervous system in healthy humans. Int J Psychophysiol 2016;103:79-87. [doi: 10.1016/j.ijpsycho.2015.02.009] [Medline: 25660308]
28. Roussel NA, Nijs J, Meeus M, Mylius V, Fayt C, Oostendorp R. Central sensitization and altered central pain processing in chronic low back pain: fact or myth? Clin J Pain 2013;29(7):625-638. [doi: 10.1097/AJP.0b013e31826f9a71] [Medline: 23739534]
29. Higgins DM, Martin AM, Baker DG, Vasterling JJ, Risbrough V. The relationship between chronic pain and neurocognitive function: a systematic review. Clin J Pain 2018;34(3):262-275 [FREE Full text] [doi: 10.1097/AJP.0000000000000536] [Medline: 28719507]
30. Seminowicz DA, Shpaner M, Keaser ML, Krauthamer GM, Mantegna J, Dumas JA, et al. Cognitive-behavioral therapy increases prefrontal cortex gray matter in patients with chronic pain. J Pain 2013;14(12):1573-1584 [FREE Full text] [doi: 10.1016/j.jpain.2013.07.020] [Medline: 24135432]
31. Schiltenwolf M, Akbar M, Neubauer E, Gantz S, Flor H, Hug A, et al. The cognitive impact of chronic low back pain: positive effect of multidisciplinary pain therapy. Scand J Pain 2017;17:273-278. [doi: 10.1016/j.sjpain.2017.07.019] [Medline: 28993113]
32. Nijs J, Kosek E, Van Oosterwijck J, Meeus M. Dysfunctional endogenous analgesia during exercise in patients with chronic pain: to exercise or not to exercise? Pain Physician 2012;15(3 Suppl):S205-S213 [FREE Full text] [Medline: 22786458]
33. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
34. Sitges C, Velasco-Roldán O, Crespí J, García-Dopico N, Segur-Ferrer J, González-Roldán AM, et al. Acute effects of a brief physical exercise intervention on somatosensory perception, lumbar strength, and flexibility in patients with nonspecific chronic low-back pain. J Pain Res 2021;14:487-500 [FREE Full text] [doi: 10.2147/JPR.S274134] [Medline: 33633462]
35. Nourbakhsh MR, Arab AM. Relationship between mechanical factors and incidence of low back pain. J Orthop Sports Phys Ther 2002;32(9):447-460. [doi: 10.2519/jospt.2002.32.9.447] [Medline: 12322811]
36. Amundsen PA, Evans DW, Rajendran D, Bright P, Bjørkli T, Eldridge S, et al. Inclusion and exclusion criteria used in non-specific low back pain trials: a review of randomised controlled trials published between 2006 and 2012. BMC Musculoskelet Disord 2018;19(1):113 [FREE Full text] [doi: 10.1186/s12891-018-2034-6] [Medline: 29650015]
37. Marrugat J. Calculadora de grandària mostral GRANMO. Version 7.12. Institut Municipal d'Investigació Mèdica. 2012. URL: https://www.imim.cat/ofertadeserveis/software-public/granmo/ [accessed 2022-03-09]

38. Suzuki H, Aono S, Inoue S, Imajo Y, Nishida N, Funaba M, et al. Clinically significant changes in pain along the Pain Intensity Numerical Rating Scale in patients with chronic low back pain. PLoS One 2020;15(3):e0229228 [FREE Full text] [doi: 10.1371/journal.pone.0229228] [Medline: 32126108]
39. Butler DS, Moseley GL. Explain pain. Adelaide, Australia: Noigroup Publications; 2003.
40. American College of Sports Medicine. In: Bayles MP, Swank AM, editors. ACSM's guidelines for exercise testing and prescription. 10th edition. Philadelphia, PA: Wolters Kluver; 2018.
41. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, COST B13 Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006;15 Suppl 2(Suppl 2):S192-S300 [FREE Full text] [doi: 10.1007/s00586-006-1072-1] [Medline: 16550448]
42. McGill S. Low back disorders: evidence-based prevention and rehabilitation. Champaign, IL: Human Kinetics; 2015.
43. Hodges PW. Core stability exercise in chronic low back pain. Orthop Clin North Am 2003;34(2):245-254. [doi: 10.1016/s0030-5898(03)00003-8] [Medline: 12914264]
44. BackFit App. Andalusian Agency for Healthcare Quality. 2020. URL: http://www.calidadappsalud.com/distintivo/info?app=Backfit [accessed 2022-03-09]
45. Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 1983;55(4):468-484. [doi: 10.1016/0013-4694(83)90135-9] [Medline: 6187540]
46. Pascual-Marqui RD, Lehmann D, Koukkou M, Kochi K, Anderer P, Saletu B, et al. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. Philos Trans A Math Phys Eng Sci 2011;369(1952):3768-3784. [doi: 10.1098/rsta.2011.0081] [Medline: 21893527]
47. Allen JJ, Chambers AS, Towers DN. The many metrics of cardiac chronotropy: a pragmatic primer and a brief comparison of metrics. Biol Psychol 2007;74(2):243-262. [doi: 10.1016/j.biopsycho.2006.08.005] [Medline: 17070982]
48. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]
49. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 2007;164(1):177-190. [doi: 10.1016/j.jneumeth.2007.03.024] [Medline: 17517438]
50. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011;2011:156869 [FREE Full text] [doi: 10.1155/2011/156869] [Medline: 21253357]
51. Albayay J, Villarroel-Gruner P, Bascour-Sandoval C, Parma V, Gálvez-García G. Psychometric properties of the Spanish version of the Edinburgh Handedness Inventory in a sample of Chilean undergraduates. Brain Cogn 2019;137:103618. [doi: 10.1016/j.bandc.2019.103618] [Medline: 31629000]
52. Alcántara-Bumbiedro S, Flórez-García MT, Echávarri-Pérez C, García-Pérez F. Escala de incapacidad por dolor lumbar de Oswestry. Rehabilitación 2006;40(3):150-158. [doi: 10.1016/S0048-7120(06)74881-2]
53. Andrade Fernández EM, Arce C, Seoane G. Adaptación al español del cuestionario 'Perfil de los Estados de Ánimo' en una muestra de deportistas. Psicothema 2002;14(4):708-713.
54. Buela-Casal G, Guillén-Riquelme A, Seisdedos Cubero N. Cuestionario de ansiedad estado-rasgo: adaptación española. Madrid, Spain: TEA Ediciones; 2011.
55. Gómez-Pérez L, López-Martínez AE, Ruiz-Párraga GT. Psychometric properties of the Spanish version of the Tampa scale for Kinesiophobia (TSK). J Pain 2011;12(4):425-435. [doi: 10.1016/j.jpain.2010.08.004] [Medline: 20926355]
56. García Campayo J, Rodero B, Alda M, Sobradiel N, Montero J, Moreno S. Validación de la versión española de la escala de la catastrofización ante el dolor. Med Clin (Barc) 2008;131(13):487-492. [doi: 10.1157/13127277] [Medline: 19007576]
57. Kovacs FM, Muriel A, Medina JM, Abraira V, Sánchez MD, Jaúregui JO, Spanish Back Pain Research Network. Psychometric characteristics of the Spanish version of the FAB questionnaire. Spine (Phila Pa 1976) 2006;31(1):104-110. [doi: 10.1097/01.brs.0000193912.36742.4f] [Medline: 16395186]
58. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 1974;16(1):143-149. [doi: 10.3758/bf03203267]
59. Kregel J, Meeus M, Malfliet A, Dolphens M, Danneels L, Nijs J, et al. Structural and functional brain abnormalities in chronic low back pain: a systematic review. Semin Arthritis Rheum 2015;45(2):229-237. [doi: 10.1016/j.semarthrit.2015.05.002] [Medline: 26092329]
60. Kim JA, Davis KD. Neural oscillations: understanding a neural code of pain. Neuroscientist 2021;27(5):544-570. [doi: 10.1177/1073858420958629] [Medline: 32981457]
61. Pinheiro ES, de Queirós FC, Montoya P, Santos CL, do Nascimento MA, Ito CH, et al. Electroencephalographic patterns in chronic pain: a systematic review of the literature. PLoS One 2016;11(2):e0149085 [FREE Full text] [doi: 10.1371/journal.pone.0149085] [Medline: 26914356]
62. Knyazev GG. EEG delta oscillations as a correlate of basic homeostatic and motivational processes. Neurosci Biobehav Rev 2012;36(1):677-695. [doi: 10.1016/j.neubiorev.2011.10.002] [Medline: 22020231]
63. Koenig J, Loerbroks A, Jarczok MN, Fischer JE, Thayer JF. Chronic pain and heart rate variability in a cross-sectional occupational sample: evidence for impaired vagal control. Clin J Pain 2016;32(3):218-225. [doi: 10.1097/AJP.0000000000000242] [Medline: 25924095]

64. Gockel M, Lindholm H, Niemistö L, Hurri H. Perceived disability but not pain is connected with autonomic nervous function among patients with chronic low back pain. J Rehabil Med 2008;40(5):355-358 [FREE Full text] [doi: 10.2340/16501977-0172] [Medline: 18461260]
65. Telles S, Sharma SK, Gupta RK, Bhardwaj AK, Balkrishna A. Heart rate variability in chronic low back pain patients randomized to yoga or standard care. BMC Complement Altern Med 2016;16(1):279 [FREE Full text] [doi: 10.1186/s12906-016-1271-1] [Medline: 27514611]
66. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
67. Wong JJ, Côté P, Sutton DA, Randhawa K, Yu H, Varatharajan S, et al. Clinical practice guidelines for the noninvasive management of low back pain: a systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. Eur J Pain 2017;21(2):201-216. [doi: 10.1002/ejp.931] [Medline: 27712027]
68. Malfliet A, Ickmans K, Huysmans E, Coppieters I, Willaert W, Van Bogaert W, et al. Best evidence rehabilitation for chronic pain part 3: low back pain. J Clin Med 2019;8(7):1063 [FREE Full text] [doi: 10.3390/jcm8071063] [Medline: 31331087]
69. Mehren A, Diaz Luque C, Brandes M, Lam AP, Thiel CM, Philipsen A, et al. Intensity-dependent effects of acute exercise on executive function. Neural Plast 2019;2019:8608317 [FREE Full text] [doi: 10.1155/2019/8608317] [Medline: 31281346]
70. Mandolesi L, Polverino A, Montuori S, Foti F, Ferraioli G, Sorrentino P, et al. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. Front Psychol 2018;9:509 [FREE Full text] [doi: 10.3389/fpsyg.2018.00509] [Medline: 29755380]

Abbreviations

ACC: anterior cingulate cortex
BA: Brodmann area
CBPT: cluster-based permutation test
CLBP: chronic low back pain
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ECG: electrocardiogram
EEG: electroencephalography
FABQ: Fear-Avoidance Beliefs Questionnaire
HF: high frequency
HR: heart rate
HRV: heart rate variability
LBP: low back pain
LF: low frequency
mHealth: mobile health
ODI: Oswestry Disability Index
PCS: Pain Catastrophizing Scale
POMS: Profile of Mood States
PPT: pressure pain threshold
RMSSD: root mean square of the successive differences
RT: reaction time
SDNN: SD of the normal-to-normal (R-R) intervals
TSK-11: Tampa Scale for Kinesiophobia
VLF: very low frequency

Edited by L Buis; submitted 29.03.21; peer-reviewed by A Cashin, E Sadeghi-Demneh, A Teles; comments to author 09.06.21; revised version received 28.07.21; accepted 07.01.22; published 15.03.22.

Please cite as:
Sitges C, Terrasa JL, García-Dopico N, Segur-Ferrer J, Velasco-Roldán O, Crespí-Palmer J, González-Roldán AM, Montoya P
An Educational and Exercise Mobile Phone–Based Intervention to Elicit Electrophysiological Changes and to Improve Psychological Functioning in Adults With Nonspecific Chronic Low Back Pain (BackFit App): Nonrandomized Clinical Trial
JMIR Mhealth Uhealth 2022;10(3):e29171
URL: https://mhealth.jmir.org/2022/3/e29171
doi:10.2196/29171
PMID:35289758

Original Paper

Digital Contact Tracing Apps for COVID-19: Development of a Citizen-Centered Evaluation Framework

Damyanka Tsvyatkova[1], BSc, PhD; Jim Buckley[1,2], BSc, PhD; Sarah Beecham[1], PhD; Muslim Chochlov[1], PhD; Ian R O'Keeffe[1], BA, BAI, MSc, PhD; Abdul Razzaq[1], PhD; Kaavya Rekanar[1], BTech, MSc; Ita Richardson[1,2], BSc, MSc, PhD; Thomas Welsh[1], PhD; Cristiano Storni[1,2,3], PhD; COVIGILANT Group[4]

[1]Lero, Science Foundation Ireland Research Centre for Software, University of Limerick, Limerick, Ireland

[2]Department of Computer Science and Information Systems, University of Limerick, Limerick, Ireland

[3]Interaction Design Centre, University of Limerick, Limerick, Ireland

[4]see Acknowledgments for contributors

Corresponding Author:
Cristiano Storni, PhD
Department of Computer Science and Information Systems
University of Limerick
Computer Science Building
Plassey Park Rd
Limerick, V94 T9PX
Ireland
Phone: 353 61 202632
Email: Cristiano.Storni@ul.ie

Abstract

Background: The silent transmission of COVID-19 has led to an exponential growth of fatal infections. With over 4 million deaths worldwide, the need to control and stem transmission has never been more critical. New COVID-19 vaccines offer hope. However, administration timelines, long-term protection, and effectiveness against potential variants are still unknown. In this context, contact tracing and digital contact tracing apps (CTAs) continue to offer a mechanism to help contain transmission, keep people safe, and help kickstart economies. However, CTAs must address a wide range of often conflicting concerns, which make their development/evolution complex. For example, the app must preserve citizens' privacy while gleaning their close contacts and as much epidemiological information as possible.

Objective: In this study, we derived a compare-and-contrast evaluative framework for CTAs that integrates and expands upon existing works in this domain, with a particular focus on citizen adoption; we call this framework the Citizen-Focused Compare-and-Contrast Evaluation Framework (C^3EF) for CTAs.

Methods: The framework was derived using an iterative approach. First, we reviewed the literature on CTAs and mobile health app evaluations, from which we derived a preliminary set of attributes and organizing pillars. These attributes and the probing questions that we formulated were iteratively validated, augmented, and refined by applying the provisional framework against a selection of CTAs. Each framework pillar was then subjected to internal cross-team scrutiny, where domain experts cross-checked sufficiency, relevancy, specificity, and nonredundancy of the attributes, and their organization in pillars. The consolidated framework was further validated on the selected CTAs to create a finalized version of C^3EF for CTAs, which we offer in this paper.

Results: The final framework presents seven pillars exploring issues related to CTA design, adoption, and use: (General) Characteristics, Usability, Data Protection, Effectiveness, Transparency, Technical Performance, and Citizen Autonomy. The pillars encompass attributes, subattributes, and a set of illustrative questions (with associated example answers) to support app design, evaluation, and evolution. An online version of the framework has been made available to developers, health authorities, and others interested in assessing CTAs.

Conclusions: Our CTA framework provides a holistic compare-and-contrast tool that supports the work of decision-makers in the development and evolution of CTAs for citizens. This framework supports reflection on design decisions to better understand and optimize the design compromises in play when evolving current CTAs for increased public adoption. We intend this framework to serve as a foundation for other researchers to build on and extend as the technology matures and new CTAs become available.

(JMIR Mhealth Uhealth 2022;10(3):e30691) doi:10.2196/30691

KEYWORDS

COVID-19; mHealth; digital contact tracing apps; framework; evaluation; mobile health; health apps; digital health; contact tracing

Introduction

The global coronavirus pandemic (COVID-19) calls for rapid measures to monitor and control the spread of the virus. Contact tracing is one of the measures adopted by health authorities. This approach has already been used with a certain level of success for other dangerous illnesses such as tuberculosis [1] and Ebola [2]. As part of the contact tracing effort in the COVID-19 pandemic, the deployment of mobile apps, and their potential in collecting, storing, and sharing citizens' contact tracing data have been examined, with early studies showing favorable results [3,4]. These studies have contributed to the impetus for using digital contact tracing apps (CTAs), and many CTAs have been developed for nations' use to facilitate community-based disease surveillance [5].

The application of CTAs in real-world settings has provoked numerous discussions regarding their design [6-9], concerns about the security and privacy of CTA data, and the barriers for their widespread acceptance and adoption by citizens [10,11]. Reflecting these discussions, CTA evaluation frameworks have emerged that specifically focus on different aspects such as the assessment of contact tracing architectures [12], sociotechnical issues [13], privacy [14,15], ethical and legal challenges [16], feasibility and effectiveness [17], usability [18], and essential attributes [19,20].

In the context of such fragmentation, a legitimate concern is for a more comprehensive evaluation framework that would encompass a variety of different aspects of CTAs, pertinent to the adopting citizens, and which would enable decision-makers (eg, developers, health authorities) to assess and possibly improve their designs. This concern drives our research question: *how to devise and organize a framework to enable a more comprehensive assessment of current CTAs, supporting the work of decision-makers (eg, developers, health authorities) in the development and evolution of CTAs, potentially increasing adoption?* In this paper, we address this question by proposing a Citizen-Focused Compare-and-Contrast Evaluation Framework for CTAs (C^3EF), which we derived by holistically bringing together existing works on the evaluation of CTAs and mobile health (mHealth) apps, and iteratively grounding and stress-testing our derivations with a number of current CTAs.

The framework proposed here is focused more on the apps themselves than on the apps' embedding in national health systems (as another important perspective [21]). The framework is organized to help in the assessment and improvement of existing CTA solutions through a taxonomy of 7 pillars that focus on clustered attributes: (General) Characteristics, Usability, Data Protection, Effectiveness, Transparency, Technical Performance, and Citizen Autonomy. This article introduces the C^3EF and presents its derivation over several iterations. As we present the framework, we will illustrate the framework's application to a selection of existing CTAs, showing how the framework can be used to assess and possibly improve aspects of CTA design.

The next section presents an overview of how we developed our framework. This is followed by an overview and discussion of the C^3EF framework itself. Our Discussion presents a summary of our contributions, limitations of this study, and future work.

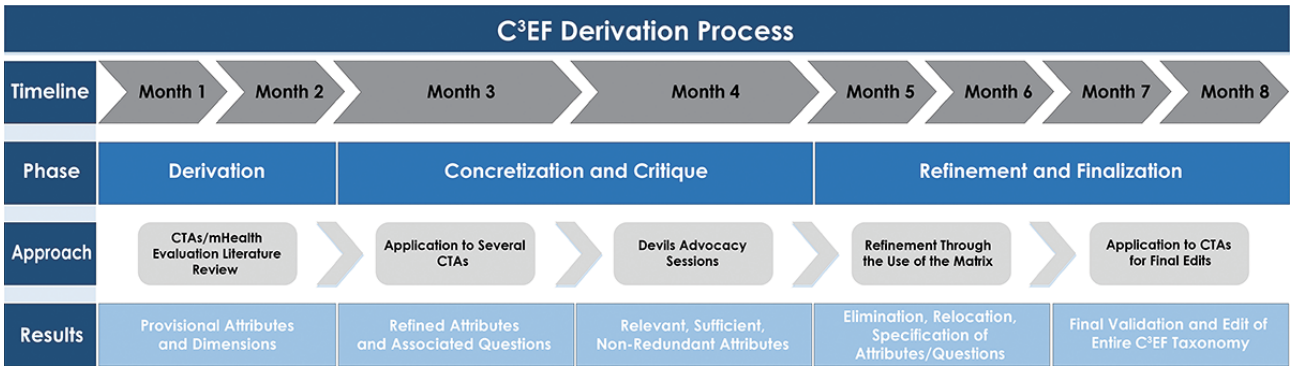
Methods

Review and Framework Derivation

To accommodate the high-complexity and multidisciplinary nature of CTAs evaluation for wide societal adoption, we used an iterative approach combining a literature review and expert opinions with an empirical application of the derived attributes, and their associated questions, to the evaluation of actual CTAs. Within this approach, the multidisciplinary nature of the framework was specifically handled by a progressive "segmentation," where 10 domain experts within the research team were allocated responsibility for individual parts of the framework.

This methodology is portrayed in Figure 1, which shows the three main phases over the period of 8 months, starting in June 2020. Phase 1 focused on initial derivation (for months 1 and 2), phase 2 focused on concretization and critique of the prototype framework derived in phase 1 (months 3 and 4), and phase 3 focused on final refinements (months 5 to 8).

Figure 1. Phases and deliverables in the development of our Citizen-Focused Compare-and-Contrast Evaluation Framework (C^3EF) for contact tracing apps (CTAs). mHealth: mobile health.



Phase 1: Derivation

Literature Review

This phase was based on a "critical (literature) review" [22] of relevant areas. Here, iterative refinement/evaluation is used to focus in on more optimal search parameters, search databases, the search string, and inclusion/exclusion criteria based on the initial research question. Ultimately, to obtain a holistic perspective, this resulted in us focusing on:

1. CTAs: these included peer-reviewed literature on existing evaluation frameworks, grey publications discussing design characteristics and functionality, and design guidelines/EU regulations for development of effective/appropriate digital CTAs.
2. Mobile apps: this included the most recent evaluation frameworks for mobile apps, evaluation frameworks for mHealth apps, accessibility principles for mobile apps, universal design (UD) for the apps, and taxonomies of usability.

To reflect this broad focus, the search sources employed were:

1. Electronic databases to search academic texts: Google Scholar, Elsevier, ACM Digital Library, Sage, IEEE Xplore, and Springer.
2. Searches of web-based grey literature (using Google).
3. Consulting the reference lists of the selected articles to identify further relevant studies, following the systematic "backward snowballing" protocol proposed by Wohlin et al [23]. This allowed us to use the original sources to recursively increase our existing set of articles. "Forward snowballing" [23] was not used, based on the relatively recent appearance of CTA-specific literature.

The search string derived from the critical review was "evaluation frameworks" AND "digital contact tracing applications" AND "COVID-19" OR "mobile applications" OR "mHealth applications" OR "accessibility" OR "universal design" OR "usability" OR "taxonomies" OR "Data protection" OR "GDPR" OR "security threats." Articles written in English and published between 2010 and 2020 were reviewed. Articles offering evaluation frameworks were selected as well as articles discussing particular aspects, qualities, or characteristics of CTAs/mHealth apps. Inclusion was assessed by reading the abstract, and, in cases where the abstract was insufficient, by reading their introduction and conclusion. All 10 researchers from the team were involved in the search and selection of the sources, with marginal papers being discussed for relevance in dedicated group meetings.

With this search strategy, we identified 44 relevant sources (a full list is available in Multimedia Appendix 1 [6,7,9,13-17,24-59]). Twenty-one of these were distinct frameworks focusing on particular aspects of CTAs, 13 provided regulations and guidelines for the design or evaluation of CTAs or mHealth apps, and 10 others described important characteristics for CTAs. From these sources, one of the researchers extracted an initial set of 111 attributes, representing a pool of attributes to be used for the derivation of the first iteration of the framework. Again, these were reviewed in a group meeting, where more marginal attributes were debated, but all ultimately persisted.

We performed a cluster analysis of this initial attribute list, which was aimed at identifying overlaps and affinities, and at grouping them into thematic areas. That is, we focused on constructing an "information architecture" of categories, but not a "navigation structure" between categories, as described by Righi et al [60]. We then juxtaposed our identified areas with those explicitly provided in the papers directed at frameworks and taxonomies for CTAs [13,14,17,24-31,58,59], and we ended up identifying 6 evaluation areas, which we call pillars: *Usability*, *Data Protection*, *Effectiveness*, *Transparency*, *Technical Performance*, and the degree of *Autonomy* the app provides to downloading citizens. To uniquely identify an app and report on its nonevaluative characteristics, "*General Characteristics*" was also added. Multimedia Appendix 1 offers a full list of the included papers, a table of the extracted attributes and categories from the selected papers, and how we grouped them into our resulting 6 pillars. At this stage, the project team was divided into domain expert subgroups, one for each pillar, working on their specific development and further refinement. Overall, these subgroups reflected a range of competencies such as software engineering, human-computer interaction, security, and data protection.

Usability

Usability refers to the ability of the CTAs to be easy to use and understood. We prioritized concerns of usability as the project centered around increasing adoption by citizens, and therefore understanding its usability for target audiences was essential. We derived the initial attributes after a review of CTA

evaluation frameworks [14-17,58], and from the usability frameworks for mobile apps/mHealth apps, accessibility, and UD literature [27,29,31,44-50]. Other sources that informed our deliberations were those discussing usability standards [61-63] in general and EU design requirements [55]. Accessibility was included as a high-level attribute under usability, and was mostly derived from the EU directive Accessibility EN 301 549 [30], from where we took initial requirements and checked them against the Web Content Accessibility Guidelines [56] to formulate probing questions attached to our identified attributes. Frameworks for designing touchscreen interfaces for children [52], evaluating apps for children [53], and General Data Protection Regulations (GDPR) regulation for minors [54] were also consulted. An early report of the work in this pillar is available [64].

Data Protection

Data Protection was chosen to accommodate societal concerns of privacy and security inspired by similar attributes in related works [28]. Although it has been noted for its complexity [65], we selected GDPR as our reference for the development of the Data Protection pillar: the GDPR of the European Union [66] is currently considered the foremost data protection legislation worldwide for protecting the rights of the individual. We retrieved the initial attributes and preliminary questions from national legal interpretations [67,68] for data-focused concerns. Since this approach excludes wider organizational attributes (now found, for example, within our Transparency pillar) and system-oriented goals such as information security, we developed a novel risk-based approach to compare the system security of CTAs, based on a review of the related literature in mobile app GDPR evaluation [65,69-76].

Effectiveness

Effectiveness measures how successful an app is in terms of the accuracy of its contact tracing, the COVID-restraining impact of the app over a jurisdiction, and the app's popularity with citizens. Concerns include detecting and sharing close contacts, providing relevant information to citizens, and assessing their reactions to that information. This pillar was informed by drawing and expanding on the definition of effectiveness in CTAs, provided by Lueks et al [77], and by considering Vokinger et al's [58] framework, which also explicitly tackles this concern.

Transparency

While transparency is officially a subset of GDPR, a separate Transparency pillar was created in the framework to consider wider aspects of transparency not specifically related to functionality. For example, while the GDPR approach to transparency considers specific data stores (such as locally stored contact information), transparency concerns such as the availability of a privacy policy or the open-sourcing of the source code would not fit into that approach. In other words, "transparency" in this context concerns how open the developing organization is with respect to its internal processes and artifacts. The initial attributes and their questions were formulated by extending our interpretation of GDPR [66], and considering already existing taxonomies [25,28].

Technical Performance

The Technical Performance pillar captures the efficiency of the contact tracing. Particularly, the Technical Performance pillar focuses on system resource utilization and execution speed, as these aspects impact use. The relevance of this pillar can be seen in how, for example, battery issues with the Exposure Notification Service provided by Google [78] and incorporated into the national Contact Tracker app in Ireland, caused battery issues over only one weekend and caused a large fall-off in app retention by the public [79]. The attributes for the Technical Performance pillar can be divided into resource utilization–related performance (eg, CPU/disk/memory usage) and efficiency-related performance (eg, response time). Because COVID-19 tracing apps are usually complex software systems (with dedicated front-end and back-end subsystems), the attributes can be applied to both subsystems respectively. The initial attributes for this pillar were derived from the "Performance Efficiency" category of ISO/IEC 25010, a software engineering quality model. The model is a standard for assessing characteristics of software systems and is widely applicable in software engineering.

Citizen Autonomy

Citizen Autonomy focuses on the citizen's ability to consent and the voluntary nature of the app. Its inclusion was inspired by the work of Gasser et al [16] studying a digital tool's ethical challenges. It was also based on the "User control/self-determination" domain in Vokinger et al's [58] assessment framework for (COVID-19) CTAs and the "autonomy" category in the checklist proposed by van Haasteren et al [28]. In these works, the authors focused on users' (existing) "data protection" concerns, which are mostly handled by our Data Protection pillar, but we wanted to extend the scope to specifically cover *initial* data access. Hence, this pillar focuses on a series of specific attributes that assess citizens' control over the app's access to phone functionalities such as the camera, microphone, and GPS.

(General) Characteristics

"General Characteristics" refers to characteristics that are nonevaluative, but serve to distinguish the app from others and other versions of the app. Thus, the static information captured by the Characteristics pillar acts as a necessary first step to conducting the more in-depth compare-and-contrast evaluation found in the other pillars. An initial set of distinguishing characteristics was derived by examining three CTAs: SwissCovid (Switzerland) [80], Apturi Covid (Latvia) [81], and Immuni (Italy) [82], and related data retrieved from their AppStore, Google Play, and app home websites. Next, we analyzed the Google and Apple Exposure Notification (GAEN) application programming interface (API)/framework [78] made available for use on Apple and Android devices, and the Decentralized Privacy Preserving Proximity Tracing (DP-3T) protocol [83,84] that inspired Google's API. We expanded our list of attributes further through a review of contact tracing protocols and frameworks listed on the Wikipedia COVID-19 Apps page [85]. Finally, we incorporated the literature review of app/mHealth app evaluations (see the grey literature in Multimedia Appendix 1 [32-39]).

At the end of the work described above, our initial list of 111 attributes had grown to 139 organized under 7 pillars.

Phase 2: Concretization and Critique

Test of the Framework Against Five CTAs

In the second phase of development, our provisional framework was tested against five CTAs that could be downloaded and activated in Ireland: Health Service Executive (HSE) COVID Tracker app (Ireland) [86], PathCheck SafePlaces (Massachusetts Institute of Technology [MIT], United States) [87], NOVID (United States) [88], Corona-Warn (Germany) [89], and Aman (Jordan) [90].

The two core considerations were to assess if the attributes could produce useful evaluation information on the CTA and how that information could be feasibly obtained. In this phase, feasibility concerns were sometimes overridden by the perceived importance of the information provided when probing a specific attribute. For example, the "number of people alerted-early to their close contact status, who then go for testing," seems like one of the core "Effectiveness" measures for CTAs. However, identifying this number involves a much wider information-gathering and reporting effort than is normally available from the app itself, and so could be quite difficult to assess [21] (ie, members of the public would need to inform authorities when they turn up for testing that they have done so based on an alert issued to them by the app, and that would need to be recorded on a national health system that ideally integrates back with the CTA).

The selected five apps reflected a *broad range* of approaches, as illustrated by their different *lead bodies*, and the different *data protection philosophies* underpinning them. In terms of "broad range," the Republic of Ireland's app has provided the basis for apps in other jurisdictions, both in Europe and the United States [79]. In terms of "lead bodies," these apps come predominantly from national health services, but PathCheck SafePlaces is an MIT-led initiative [91] and NOVID is crowd-sourced, originating from Carnegie Mellon University [88]. In terms of *data protection philosophies*, two of these apps originate in GDPR jurisdictions, but two originate from the United States, and one originates from Jordan.

The process followed in this phase was that the domain experts would apply their pillar to the five chosen apps to stress-test the ability of the framework to identify criticalities and key differences among apps. For each of the identified attributes, they formulated appropriate questions, and assessed the answers obtained to see if the attributes and the related questions had evaluative merit. Attributes were added where necessary, sometimes merged or reorganized. For example, in *General Characteristics*, the version number was identified in this phase as an important identifier, as CTAs, like other apps, tend to receive regular updates. Likewise, in *Effectiveness*, the effort/speed with which close contacts are alerted by CTAs was also identified as important. In contrast, *Usability* made sure that "accessibility" aspects were treated specifically and separately from more general usability and interaction aspects, which resulted in reorganizing some of the subattributes.

This concretization was sometimes complemented and reinforced with further, targeted reviews of the literature where deemed necessary.

Devil's Advocates Sessions

As the pillars were developed independently, a series of meetings across the expert subgroups were held to retain a wider perspective. Specifically, the meetings were organized so that domain experts, who were testing and refining a specific pillar, presented and defended their work to the other team members (see Table 1) who dissected the pillar and questioned its attributes under the headings of*:*

- *Relevance and Sufficiency*, where the team was encouraged to ask questions such as "why is this important?" and "what else might be important?";
- *Specificity*, where domain experts were encouraged to hypothetically answer each of the associated questions in the pillar and to (thus) probe it for any ambiguity; and
- *Cross-checking* with their own pillars to identify possible overlaps in the framework.

Table 1. Distribution of team members as pillar owners and devil's advocates in phase 2.

Pillar name	Pillar owner(s)	Devil's advocate(s)
(General) Characteristics	IO and SB	JB
Usability and Accessibility	CS, IR, and DT	IO and JB
Data Protection	TW	KR
Effectiveness	AR	DT
Technical Performance	MC	KR
Transparency	KR	MC
Citizen Autonomy	JB	IR

These meetings were in the form of "devil's advocate" sessions (7 in total), where 1 participant actively tried to identify/exaggerate flaws in the current attributes. This is because such an approach has been shown to increase the "accuracy of group solutions" [92]. The activity highlighted a number of changes mostly concerned with clarifying potential overlaps or redundancies, clarifying terminology and questions, and improving organization. At the end of the grounding and critiquing exercises, we ended up with a total of 163 "grounded" attributes and an initial formulation of 199 related questions. Additionally, we identified some cases to be discussed by the entire team during the third and final phase of our development.

Phase 3: Refinement and Finalization

To support the refinement work of the entire team (especially in consideration of the remote work environment demanded by COVID-19), we created a cross-pillar analysis matrix (Multimedia Appendix 2), which listed the ordered attributes and subattributes for each of the 7 pillars, assigning a unique identifier to each attribute (eg, the first General Characteristic attribute was assigned the identifier C01) and a color code to each pillar. To further support understanding of the attributes and to clarify their evaluative merit, we decided to also add sample answers (based on our grounding analysis of the five apps). We discussed the difficult cases identified in the previous phase and noted further (relevancy, sufficiency, specificity, and redundancy) issues for each attribute over a total of 15 refinement sessions around this cross-pillar matrix. The identified changes were progressively included in the framework, and 9 of the 10 authors were involved. For instance, the framework probes *Citizen Autonomy* in terms of whether there is an official discussion forum for citizens using the app and whether that forum can be used to prompt change (CA01, CA02; see Multimedia Appendix 2). It was noted that these overlap with attribute C16, a *Characteristic* attribute that probed the form of technical support, and U73, a *Usability* attribute probing the existence of interactive assistance for technical support or any other mechanism to submit feedback on technical issues, bugs, and errors detected. A reorganization was proposed, deleting CA11 (redundant with CA01 and CA02); changing C16 from "*Does the app offer technical support?*" into "*What form of technical support is available for the users, to include synchronous and asynchronous support?*"; and simplifying CA02 by removing its reference to *any other mechanism (to obtain technical support)*, as this was covered by the new phrasing of C16.

At the end of our 15 sessions, the refined pillars were applied to the newest versions of two of the five apps employed in the "grounding" phase (HSE's COVID Tracker [86] and NOVID [88]) to systematically double-check all attributes, questions, and answers so that we could either confirm or implement final edits. The main goal was to make sure the questions were clear and understandable. A number of other apps were also assessed less systematically to the same end.

At the end of this last test, our consolidated framework was restructured to 161 attributes and 180 related questions (with sample answers), which now had internal consistency and no overlap. Graphical visualizations of the refined pillars and their structures were also generated (Multimedia Appendix 3).

The consolidated framework (with 7 pillars and 161 attributes) was presented for feedback to medical researchers and practitioners from the wider "COVIGILANT" group, the (Irish) Department of Health, and the (Irish) HSE's "App Advisory Group," which included representative from Nearform, the company charged with creating the Irish national CTA. Likewise, informal discussions were held around the Effectiveness pillar with the European Centre for Disease Control (CDC), all serving to suggest a number of minor edits and tweaks to create the final version of the C^{3}EF for CTAs, as presented here.

At this stage, we also created a web-based application [93] to make our framework available in the form of an online survey. This acts as a demonstrator of our framework, but it has been devised to possibly assist relevant stakeholders of CTAs in independently evaluating their work and/or to share any feedback with us. This online tool offers visual overviews of the framework and gives access to the entire framework. With the depth and range of questions included in our C^{3}EF, the evaluation process may appear daunting and time-consuming. Consequently, we decided to provide access to individual pillars to enable breaking down the assessment, and allow stakeholders to select and prioritize their own assessment focus.

Results

Overview of the C^{3}EF Framework

In this description of our final framework, we define each pillar and provide an overview of its specific attributes, subattributes, and questions. We then offer a selection of sample questions and answers to illustrate how we used the framework to evaluate, compare, and contrast CTAs, and how this could be conductive of possible improvements in the apps, as questions often probe the desirable or best practice options. Table 2 offers a top-level view of the 7 pillars and the high-level attributes. (General) Characteristics is presented first, as it provides important contextualization/identification information for the other six pillars.

Table 2. The 7 pillars with their first- and second-level attributes (only).

First-level attributes	Second-level attributes
Characteristics Pillar	
1. General characteristics	
	Name of app
	Country
	Current versions
	Language support
	Age of users
2. Availability	
	Internet connectivity: app (other)
	Platform dependency
3. Organizational reputation	
	App status
	Development
4. App content	
	Processing overview
	Sensor employed
	App running state
	Contact tracing definition
	App data
	App permissions
	Notification method
	Diagnosis status
Usability Pillar	
1. Subjective satisfaction	
	Rating
	Motivations for high/low scores
2. Universality	
	Accessibility
	Cultural universality
3. Design effectiveness	
	Completeness
	Configurability
	User interface
	Helpfulness
4. User interaction	
	Efficiency
	Robustness
	Clarity of interaction with elements
	Consistency of interaction with elements
	Alerts and notifications messages
5. Ongoing app evaluation	Frequency of upgrade

First-level attributes	Second-level attributes
Data Protection Pillar	
1. Security	
	STRIDE[a] taxonomy/vulnerabilities
	CT[b]-specific threats
	Software architecture security
	SDLC[c] and security
2. GDPR[d]	
	Preliminaries
	GDPR principles
	Rights
Effectiveness Pillar	
1. Effective reporting	
	Detecting close contacts
	Reporting positive close contacts
	Reporting all close contacts
	Reporting hotspots
2. Effective results	
	Users who share their data
	Number of (additional) contacts/week found
	Number of those contacts found positive
	Relative effort per contact found versus manual CT
3. Effective engagement	
	Population uptake
	Population retention
	Population engagement
Transparency Pillar	
1. App transparency	
	App purpose
	App permission
2. User participation	App participation knowledge
3. Data transparency	
	Minimization, gathering, storing, accessibility, etc
	GDPR applicability
	Life cycle
Technical Performance Pillar	
1. Speed	Response time (frontend)
2. Efficiency	Response time
3. Consumption	
	Battery
	Disk space
4. Resource/troubleshooting and trust	
	CPU/memory usage

First-level attributes	Second-level attributes
	Bandwidth usage
	Throughput (backend)
Citizen Autonomy Pillar	
1. App discussion authority	
	Official discussion forums
	Empowered moderators
2. Phone functionality	
	GPS access
	Bluetooth
	ENS[e] access
	Notifications
	Microphone
3. Data control	
	Data upload authority
	Uploaded data location visibility

[a]STRIDE: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege.

[b]CT: contact tracing.

[c]SDLC: Software Development Life Cycle.

[d]GDPR: General Data Protection Regulation.

[e]ENS: Enhanced Network Selection.

Characteristics

Characteristics captures nonjudgmental criteria and factual information that are important to identify and differentiate a given app and its main functionalities. The App Characteristics pillar is organized according to four headings: *General Characteristics*, *Availability*, *Organizational Reputation*, and *App Content* (Table 2). Under these headings, there are a total of 25 specific questions that elaborate upon these app characteristics, all of which can be answered by direct inspection of the working app, through the information available on Google Play and Apple Store, or through the developer website.

General Characteristics captures four high-level app attributes, including the name of the app and the country. *Availability* looks at connectivity and platform dependency. The first aspect of availability questions whether an internet connection is needed to use the app, as some apps appear to require an internet connection even if they do not use the internet or location-based information for their contact tracing (eg, the Jordanian app AMAN [90]). The second questions are related to what platforms (Android/iOS) are supported and the download size.

Organizational Reputation looks at the status of the app, including whether the app is national and what official documentation is available. It also considers the organization that developed the app and whether any third-party or partners are involved in its development. This attribute examines the history of development, through an examination of the developers' prior experience of developing data-sensitive apps, along with evidence of updates, enhancements, and maintenance of the actual product. Even if we are aware that questions concerning organizational reputation (such as history) may disadvantage apps from new startups, we believe that this attribute captures an aspect that contributes to the confidence users will have in app adoption. Finally, the ability for users to ask questions and seek technical support is also probed.

App Content refers to what the app includes in terms of functionality and management of information. Definitions of contact tracing are queried, and information is given as to when and how contact tracing notifications are managed, since this may vary from country to country, or even from app to app. This is where key distinctions can be made between apps that use a different approach to contact tracing and notification. For instance, the HSE's app [86] states that the close-contact notifications will be activated when there is "direct exposure" to a positive case, where "direct exposure" refers to "within two meters for 15 minutes or more." In contrast, NOVID [88] notifies users when other NOVID users close in their social network are positive cases. The former, a more common approach, supports health authorities in warning citizens that they have been in contact with a positive case, while the latter notifies when the infection is close and aims at warning the citizens ahead of being in contact with the virus. NOVID classifies physical proximity as "near" or "far," with 6 feet (2 meters) or under being "near" and 12 feet (4 meters) or over being "far." Definitions of physical proximity in both systems (the HSE's app and NOVID) are based on parameters for proximity classification that can change in updated versions of the apps, depending on variants and infection events. However, at their core, these apps offer a different benefit to users. NOVID (which is not a national app and thus might be difficult to reach

the required critical mass of citizens adopting the system to help it perform at its best) seems to have an obvious advantage from the perspective of citizens' adoption. However, it would raise some issues under our Data Protection pillar because, even though this approach captures no personalized citizens' data and so users can be entirely anonymous, the social network data would need to be centrally collected and aggregated, providing indirect, yet not impossible, opportunities to deanonymize the information. We note that while such widely differing approaches make the application of any framework difficult, this attribute highlights the divergence at an elevated level, and many of the attributes proved resilient to this (NOVID's) different paradigm [94], though not all.

Usability

Five high-level attributes have been identified under *Usability: Subjective Satisfaction*, *Universality*, *Design Effectiveness*, *User Interaction*, and *Ongoing App Evaluation* (Table 2). Together, they offer the opportunity to ask 86 specific questions about usability aspects of CTAs.

Subjective Satisfaction looks at the perceived level of comfort experienced in using the app. As user retention is important in CTAs, this attribute captures how citizens rate their experience in using the app. It includes a rating attribute (1 to 5), and two attributes looking at motivations for high or low scores. To inform our answers to these questions, we typically look at the rating and reviews available on Google Play and Apple Store, although satisfaction could be better captured with longitudinal surveys.

Universality addresses population penetration, which is also key in the successful implementation of a CTA strategy. Specifically, universality aims to capture the ability of the app to be used by a variety of different users: users with potential impairments (physical or mental), but also users with different cultures/levels of education or of different ages. The first is captured by the subattribute Accessibility, and the latter is captured by Cultural Universality. Accessibility refers to the quality of being "easy to reach and use," and it mostly refers to users who might have a form of disability, impairment, or limitation (either mental or physical). We covered three aspects related to Accessibility: *Functional Performance*, *User Interface Elements*, and *Accessible Interactions*. The first two look at the interface and how its elements adhere to general accessibility guidelines and EU regulations [6,29,30,45,46,56]. Questions can be used as a checklist to make sure the app meets basic accessibility requirements. *Accessible Interaction*, the last subattribute under Accessibility, covers aspects such as onboarding (ie, features helping new users understand what the app does and learning how to use it) and the design of interactive elements to support low physical effort (eg, completing a task without scrolling, one-hand use, radio buttons). *Cultural Universality* helps to assess the extent to which the system can be used by different users regardless of their cultural background and beliefs. We developed attributes and questions to cover aspects such as (1) availability of different languages; (2) meanings that are evoked by the name of the app; (3) information on the age groups that the app targets, usually described in the "Terms and Conditions" (see Figure 2, Example 1); and (4) design elements such as logos, colors, national flags, and symbols for expressing cultural conventions.

Design Effectiveness covers several aspects concerning the capacity of the system, user interface, and interaction design to provide citizens with the necessary functionalities, options, commands, and supports. This attribute includes four dimensions that are found to be key in conveying the correct utilization of the system and its adaptation to different contexts of use and user preferences: *Completeness*, *Configurability*, *User Interface*, and *Helpfulness*. Completeness was formulated to identify both essential and optional functionalities offered by the app and those features that are not included in the app, but that users (eg, as voiced on online reviews) would like to have. Identification of core and optional functionalities help to identify user tasks that can be carried out by the user in interacting with the CTA interface, and this forms the basis for task-specific questions in our framework, which will come later. The identification of optional and emerging new functionalities is important as CTAs continue to evolve and offer new uses beyond contact tracing (eg, check-in, digital vaccination certificates, travel passes), and our question about optional functionalities allows our framework to incorporate them in the assessment. However, there is a tradeoff is in play here, which our framework can help to capture: offering a number of different functionalities could potentially increase the attractiveness of a CTA, thus resulting in higher adoption and satisfaction, while accommodating more functionalities within the same app might compromise usability (in our case resulting in potential issues identified under our accessibility, design effectiveness, and user interaction attributes). The next attribute (*Configurability*) looks at a variety of aspects concerning the capacity of the system to be personalized in terms of the technology in use (eg, allowing independent activation or deactivation of GPS, Bluetooth, and other technologies) or its design (eg, allowing personalization of visual, acoustic, and haptic feedback). *User Interface* deals with the assessment of the design elements used in the user interface with its *Aesthetic and Attractiveness* (concerned with the look and feel, color palette, and name of the app), *Responsiveness* (concerned with the ability to adapt to different phone models, screen sizes, and operating systems), and *Clarity* and *Consistency* of the design elements. These two last attributes offer *element-specific* subattributes, in that they refer to and evaluate specific elements of the interface and not the app as a whole. An element can vary from a button to a menu, a slider, or a table (see [95] for a full list and glossary of user interface elements).

Our questions explore the apps' perceptual and conceptual clarity, looking at their visibility, understandability (Figure 2, Example 2), and consistency (again both in terms of how they look and that their meaning is consistent throughout the app). In this sense, we intend for the framework to help to find potential flaws, hindering satisfaction, use, and adoption. However, both *Clarity* and *Consistency* also apply to structures of elements (eg, a button-bar containing buttons) in terms of how elements are logically grouped and whether these logic-based groupings are consistent. The last subattribute of *Design Effectiveness* is *Helpfulness*, looking at the suitability of documentation available to use and understand the app, and

whether interactive assistance is available. Subattributes look at the availability of supportive information such as definitions of the terms used (eg, what counts as a contact), descriptions (see Figure 2, Example 3), examples (eg, tutorial, walk-through or explanatory videos showing how to use the app), and the availability of interactive assistance for troubleshooting (eg, chatbots).

User Interaction helps to assess the user's interaction with the app interface in the execution of specific tasks, which are identified under the above-mentioned *Completeness* attribute. Similar to *Design Effectiveness*, *User Interaction* is important to ensure correct use of the CTA and lessening user frustration; however, in contrast to design effectiveness that considers what the interface statically offers, looks, and conveys (eg, affords), user interaction considers how the interface behaves when users interact with it (eg, feedback). It includes five subattributes: *Efficiency*, *Robustness*, *Clarity of Interaction with Elements*, *Consistency of Interaction with Elements*, and *Alerts and Notification Messages*. Most of the selected subattributes and their related questions are *task-specific*: similar to element-specific attributes, task-specific attributes and questions are specific to users carrying out one specific task from beginning to end (eg, activating/deactivating the contact tracing functionality). Therefore, to answer task-specific questions, app inspection is needed. *Efficiency* explores the capacity of the system to produce appropriate results in return for the resources that are invested. Here, we considered three elements: Human Effort, as the number of steps that are needed to carry out a core task; Time, which is the time needed to perform that task; and the Tied-up Resources, representing the potential need for external resources (eg, power or internet) to perform the task. *Robustness* deals with the capacity of the system to adapt to different user preferences and contexts of use but also its ability to deal with user errors. With our attributes, we look at: landscape/portrait mode, multitasking when using technologies such as Bluetooth or GPS for more than one app/task (eg, while using Bluetooth earphones), and the availability of multiple ways to achieve task execution (eg, shortcuts). Adaptability looks at supporting task execution in different environments (eg, in the dark), while Errors looks at error messages available in the app as a result of inappropriate interaction and the availability of error recovery options (eg, undo, redo) or the reversibility of user actions. *Clarity of Interaction with Elements* is concerned with the clarity of what can be done with the elements available in the app interface (namely their affordances) and what happens when users interact with these elements (eg, with respect to clarity/confusion of app feedback). *Consistency of Interaction with Elements* is next, which looks at potential consistencies of Actions across the elements, the Inconsistency of Feedback, and the use of Design Constraints (if any) to prevent human errors/guide users toward correct use (Figure 2, Example 4). In our analysis, we realized the importance of feedback on the contact tracing functionalities, as we noted a number of apps, especially in their early versions, that failed to offer clear feedback to the users after the contact tracing functionality was enabled. In most cases, users had to exit the app to enable, for instance, Bluetooth, and this can create confusion for the user. In some of these cases (Figure 3, Example 5), the button enabling contact tracing did not allow reversing the action (eg, click again to disable contact tracing) that could only be reversed by disabling Bluetooth from phone settings (not from within the app), which is also problematic in terms of users felling in control of the app.

Alerts and Notification Messages is the last subattribute of the User Interaction attribute, which refers to the alert messages and notifications used in the app. We included attributes and questions to assess various types of Alert Messages used in the app and to assess the availability of Notification Controls, particularly for built-in notification settings in the app and for Notification Messages that alert users who have been in close contact with someone that reported a positive COVID-19 test. This part closes with a question concerning the ability of the user to access and perhaps even manipulate or visualize the generated contact tracing data (eg, number of contacts in a day, week, etc).

Ongoing App Evaluation refers to the app's maintenance and upgrading, as these are important to maintain retention while also targeting new emerging needs. It includes only one subattribute looking at the Frequency of Upgrade: this can be found in Google Play and App Store, where the app can be downloaded and installed.

Figure 2. Example 1, PathCheck SafePlaces (United States) [87]: age in "Terms of Use." Example 2, Corona-Warn (Germany) [89]: understandability of interface elements. Example 3, NOVID app (United States) [88]: descriptions offered. Example 4, COVID Tracker app (Ireland) [86]: constraints for preventing errors.

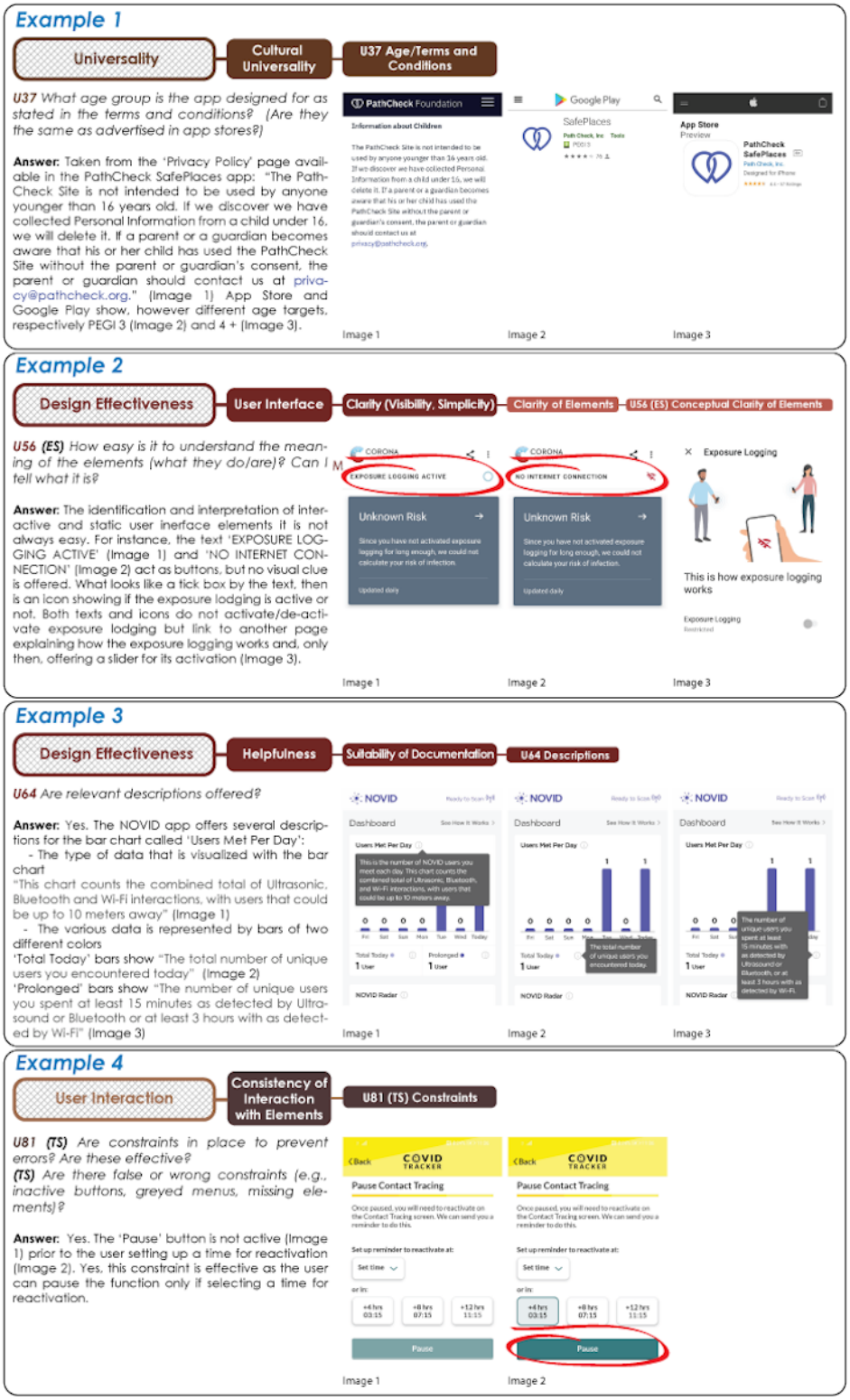
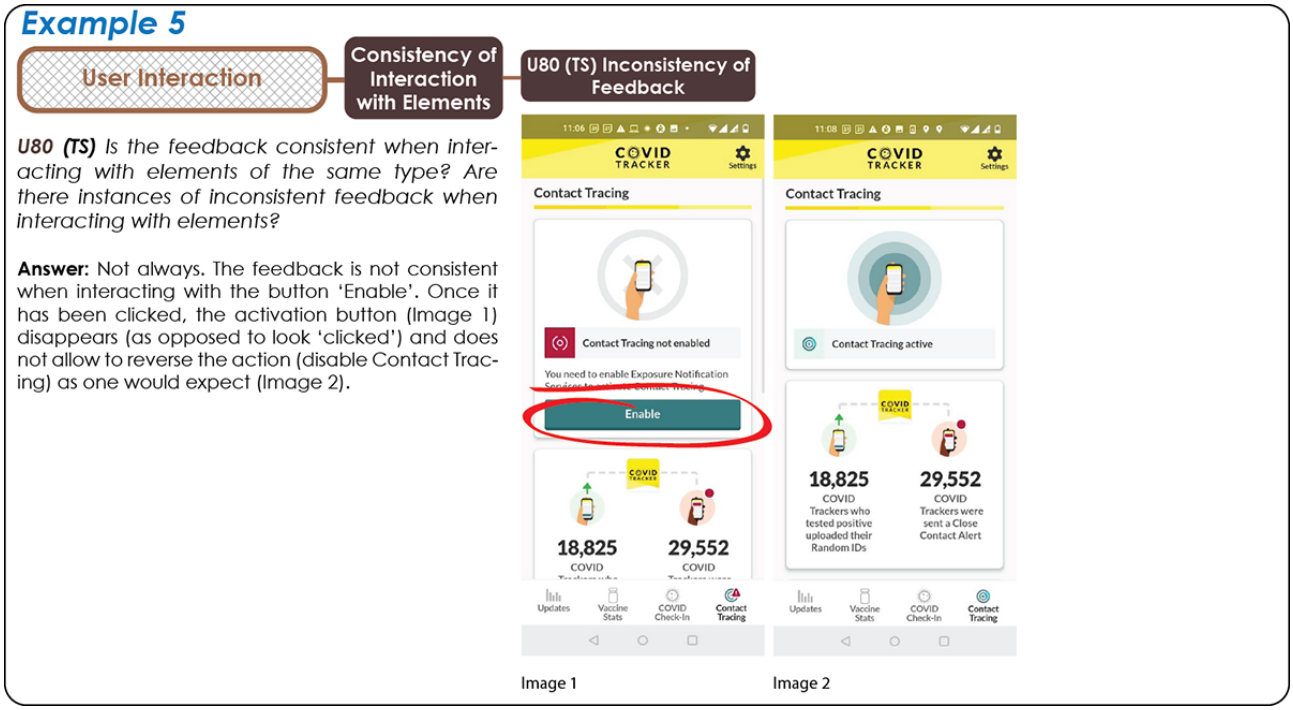


Figure 3. Example 5, COVID Tracker app (Ireland) [86]: inconsistency of feedback when clicking on the button "Enable".



Data Protection

The *Data Protection* pillar consists of two subcategories: *Security* and *GDPR*. The GDPR category focuses upon the rights of the individual citizens, while the security category takes a more data-centric view (Table 2).

Security consists of 4 criteria, which center around contract tracing–specific security threats and vulnerabilities. They are scoped to ensure that CTAs are compared fairly, such that security vulnerabilities related to software or system components that cannot be changed by the CTA development are not considered; for example, those related to the system security of third-party providers (third-party vendors are noted in "General Characteristics," under "development partners," so that an indirect warning flag is retained). These attributes incorporate a novel approach to CTA evaluation that was developed to ensure a lightweight comparison using the potentially incomplete data available for each app: analyzing vulnerabilities of distinct app functionalities against a common threat assessment model [96]. Attributes under Security (namely, STRIDE [Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege] taxonomy/vulnerabilities, contact tracing–specific threats, software architecture security, and Software Development Life Cycle [SDLC] and Security) are designed to indicate whether these vulnerabilities are bugs in the code, which can be fixed or would require a redesign of the architecture to address. For example, when we used this approach to compare the security of two CTAs (Corona-Warn [89] and MyTrace [97]), we first used automated tools to identify Common Weakness Enumerators (CWEs; a categorized "encyclopedia" of over 600 types of software weaknesses), and then we manually confirmed them using in-house security expertise. We compared the identified enumerators for both apps (Table 3) and then against the predefined common threat assessment model (Table 4), providing an answer to our Security questions. Our analysis showed that while both apps may suffer from similar concerns related to information disclosure and deanonymization, the CWEs that enable these are different, with our questions under SDLC and Security allowing us to capture if the identified vulnerabilities can easily be patched/fixed or would be more difficult to correct.

GDPR considerations are important because they speak to the essential user concern of data privacy. The GDPR attribute has three subattributes, including Preliminaries, GDPR Principles, and Rights. Preliminaries involves information required for evaluation of the individual data stored later: Data Stored, Data Type, and Basis for Processing. It also includes Withdrawal and whether the organization declared consent and has a legal requirement as their basis for processing the data. For instance, applying question DP06 (Data Stored): "What personal data are collected?" to the COVID Tracker App [86] will generate the following list: *phone number, date of last exposure, sex, age range, county, town, symptoms, diagnosis keys, date of symptom onset, app metrics, IP address, and app security tokens*. GDPR Principles refer to the key principles of GDPR (such as Minimization, Fairness, and Storage Limitation), which are not under scrutiny in other dimensions of our C^3EF framework. It consists of 5 attributes that are evaluated across the stored data retrieved from the question on Data Stored. Most of these criteria require details from the data controller, and must rely on those details being truthful and accurate. Our final attribute under GDPR is Rights, which refer to the rights of the individual that must be upheld if they are the subject to data processing by an organization. Therefore, they refer to the availability of organizational procedures to ensure these rights. We have 5 attributes under GDPR Rights: Access; Object to Reuse; Portability; Automated Processing Rejection; and Rectified, Restricted, or Erased (data). This is where a key distinction can be captured between CTAs that notify direct exposure to positive cases and the NOVID approach, which we discussed under

Characteristics above. The way data are aggregated in NOVID creates a conflict with an individual's right under the GDPR, as exercising a right that results in removal or change of these data (eg, the right to withdraw) will affect the wider data set. Additionally, such a data model would create conflicts with core GDPR principles such as data minimization, and avoidance of user may result in reidentification, through the combination of multiple data sources. Similarly, the AMAN app [90], which is not based in an EU state and, as such, is only required to adhere to the GDPR if it is used by EU citizens, had a distinct lack of documentation to support GDPR rights, as is to be expected.

Table 3. Comparison of Common Weakness Enumerators (CWEs) in the Corona-Warn [89] and MyTrace [97] apps.

CWE	Corona-Warn	MyTrace
89: A (SQL[a]) Command	Local SQL injection possible but data encrypted	Local SQL injection possible and data not encrypted
276: Incorrect Default Permissions	N/A[b]	Permissions for tasks, Bluetooth administration, and external storage
295: Improper Certificate Validation	Vulnerable to SSL[c] MITM[d] attack	N/A
532: Insertion of Sensitive Information into Log File	Sensitive information is encrypted	Excessive information logged
327: Use of a Broken or Risky Cryptographic Algorithm	Weak hash function in SSL	N/A

[a]SQL: Structured Query Language.

[b]N/A: not applicable.

[c]SSL: Secure Socket Layer.

[d]MITM: man in the middle.

Table 4. Comparison of threats to Corona-Warn [89] and MyTrace [97] using the Common Weakness Enumerators (CWEs) listed in Table 3 (with a severity rating: H=high, M=medium, and L=low) against the common threat assessment model.

Threat	Corona-Warn Matched CWEs	MyTrace Matched CWEs
Fake alert injection	N/A[a]	CWE-327-H
False report	CWE-295-H, CWE-327-L	N/A
Proximity beacons altered	N/A	CWE-89-H
User can deny or retract infection report or contact details	CWE-295-H	N/A
Personal information disclosed	CWE-327-L, CWE295-H	CWE-89-H, CWE-276-H
User deanonymized and tracked	CWE327-L, CWE295-H	CWE-89-H, CWE-276-H, CWE-532-H
Energy resource drain attack	N/A	CWE-276-H
System resource contention	N/A	CWE276-H

[a]N/A: not applicable.

Effectiveness

The *Effectiveness* pillar refers to the degree to which the app is successful in its core aims: accurately detecting close contacts and thus providing "notification to other app users with potential exposure risks to an infected app user" [98]. It contains three high-level attributes (see Table 2), the first of which (*Effective Reporting*) refers to concerns related to accurate detection, and the second of which (*Effective Results*) refers to providing notification to other app users with potential exposure risks, a concern provisionally referred to as "performance" by other commentators in the field [77]. The third attribute (*Effective Engagement*) refers to the "other app users" and "infected app users" in the definition, specifically focusing on the level of app adoption by citizens.

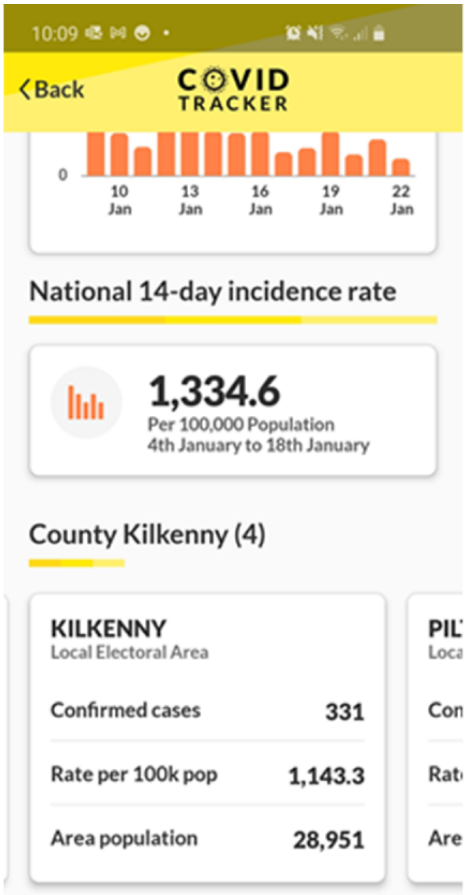
Effective Reporting focuses on the ability of the app to report accurately on close contacts and the location of virus hotspots to individual users. It first assesses the accuracy of close contact detection: often, this will have to be reported at the protocol level, for example, stating that the app reports at GAEN-level accuracy [78]. The framework then breaks down the reporting of close contacts into two categories: reporting on COVID-19–positive close contacts and reporting on a user's total number of close contacts over time periods. This latter category is sometimes reported in an effort to highlight and refine users' behavior, as in the case of PathCheck SafePlaces [87]. Finally, several of the apps report on prevalence information by locale to let users know areas where the virus is more (or less) prevalent. This is a form of hotspot identification for users (see Figure 4 for an example from the HSE's COVID Tracker app [86]). The final question in this

section probes the availability/granularity of this hotspot facility ("electoral division" in the case of Figure 4).

Effective Results focuses on the ability of the app to meet its wider objectives across the jurisdiction. Hence, it concentrates on the (total) number of users who choose to share their data after being told of their positive status and the number of additional close contacts that are informed, based on that sharing. It then looks at the number of those contacts who were subsequently identified as positive. Finally, it aims to assess the relative time and effort in identifying a close contact via the app, as opposed to via the manual-tracing effort.

Effective Engagement focuses on the population's adoption of the app, probing the uptake of the app across the population; citizens' retention of the app over time; and, in cases where the app contains interactive features, the population's engagement with the app, as possibly measured by their usage of these interactive features. This is important because digital contact tracing is very dependent on the proportion of the population who upload the associated app and retain it over time.

Figure 4. Electoral district–level COVID-19 statistics on the Health Service Executive's COVID Tracker app (Ireland) [86].



Transparency

Transparency is discussed here as an independent pillar, despite the obvious overlap with GDPR. This is because here it addresses the transparency of the processes and artifacts utilized/formed during development of the app specifically. In this context, transparency is an important aspect from the perspective of adopting citizens' confidence. The pillar has been divided into categories looking at *App Transparency*, *User Participation*, and *Data Transparency* (Table 2).

App Transparency includes App Purpose and App Permission. App Purpose offers the attributes (1) App-Purpose Knowledge, which refers to the purpose of the app being made accurately and accessibly explicit to the adopting citizen; (2) App Participation Knowledge, which looks at whether the citizens receive a clear explanation of the voluntary nature of participation; (3) App Development Knowledge, which looks at the mechanisms employed to guarantee community feedback; and (4) Open Source Repository to assess if the source code is made available (eg, on GitHub), as this too shows transparency at a high level. App Permission investigates all the permissions that are being asked for by the app, such as permission to access Bluetooth/the camera, in terms of how transparent the app is about these phone functionalities accessed. Modus Operandi probes the CTA's transparency regarding the permissions required for its functionality, and looks at the time period over which the services are being used as well as the contact tracing accuracy claimed by the developers (eg, with questions such as "Is the app being transparent about the contact tracing accuracy that they are achieving?").

User Participation consists of only one question, ensuring user consent: "Does the app indicate and explain to the end user about the voluntary nature of participation?"

Data Transparency focuses on whether the app has been designed following a privacy-by-design principle (under Data

Capture Knowledge), as this will heighten confidence as to its data privacy nature. It assesses if the users are made accurately and accessibly aware of the data accessible to other bodies, both in terms of the data and the accessing bodies (Data-Access Knowledge). It probes if the users are made explicitly aware of where and for how long their data are stored. "Privacy policy knowledge" is also of concern here, looking at if, how, and when the citizen is informed about the data being collected in a Data Protection Impact Assessment. Two more attributes look at the "minimality" of the collected personal data and at Data Protection, which focuses exclusively on the transparency section of data encryption and data anonymity. It has several questions that assess the anonymity, encryption protocol, and the end-of-life conditions for the data.

Technical Performance

Technical Performance defines how efficiently a software system operates (in contrast to effectiveness), and it includes attributes and questions that help to capture this operational efficiency. The main attributes populating the Technical Performance pillar are: *Speed*, *Efficiency*, *Consumption*, and *Resource/Troubleshooting and Trust*.

The *Speed* subattribute captures how quickly a software system's frontend app responds to a user's requests, as delays may cause user frustration. This subattribute probes two issues. The first queries how fast the app responds to a user's interaction. The response here is measured in time units (eg, milliseconds), and can be influenced by several aspects such as third-party apps (including an operating system's libraries and API), hardware and its configurations, various components of the app, and how they work together.

The second *Efficiency* question focuses on the algorithms of the app that are responsible for answering a user's requests.

Consumption and *Resources/Troubleshooting and Trust* capture how efficiently a software system consumes available hardware resources, including efficiency of battery usage, disk usage, CPU and memory usage, and bandwidth consumption. These attributes are particularly important with respect to retention if users perceive a battery/storage/CPU drain on their device [79] and so should probably be assessed as above or below according to some sort of "noticeable-threshold" level.

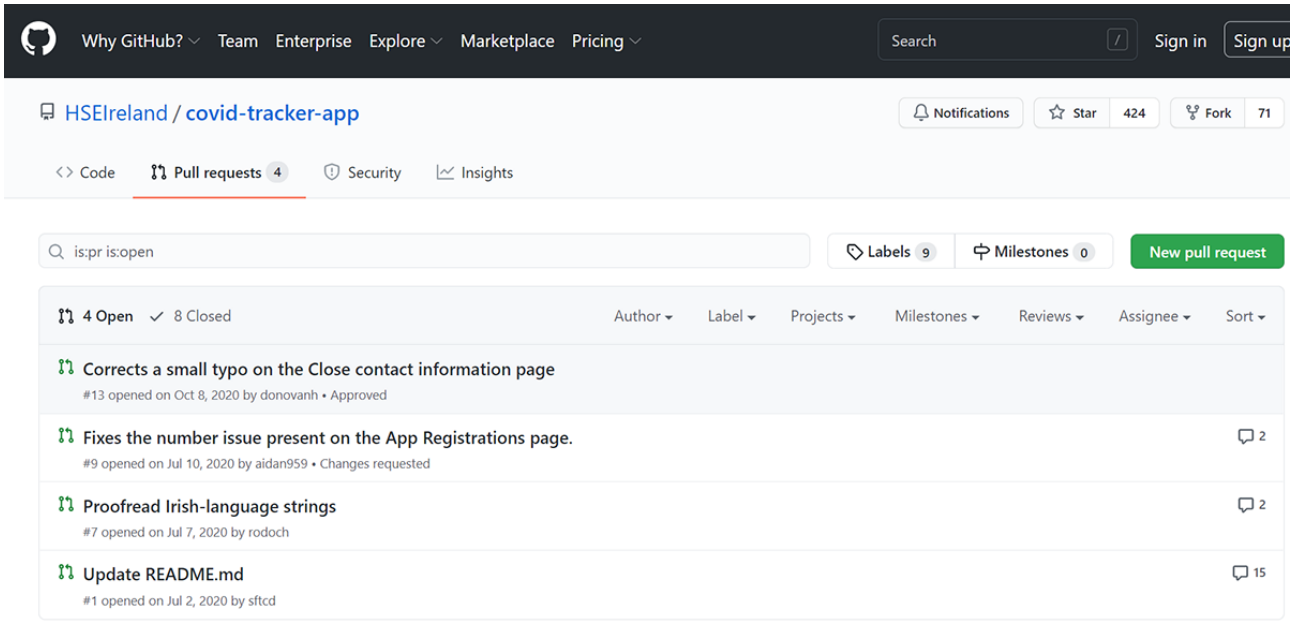
Citizen Autonomy

The Citizen Autonomy pillar refers to the degree to which a user has the ability to define their own control levels in terms of the rights and accesses they grant the app. Additionally, it is concerned with the user's ability to influence the evolution of the app going forward: an important element of autonomy, given that jurisdictions are asking users to retain the app for the duration of the emergency.

The pillar has three high-level attributes: *App Discussion Authority*, *Phone Functionality*, and *Data Control* (see Table 2). Cumulatively, these three categories consist of 9 questions.

App Discussion Authority focuses on the ability of the user to influence the future direction of the app and thus feel a sense of ownership. It first checks if there is a discussion forum where users are free to leave opinions and requests for change (most apps have at least a review section on Google Play or Apple's App Store by default). An important consideration then is whether the available review sites are curated or moderated by representatives of the app development team. For example, the HSE's COVID Tracker app [86] has reviews on Google Play and Apple Store, but it also has a GitHub repository [99] where users can leave their push requests for developers, as illustrated in Figure 5.

Figure 5. The "pull requests" GitHub page [99] for the Health Service Executive's COVID Tracker app [86].



Phone Functionality focuses on the ability of the user to control the app's access to phone services. Typically, these are relevant services such as GPS, GAEN [78], Bluetooth, and notifications, where their role in the app's functioning is apparent. However, occasionally, additional services may be required. For example, NOVID [88] uses ultrasound in an attempt to make close-contact detection more accurate and, as a result, requires access to the phone's microphone.

Data Control focuses primarily on the data that are uploaded from the user's app, typically in an instance of a positive diagnosis for the virus. It checks if the user is explicitly asked their permission for the upload of these sensitive data to happen and if the user is made aware of where the resultant data are stored (and by whom).

Discussion

Significance of the Framework

The research question of this study revolved around devising and organizing a framework to enable a more comprehensive assessment of current CTAs, thus supporting the work of decision-makers (eg, developers, health authorities) in the development and evolution of CTAs, and potentially increasing citizens' adoption. While evaluation frameworks exist, they tend to focus on specific aspects of CTAs, and we wanted to enable a more holistic assessment that could help to compare and improve the design of CTAs across a series of dimensions important for citizen adoption. For this purpose, we also wanted to assess what is available in the literature and ground what we derived with empirical and iterative tests run using a selection of CTAs. In doing this, we formulated and added probing questions (and sample answers) to the attributes derived for our framework, which we consider another key contribution, in terms of its application by interested stakeholders.

Open-Ended Nature of C^3EF

Questions in C^3EF were mainly formulated to assess the essential functionality of the CTAs (ie, the contact-tracing function). Nonetheless, one important consideration at this stage concerns the open-ended nature of C^3EF. Their iterative derivation and refinement make particularly apparent how CTAs are constantly changing and evolving. Not only are CTAs (more or less) frequently updated to fix issues and improve functionalities, but they also operate in a changing scenario, which offers new requirements and design opportunities over time. For instance, at the beginning of our project, vaccines were not available. However, at the time of writing, a number of vaccines are being administered, which opened up new needs (digital vaccine passports, interoperability between systems when traveling abroad) that might extend the scope of current CTAs as we know and evaluate them now. We do not see this as a limitation of our work, but rather as an invitation to progress research on this topic and to extend the framework. Moreover, pillars such as Usability, Data Protection, Citizen Autonomy, and Transparency offer several questions that are agnostic to the types of functionalities under scrutiny, and can be easily applied to nonessential functionalities such as check-in functions, statistic dashboards, displaying tests, and vaccine certificates, among others. In this sense, our framework can support assessment of mHealth apps that are not necessarily focused on contact tracing, although a number of questions remain specific to CTAs and are not necessarily applicable outside this domain.

Reporting transparently on our methodology, while also sharing a supporting website that offers interested stakeholders public access to the framework and the ability to send us feedback [93], represent strategies to facilitate adoption of our framework and adapting it to new emerging questions and needs. In this sense, we hope that researchers will engage, criticize, and improve on those attributes to support improved and expanded CTAs over time, but we also hope that stakeholders can use the included evaluation concerns as guidance when designing, evolving, or evaluating CTAs.

Use of the Framework

The examples we share in our Results section (Figures 2-5, Tables 3 and 4) are meant to illustrate how the framework can be used to assess or compare apps. The adding of attribute-related questions, sample answers from our own tests, visual representations of the structure of our taxonomy, and, once again, a website that allows others to use the framework in the form of a survey (with the opportunity to leave feedback on the application of the framework) are all devices that we devised to support potential stakeholders (eg, developers and health authorities) in using our framework to develop, evolve, and improve the design of CTAs with an eye on citizen adoption.

Although the questions tend to be descriptive in nature, the framework can also be used as a checklist of important elements to be considered. In this sense, the framework is prescriptive, as it helps developers appreciate either desirable qualities (eg, asking if there is a forum for citizens' feedback also implies that it would be desirable for CTAs to offer such a forum to citizens). Likewise, it can inform them of design tradeoffs. For example, while more functionalities and actionable information might make the CTAs more attractive, these might also make them more difficult to use. Another example is the alternative (CTA) approach that warns users in a social network *prior* to exposure to a positive case (as implemented in NOVID). This could be more appealing to citizens, but it is potentially more vulnerable to data deanonymization.

Most of our questions are devised to be answered via app inspection, inspection of the app website, or inspection of the page in the app stores (eg, Characteristic, Transparency, Usability, and Citizens' Autonomy pillars). We assume that interested stakeholders possess the required skills to independently answer the questions of these pillars. In some other cases, answering questions could be more difficult in terms of feasibility but also in terms of required competences. For example, the security section of the Data Protection pillar analyzes vulnerabilities of distinct app functionalities against a common threat assessment model [96], and this requires technical skills and familiarity with system security tools. Similarly, Technical Performance assessment assumes a user with a background in software engineering.

In terms of feasibility, we often noted a tension between the importance of an attribute and the feasibility of getting the information to answer the attribute's question. A good example is presented under Effectiveness, where the number of users tested early in response to an app notification is considered a core measure. However, this is difficult to ascertain without significant buy-in from and integration with the national health authority working to capture and make such data available. However, we believe that the framework should ultimately contain important attributes, even when it is difficult to obtain

an answer, as they often highlight the core criteria that should be used to assess the apps, and thus reflect a need for wider buy-in by the associated health authority and/or citizens.

General CTA Design Considerations

As part of our derivation and consolidation phase, we had to apply our framework (its early, iterated, and finalized versions) to the 5 apps, and this put us in a position to appreciate a number of issues we found across all CTAs and that further confirm the ability of our framework to capture critical aspects. For instance, some of the questions to assess Universality and Accessibility (under Usability) revealed that most apps do not support minors as users under parental/legal guardian consent, and that interfaces and functions do not seem to be designed to accommodate use by children under the age of 13 years. We consider this to be an issue given that the vaccination campaign for individuals under 13 is still not clear and they are a large cohort in most national populations.

As mentioned above, we also found that most apps lack clear and actionable feedback, which can sometimes be conducive to confusion (eg, is my contact tracing active?) and lack of sense of control. Similarly, we found an overall scarcity of information push and synchronous interactive features, including that of synchronous assistance (only offered by Corona-Warn in Germany). This lack of synchronous, interactive features has the potential to negatively affect adoption because the perceived benefits of CTAs are not made sufficiently apparent. In a related work, we argued how feedback and providing more diverse and actionable information to citizens (eg, number of daily contacts, proximity buzz, hotspots) was seen as key to maintain engagement in CTAs.

Finally, we also note that, in terms of security, individual CWE issues are not standard across CTAs. Even so, individual security issues seem to arise in CTAs, based on individual CWE issues specific to the CTAs reviewed. Exacerbating this concern is the fact that the framework does not assess the complete CTA, ignoring third-party components with respect to security.

Conclusions

Our stated research question was "how to devise and organize a framework to enable a more comprehensive assessment of current CTAs, supporting the work of decision-makers (eg, developers, health authorities) in the development and evolution of CTAs, potentially increasing adoption?" While the "how" part of this question is largely addressed in our Methods section, the resultant citizen-focused CTA framework provides a holistic compare-and-contrast tool that supports the work of decision-makers in this sphere. It aims to support reflection on developers' design decisions so as to better understand and optimize the design compromises in play. As such, we see it as a vital tool for designers designing and evolving current CTAs for increased public adoption. However, it can also be used by commentators in the assessment of CTAs, more generally, across jurisdictions to identify more optimal alternatives and prevalent, problematic issues. Such commentaries can be an important tool when governments are looking for CTA solutions to adopt or mimic.

For these purposes, it is important that we continue to assess existing and new CTAs against the framework, and document the results. As a prerequisite, we must also develop "best-practice" guidelines for performing app evaluations when using the framework. Both of these agendas are areas of ongoing concern for us in the "COVIGILANT" project. Ideally, this work will result in an openly accessible protocol and a database of framework application results, where CTA designers/developers will also have the right to reply (to heighten traceability), and these facilities would act as a single resource where designers could go to obtain a broad comparison over the CTAs available.

Nevertheless, there are still issues to address: the example discussed earlier assumes, for instance, that data about the number of citizens alerted of their COVID-19 positivity and the number who decide to get tested because of a CTA notification exist, and that these numbers are monitored. This is only possible if certain procedures are established by health authorities (such as keeping records of individuals who ask for a test because of a notification from their CTAs). Our framework focuses on the app itself; however, to have a significant impact on the spread of the virus, CTAs need to be integrated as part of the pandemic response through an ecosystem of organizational, political, and social entities, which goes well beyond the scope of our framework. The temporary failure of health systems in dealing with digital contact tracing has been well documented [21], and this means that *scoring well* on our C^3EF does not necessarily guarantee that the CTAs will have an evaluable impact on containing the spread of the virus.

Another limitation is that target users were not directly involved in the initial derivation of our C^3EF. The project was run during the first lockdown in Europe and, despite the team having connections with CTA developers in Ireland, the Irish health authorities, and the European CDC, we acknowledge that the interactions held with these bodies were informal sessions toward the end of the derivation process. In those sessions, they provided feedback to us on the framework and, in the case of the CDC, we contributed to their effectiveness framework [100] using the insights obtained to identify small refinements in our own framework. We desisted from detailing these interactions in our Methods section based on their lack of formality, and acknowledge that more should have been done to incorporate these central stakeholders in the derivation. To partially address this, we have made our framework publicly available and offer the opportunity for users to leave feedback in an attempt to support future engagement with stakeholders. We appreciate that this is unlikely to happen without further directed work, and this too is an area of future work for us.

A final potential limitation of this work is the methodology employed. While the differing approaches employed can be considered a form of data triangulation, and thus a positive attribute [101,102], the data were nonindependent, with earlier data being used as the foundation for later data generation/refinement. In addition, the protocol we used for the first part of phase 2 (applying the provisional framework to 5 CTAs) and phase 3 of the analysis (Refinement and Finalization) were derived specifically for this study, and not based on any

established method. However, this was because of the specific goals of these analysis phases: determining relevancy, sufficiency, specificity, and redundancy issues in the framework. In addition, the extensive, immersive, and targeted nature of the application/refinement sessions, along with the final evaluation of the resultant framework against existing CTAs suggest a level of rigor in these phases that is defensible.

Acknowledgments

The work undertaken was supported, in part, by Science Foundation Ireland (SFI) grant 13/RC/2094 and the "COVIGILANT" Science Foundation Ireland grant 20/COV/0133. We would like to express our gratitude to all members of the "COVIGILANT" Group, including Manzar Abbas: Lero, Science Foundation Ireland Research Centre for Ireland, University of Limerick, Limerick, IE; Brian Fitzgerald: Lero, Science Foundation Ireland Research Centre for Ireland, University of Limerick, Limerick, IE; Liam Glynn: School of Medicine, University of Limerick, Limerick, IE; Kevin Johnson: Department of Nursing and Midwifery, University of Limerick, Limerick, IE; John Laffey: School of Medicine, National University of Ireland Galway, Galway, IE; Bairbre McNicholas: School of Medicine, National University of Ireland Galway, Galway, IE; Bashar Nuseibeh: Lero, Science Foundation Ireland Research Centre for Ireland, University of Limerick, Limerick, IE; Michael O'Callaghan: School of Medicine, University of Limerick, Limerick, IE; James O'Connell: Lero, Science Foundation Ireland Research Centre for Ireland, University of Limerick, Limerick, IE; Derek O'Keeffe: School of Medicine, National University of Ireland Galway, Galway, IE; Andrew Simpkin: School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, IE; Jane Walsh: School of Psychology, National University of Ireland Galway, Galway, IE.

Authors' Contributions

Twenty-two researchers worked in the "COVIGILANT" project, divided into 3 streams. Stream 2 (10 researchers) is responsible for the development of the C^3EF as follows: Characteristics (IO and SB), Usability (CS and DT), Accessibility (IR, CS, and DT), Data Protection (TW), Effectiveness (AR and JB), Transparency (KR), Technical, Performance (MC), and Citizen Autonomy (JB). DT, CS, and KR designed pillar visualizations and a web-based digital tool. CS, JB, and DT wrote and revised this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A summary of all papers identified for the literature review. Initial list of 111 extracted attributes and categories, our analysis, and grouping into 6 categories.
[DOCX File , 36 KB - mhealth_v10i3e30691_app1.docx]

Multimedia Appendix 2
Snippet of the cross-pillar/ambiguity analysis matrix.
[DOCX File , 73 KB - mhealth_v10i3e30691_app2.docx]

Multimedia Appendix 3
Visualization of all pillars and questions.
[PDF File (Adobe PDF File), 343 KB - mhealth_v10i3e30691_app3.pdf]

References

1. Baxter S, Goyder E, Chambers D, Johnson M, Preston L, Booth A. Interventions to improve contact tracing for tuberculosis in specific groups and in wider populations: an evidence synthesis. Health Serv Deliv Res 2017 Jan;5(1):1-102. [doi: 10.3310/hsdr05010] [Medline: 28121090]
2. Swanson KC, Altare C, Wesseh CS, Nyenswah T, Ahmed T, Eyal N, et al. Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015. PLoS Negl Trop Dis 2018 Sep;12(9):e0006762 [FREE Full text] [doi: 10.1371/journal.pntd.0006762] [Medline: 30208032]
3. Anglemyer A, Moore TH, Parker L, Chambers T, Grady A, Chiu K, et al. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst Rev 2020 Aug 18;8:CD013699 [FREE Full text] [doi: 10.1002/14651858.CD013699] [Medline: 33502000]
4. Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, et al. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis 2019 Sep 18;19(1):810 [FREE Full text] [doi: 10.1186/s12879-019-4354-z] [Medline: 31533659]

5. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
6. eHealth Network. Mobile applications to support contact tracing in the EU's fight against COVID-19: Common EU Toolbox for Member States. European Comission. 2020 Apr 15. URL: https://ec.europa.eu/health/system/files/2020-04/covid-19_apps_en_0.pdf [accessed 2020-07-02]
7. COVID-19 Contact tracing: data protection expectations on app development. Information Commission Office (ICO). URL: https://ico.org.uk/media/for-organisations/documents/2617676/ico-contact-tracing-recommendations.pdf [accessed 2020-07-15]
8. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul 04;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
9. Digital tools for COVID-19 contact tracing. World Health Organization. 2020 Jun 04. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Contact_Tracing-Tools_Annex-2020.1 [accessed 2020-08-16]
10. Jacob S, Lawarée J. The adoption of contact tracing applications of COVID-19 by European governments. Policy Design Pract 2020 Nov 28;4(1):44-58. [doi: 10.1080/25741292.2020.1850404]
11. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
12. Martin T, Karopoulos G, Hernández-Ramos JL, Kambourakis G, Nai Fovino I. Demystifying COVID-19 digital contact tracing: a survey on frameworks and mobile apps. Wirel Commun Mob Comput 2020 Oct 17;2020:8851429. [doi: 10.1155/2020/8851429]
13. Vinuesa R, Theodorou A, Battaglini M, Dignum V. A socio-technical framework for digital contact tracing. Results Eng 2020 Dec;8:100163. [doi: 10.1016/j.rineng.2020.100163]
14. Cho H, Ippolito D, Yu Y. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv. 2020 Mar 24. URL: http://arxiv.org/abs/2003.11511 [accessed 2020-08-23]
15. De Carli A, Franco M, Gassmann A, Killer C, Rodrigues B, Scheid E, et al. WeTrace -- A privacy-preserving mobile COVID-19 tracing approach and application. arXiv. 2020 May 19. URL: http://arxiv.org/abs/2004.08812 [accessed 2020-07-07]
16. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: Framing the ethical challenges and how to address them. arXiv. 2020 May 21. URL: http://arxiv.org/abs/2004.10236 [accessed 2020-07-30]
17. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
18. Bente BE, van 't Klooster JWJR, Schreijer MA, Berkemeier L, van Gend JE, Slijkhuis PJH, et al. The Dutch COVID-19 contact tracing app (the CoronaMelder): usability study. JMIR Form Res 2021 Mar 26;5(3):e27882 [FREE Full text] [doi: 10.2196/27882] [Medline: 33724198]
19. Weiß JP, Esdar M, Hübner U. Analyzing the essential attributes of nationally issued COVID-19 contact tracing apps: open-source intelligence approach and content analysis. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e27232 [FREE Full text] [doi: 10.2196/27232] [Medline: 33724920]
20. Europe for Privacy-Preserving Pandemic Protection (E4P); Comparison of existing pandemic contact tracing systems. ETSI. 2021. URL: https://www.etsi.org/deliver/etsi_gr/E4P/001_099/002/01.01.01_60/gr_E4P002v010101p.pdf [accessed 2021-02-20]
21. von Wyl V. Challenges for nontechnical implementation of digital proximity tracing during the COVID-19 pandemic: media analysis of the SwissCovid app. JMIR Mhealth Uhealth 2021 Feb 26;9(2):e25345 [FREE Full text] [doi: 10.2196/25345] [Medline: 33606658]
22. Saunders M, Lewis P, Thornhill A. Research Methods For Business Students. 8th ed. New York, NY: Pearson Education; May 2019.
23. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. 2014 Presented at: EASE '14: 18th International Conference on Evaluation and Assessment in Software Engineering; May 2014; London, England p. 13-14. [doi: 10.1145/2601248.2601268]
24. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for SARS-CoV-2 digital tracking and tracing systems. In: Cowls J, Morley J, editors. The 2020 Yearbook of the Digital Ethics Lab. Digital Ethics Lab Yearbook. Cham: Springer; 2021.
25. Lodders A, Paterson JM. Scrutinising COVIDSafe: frameworks for evaluating digital contact tracing technologies. Alt Law J 2020 Aug 18;45(3):153-161. [doi: 10.1177/1037969x20948262]
26. Sanwikarja P. Contact tracing: How do you design an app millions of people will trust? UX Collective. 2020 May 25. URL: https://uxdesign.cc/how-do-you-design-an-app-millions-of-people-will-trust-8a63f5a5660a [accessed 2020-07-02]
27. Kasali F, Taiwo O, Akinyemi I, Alaba O, Awodele O, Kuyoro S. An enhanced usability model for mobile health application. Int J Comput Sci Inf Security 2019 Feb;17(2):20-29.

28. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth app trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
29. Harrington C, Ruzic L, Sanford J. Universally accessible mHealth apps for older adults: towards increasing adoption and sustained engagement. In: Antona M, Stephanidis C, editors. Universal Access in Human–Computer Interaction. Human and Technological Environments. UAHCI 2017. Lecture Notes in Computer Science, vol 10279. Cham: Springer; 2017.
30. Accessibility requirements for ICT products and services. EN 301 549 V2.1.2. ETSI. 2018 Aug. URL: https://www.etsi.org/deliver/etsi_en/301500_301599/301549/02.01.02_60/en_301549v020102p.pdf [accessed 2020-07-17]
31. Xcertia mHealth App Guidelines. Healthcare Information and Management Systems Society (HIMSS). 2019 Aug 12. URL: https://www.himss.org/sites/hde/files/media/file/2020/04/17/xcertia-guidelines-2019-final.pdf [accessed 2020-09-12]
32. García-Vidal J. Interoperable Digital Proximity Tracing Protocol (IDPT). UP Commons. URL: https://upcommons.upc.edu/bitstream/handle/2117/189356/IDPT-v2.pdf;jsessionid=780E56168F0398C4A8864AFA238578F6?sequence=1 [accessed 2020-07-29]
33. Rossignol J. Apple and Google reveal how COVID-19 exposure notification apps will function. MacRumors. 2020. URL: https://www.macrumors.com/2020/05/04/apple-google-covid-19-app-resources/ [accessed 2020-07-25]
34. Clover J. Apple's exposure notification system: everything you need to know. MacRumors. 2021 Oct 08. URL: https://www.macrumors.com/guide/exposure-notification/ [accessed 2020-07-23]
35. Principles for legislators on the implementation of new technologies. Irish Council for Civil Liberties (ICCL). 2020 Jun 03. URL: https://www.iccl.ie/principles-for-legislators-on-the-implementation-of-new-technologies/ [accessed 2020-07-12]
36. HSE Covid Tracker App: Pre-Release Report Card. Irish Council for Civil Liberties (ICCL). URL: https://www.iccl.ie/wp-content/uploads/2020/07/ICCL-DRI-HSE-App-Pre-Release-Report-Card.pdf [accessed 2020-08-22]
37. Mobile applications to support contact tracing in the EU's fight against COVID-19. Progress reporting June 2020. European Commission. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/mobileapps_202006progressreport_en.pdf [accessed 2020-07-29]
38. COVID-19 exposure notification using bluetooth low energy. Apple/Google. URL: https://blog.google/documents/66/Overview_of_COVID-19_Contact_Tracing_Using_BLE_1.pdf [accessed 2020-08-02]
39. Lillington K. Use Covid tracker app if you wish, but be sure to wear a face covering. The Irish Times. URL: https://www.irishtimes.com/opinion/use-covid-tracker-app-if-you-wish-but-be-sure-to-wear-a-face-covering-1.4298624 [accessed 2020-08-10]
40. Criddle C, Kelion L. Coronavirus contact-tracing: World split between two types of app - BBC News. BBC. URL: https://www.bbc.com/news/technology-52355028 [accessed 2020-06-23]
41. Moss A, Spelliscy C, Borthwick J. Demonstrating 15 contact tracing and other tools built to mitigate the impact of COVID-19. TechCrunch. 2020 Jun 05. URL: https://techcrunch.com/2020/06/05/demonstrating-15-contact-tracing-and-other-tools-built-to-mitigate-the-impact-of-covid-19/?guccounter=1 [accessed 2020-06-23]
42. COVID-19 contact tracing apps: Here is how different countries are taking different approaches. Firstpost. 2020 May 15. URL: https://www.firstpost.com/tech/news-analysis/covid-19-contact-tracing-apps-here-is-how-different-countries-are-taking-different-approaches-8369731.html [accessed 2020-06-23]
43. O'Neill P, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. MIT Technology Review. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2020-06-24]
44. Kaur E, Haghighi P. A context-aware usability model for mobile health applications. 2016 Presented at: MoMM '16: 14th International Conference on Advances in Mobile Computing and Multimedia; November 28-30, 2016; Singapore. [doi: 10.1145/3007120.3007135]
45. Kascak L, Rebola C, Sanford J. Integrating Universal Design (UD) principles and mobile design guidelines to improve design of mobile health applications for older adults. 2014 Presented at: IEEE International Conference of Healthcare Informatics; September 15-17, 2014; Verona, Italy. [doi: 10.1109/ichi.2014.54]
46. Ballantyne M, Jha A, Jacobsen A, Scott HJ, El-Glaly Y. Study of accessibility guidelines of mobile applications. 2018 Presented at: MUM 2018: 17th International Conference on Mobile and Ubiquitous Multimedia; November 25-28, 2018; Cairo, Egypt. [doi: 10.1145/3282894.3282921]
47. Alturki R, Gay V. Usability attributes for mobile applications: a systematic review. In: Jan M, Khan F, Alam M, editors. Recent Trends and Advances in Wireless and IoT-enabled Networks. EAI/Springer Innovations in Communication and Computing. Cham: Springer; 2019.
48. Goel S, Nagpal R, Mehrotra D. Mobile applications usability parameters: taking an insight view. In: Mishra D, Nayak M, Joshi A, editors. Information and Communication Technology for Sustainable Development. Lecture Notes in Networks and Systems, vol 9. Singapore: Springer; 2018:35-43.
49. Shitkova M, Holler J, Heide T, Clever N, Becker J. Towards usability guidelines for mobile websites and applications. 2015 Presented at: 12th International Conference on Wirtschaftsinformatik; March 2015; Osnabrück, Germany.

50. Weichbroth P. Usability of mobile applications: a systematic literature study. IEEE Access 2020;8:55563-55577. [doi: 10.1109/access.2020.2981892]
51. Alonso-Ríos D, Vázquez-García A, Mosqueira-Rey E, Moret-Bonillo V. Usability: a critical analysis and a taxonomy. Int J Hum-Comput Interact 2010 Jan 06;26(1):53-74. [doi: 10.1080/10447310903025552]
52. Soni N, Aloba A, Morga K, Wisniewski P, Anthony L. A framework of Touchscreen interaction design recommendations for children (TIDRC): Characterizing the gap between research evidence and design practice. 2019 Presented at: IDC '19: Proceedings of the 18th ACM International Conference on Interaction Design and Children; June 12-15, 2019; Boise, ID p. 419-431. [doi: 10.1145/3311927.3323149]
53. Gelman D. Design for kids: digital products for playing and learning. 1st ed. Brooklyn, New York: Rosenfeld Media; 2014.
54. Art. 8 GDPR - Conditions applicable to child's consent in relation to information society services. Intersoft Consulting. URL: https://gdpr-info.eu/art-8-gdpr/ [accessed 2020-10-01]
55. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. Eur-Lex. URL: https://eur-lex.europa.eu/eli/dir/2016/2102/oj [accessed 2020-07-30]
56. Web Content Accessibility Guidelines (WCAG) Overview. Web Accessibility Initiative (WAI), W3C. URL: https://www.w3.org/WAI/standards-guidelines/wcag/ [accessed 2020-08-01]
57. Silver A. Improving The Color Accessibility For Color-Blind Users. Smashing Magazine. 2016 Jun 21. URL: https://www.smashingmagazine.com/2016/06/improving-color-accessibility-for-color-blind-users/ [accessed 2020-08-23]
58. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19 epidemic: an assessment framework for apps from an epidemiological and legal perspective. Swiss Med Wkly 2020 May 04;150:w20282. [doi: 10.4414/smw.2020.20282] [Medline: 32418194]
59. Lo B, Sim I. Ethical framework for assessing manual and digital contact tracing for COVID-19. Ann Intern Med 2021 Mar;174(3):395-400. [doi: 10.7326/m20-5834]
60. Righi C, James J, Beasley M, Day D, Fox J, Gieber J, et al. Card sort analysis best practices. J Usability Stud 2013;8(3):89 [FREE Full text]
61. Nielsen J. Usability engineering. Morgan Kaufmann: San Francisco; 1993.
62. Norman D. The design of everyday things: revised and expanded edition. 2nd edition. New York: Basic Books; 2013.
63. Preece J, Sharp H, Rogers Y. Interaction design: beyond human-computer interaction. 4th edition. Chicago: John Wiley & Sons; 2015.
64. Storni C, Tsvyatkova D, Richardson I, Buckley J, Abbas M, Beecham S, et al. Toward a compare and contrast framework for COVID-19 contact tracing mobile applications: a look at usability. 2021 Presented at: 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021); February 11-14, 2021; online. [doi: 10.5220/0010307005570565]
65. Ayala-Rivera V, Pasquale L. The grace period has ended: An approach to operationalize GDPR requirements. 2018 Presented at: IEEE International Conference on Requirements Engineering; August 20-24, 2018; Banff, AB. [doi: 10.1109/re.2018.00023]
66. General Data Protection Regulation (GDPR). Intersoft Consulting. URL: https://gdpr-info.eu/ [accessed 2020-08-25]
67. Know Your Obligations. Data Protection Commission. URL: https://www.dataprotection.ie/en/organisations/know-your-obligations [accessed 2020-07-27]
68. Guide to the UK General Data Protection Regulation (UK GDPR). Information Commissioner's Office (ICO). URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/ [accessed 2020-07-27]
69. Horák M, Stupka V, Husák M. GDPR compliance in cybersecurity software: a case study of DPIA in information sharing platform. 2019 Presented at: ARES '19th International Conference on Availability, Reliability, and Security; August 26-29, 2019; Canterbury CA, UK. [doi: 10.1145/3339252.3340516]
70. Rios E, Iturbe E, Larrucea X, Rak M, Mallouli W, Dominiak J, et al. Service level agreement‐based GDPR compliance and security assurance in(multi)Cloud‐based systems. IET Softw 2019 Jun;13(3):213-222. [doi: 10.1049/iet-sen.2018.5293]
71. Torre D, Soltana G, Sabetzadeh M, Briand L, Auffinger Y, Goes P. Using models to enable compliance checking against the GDPR: an experience report. 2019 Presented at: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS); September 15-20, 2019; Munich, Germany. [doi: 10.1109/models.2019.00-20]
72. Robol M, Salnitri M, Giorgini P. Toward GDPR-compliant socio-technical systems: modeling language reasoning framework. In: Poels G, Gailly F, Serral Asenio E, Snoeck M, editors. The practice of enterprise modeling. Verlag: Springer; 2017:236-250.
73. Tom J, Sing E, Matulevičius R. Conceptual representation of the GDPR: model and application directions. In: Zdravkovic J, Grabis J, Nurcan S, Stirna J, editors. Perspectives in business informatics research. Verlag: Springer; 2018:18-28.
74. Allegue S, Rhahla M, Abdellatif T. Toward GDPR compliance in IoT systems. In: Yangui S, editor. Service-Oriented Computing – ICSOC 2019 Workshops. ICSOC 2019. Lecture Notes in Computer Science, vol 12019. Cham: Springer; 2020:130-141.

75. Tamburri D. Design principles for the General Data Protection Regulation (GDPR): a formal concept analysis and its evaluation. Inf Syst 2020 Jul;91:101469. [doi: 10.1016/j.is.2019.101469]
76. Kammueller F. Formal modeling and analysis of data protection for GDPR compliance of IoT healthcare systems. 2018 Presented at: 018 IEEE International Conference on Systems, Man, and Cybernetics (SMC); October 7-10, 2018; Miyazaki, Japan p. 3319-3324. [doi: 10.1109/smc.2018.00562]
77. Lueks W, Benzler J, Bogdanov D, Kirchner G, Lucas R, Oliveira R, et al. Toward a common performance and effectiveness terminology for digital proximity tracing applications. Front Digit Health 2021 Aug 5;3:677929 [FREE Full text] [doi: 10.3389/fdgth.2021.677929] [Medline: 34713149]
78. Exposure notifications: using technology to help public health authorities fight COVID-19. Google. URL: https://www.google.com/covid19/exposurenotifications/ [accessed 2020-08-20]
79. Rekanar K, O'Keeffe IR, Buckley S, Abbas M, Beecham S, Chochlov M, et al. Sentiment analysis of user feedback on the HSE's Covid-19 contact tracing app. Ir J Med Sci 2022 Feb 18;191(1):103-112 [FREE Full text] [doi: 10.1007/s11845-021-02529-y] [Medline: 33604836]
80. SwissCovid. Google Play. URL: https://play.google.com/store/apps/details?id=ch.admin.bag.dp3t&hl=en [accessed 2020-07-20]
81. Apturi Covid. Ministry of Health and Centre for Disease Prevention and Control. URL: https://apturicovid.lv/#en [accessed 2020-06-27]
82. Immuni. Government of Italy. URL: https://www.immuni.italia.it/ [accessed 2020-07-20]
83. Troncoso C, Payer M, Hubaux J, Salathé M, Larus J, Bugnion E, et al. DP-3T documents: Decentralized Privacy-Preserving Proximity Tracing. GitHub. URL: https://github.com/DP-3T/documents [accessed 2020-08-21]
84. Decentralized privacy-preserving proximity tracing. Wikipedia. URL: https://en.wikipedia.org/wiki/Decentralized_Privacy-Preserving_Proximity_Tracing [accessed 2020-06-18]
85. COVID-19 apps. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_apps [accessed 2020-06-18]
86. Covid Tracker App. Ireland Health Service Executive (HSE). URL: https://www.hse.ie/eng/services/news/newsfeatures/covid19-updates/covid-tracker-app/covid-tracker-app.html [accessed 2020-06-30]
87. COVID Safe Paths. PathCheck Foundation. URL: https://www.pathcheck.org/ [accessed 2020-06-27]
88. NOVID. URL: https://www.novid.org/ [accessed 2020-06-16]
89. Corona-Warn-App Open Source Project. 2020. URL: https://www.coronawarn.app/en/ [accessed 2020-06-29]
90. Aman. 2020. URL: https://amanapp.jo/en [accessed 2020-07-01]
91. Safe Paths. MIT Media Lab. URL: https://www.media.mit.edu/projects/safepaths/overview/ [accessed 2020-09-25]
92. Kasi N. Implications of an assigned devil's advocate role in a negotiations context. Thesis. Penn Libraries. 2018. URL: https://repository.upenn.edu/wharton_research_scholars/172/ [accessed 2020-07-27]
93. Digital Contact Tracing Taxonomy. URL: https://site-3663423-408-3579.mystrikingly.com/ [accessed 2021-01-30]
94. Buckley J, Storni C. The challenge in striking the right balance with Covid apps. RTE. URL: https://www.rte.ie/brainstorm/2021/0407/1208324-contact-tracing-apps-challenges/ [accessed 2021-05-26]
95. User interface elements. Usability.gov. URL: https://www.usability.gov/how-to-and-tools/methods/user-interface-elements.html [accessed 2020-09-20]
96. Welsh T, Rekanar K, Abbas M, Chochlov M, Fitzgerald B, Glynn L, et al. Towards a taxonomy for evaluating societal concerns of contact tracing Apps. 2020 Presented at: 7th International Conference on Behavior, Economic and Social Computing (BESC); November 5-7, 2020; Bournemouth, UK p. 5-7. [doi: 10.1109/besc51023.2020.9348293]
97. MyTrace, a Preventive Counter Measure and Contact Tracing Application for COVID-19. Malaysia Ministry of Science and Technological Innovation. URL: https://www.mosti.gov.my/web/en/mytrace/ [accessed 2020-07-25]
98. Policy briefs. Swiss National COVID-19 Science Task Force. 2020. URL: https://sciencetaskforce.ch/policy-briefs/ [accessed 2020-10-20]
99. HSEIreland covid-tracker-app. GitHub. 2020. URL: https://github.com/HSEIreland/covid-tracker-app/pulls [accessed 2020-07-20]
100. Indicator framework to evaluate the public health effectiveness of digital proximity tracing solutions. Geneva, Switzerland: World Health Organization; 2021.
101. Regmi K. Triangulation in healthcare research: what does it achieve? In: SAGE Research Methods Cases Part 1. Thousand Oaks, CA: SAGE Publications Ltd; 2014.
102. Rutherford GW, McFarland W, Spindler H, White K, Patel SV, Aberle-Grasse J, et al. Public health triangulation: approach and application to synthesizing data to understand national and local HIV epidemics. BMC Public Health 2010 Jul 29;10:447 [FREE Full text] [doi: 10.1186/1471-2458-10-447] [Medline: 20670448]

Abbreviations

API: application programming interface
C^{3}EF: Citizen-Focused Compare-and-Contrast Evaluation Framework
CDC: European Centre for Disease Control

CTA: contact tracing app
CWE: Common Weakness Enumerator
DP-3T: Decentralized Privacy Preserving Proximity Tracing
GAEN: Google and Apple Exposure Notification
GDPR: General Data Protection Regulations
HSE: Health Service Executive (Ireland)
mHealth: mobile health
MIT: Massachusetts Institute of Technology
SDLC: Software Development Life Cycle
STRIDE: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege
UD: Universal Design

Edited by L Buis; submitted 25.05.21; peer-reviewed by V von Wyl, JP Weiß, C Jonassaint; comments to author 21.07.21; revised version received 31.10.21; accepted 15.12.21; published 11.03.22.

Please cite as:
Tsvyatkova D, Buckley J, Beecham S, Chochlov M, O'Keeffe IR, Razzaq A, Rekanar K, Richardson I, Welsh T, Storni C, COVIGILANT Group
Digital Contact Tracing Apps for COVID-19: Development of a Citizen-Centered Evaluation Framework
JMIR Mhealth Uhealth 2022;10(3):e30691
URL: https://mhealth.jmir.org/2022/3/e30691
doi:10.2196/30691
PMID:35084338

Original Paper

Real-World Effectiveness of Wearable Augmented Reality Device for Patients With Hearing Loss: Prospective Study

Ul Gyu Han[1,2*], BS; Jung-Yup Lee[3*], MD; Ga-Young Kim[2], PhD; Mini Jo[2], BS; Jaeseong Lee[4,5], PhD; Kyoung Ho Bang[4,5], PhD; Young Sang Cho[3], MD; Sung Hwa Hong[6], MD, PhD; Il Joon Moon[2,3], MD, PhD

[1]Samsung Advanced Institute for Health Sciences & Technology, Department of Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

[2]Hearing Research Laboratory, Samsung Medical Center, Seoul, Republic of Korea

[3]Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[4]Advanced Lab - Audio, Samsung Electronics, Suwon, Republic of Korea

[5]Department of Electrical and Electronic Engineering, Yonsei University, Seodaemun-gu, Seoul, Republic of Korea

[6]Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Changwon Hospital, Sungkyunkwan University School of Medicine, Changwon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Il Joon Moon, MD, PhD
Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 10 3789 7505
Email: moon.iljoon@gmail.com

Abstract

Background: Hearing loss limits communication and social activity, and hearing aids (HAs) are an efficient rehabilitative option for improving oral communication and speech comprehension, as well as the psychosocial comfort of people with hearing loss. To overcome this problem, over-the-counter amplification devices including personal sound amplification products and wearable augmented reality devices (WARDs) have been introduced.

Objective: This study aimed to evaluate the clinical effectiveness of WARDs for patients with mild to moderate hearing loss.

Methods: A total of 40 patients (18 men and 22 women) with mild to moderate hearing loss were enrolled prospectively in this study. All participants were instructed to wear a WARD, Galaxy Buds Pro (Samsung Electronics), at least 4 hours a day for 2 weeks, for amplifying ambient sounds. Questionnaires including the Korean version of the abbreviated profile of hearing aid benefit (K-APHAB) and the Korean adaptation of the international outcome inventory for hearing aids (K-IOI-HA) were used to assess personal satisfaction in all participants. Audiologic tests, including sound field audiometry, sound field word recognition score (WRS), and the Korean version of hearing in noise test (K-HINT), were administered to 14 of 40 patients. The tests were performed under two conditions: unaided and aided with WARDs.

Results: The mean age of the participants was 55.4 (SD 10.7) years. After 2 weeks of the field trial, participants demonstrated a benefit of WARDs on the K-APHAB. Scores of 3 subscales of ease of communication, reverberation, and background noise were improved significantly (P<.001). However, scores regarding aversiveness were worse under the aided condition (P<.001). K-IOI-HA findings indicated high user satisfaction after the 2-week field trial. On audiologic evaluation, the K-HINT did not show significant differences between unaided and aided conditions (P=.97). However, the hearing threshold on sound field audiometry (P=.001) and the WRS (P=.002) showed significant improvements under the aided condition.

Conclusions: WARDs can be beneficial for patients with mild to moderate hearing loss as a cost-effective alternative to conventional hearing aids.

(JMIR Mhealth Uhealth 2022;10(3):e33476) doi:10.2196/33476

KEYWORDS

hearing loss; hearing aids; personal sound amplification product; wearable augmented reality device

Introduction

Background

Hearing loss limits communication and social activity, leading to disorders in language and cognitive impairment [1]. According to the World Health Organization, approximately 5% of the world's population has hearing loss, and the number is anticipated to increase to one in every four people by 2050 because of rapidly aging populations [2]. Oral communication is crucial for contact with other people, but people with hearing loss have reduced speech understanding compared to people with normal hearing. Therefore, active hearing rehabilitation is needed for people with hearing loss [3].

Hearing aids (HAs) are an efficient rehabilitative option for improving oral communication and speech comprehension as well as the psychosocial comfort of people with hearing loss [4]. Although the benefits of HAs have been well documented, the uptake rate of HAs remains relatively low [5]. Furthermore, only 0.47% of individuals with minimal hearing loss use HAs even when experiencing subjective hearing difficulty [6]. One of the main reasons for low uptake of HAs is high cost. The average set of HAs costs from US $1000-$5000, which can inflict a financial burden on many individuals with hearing loss [7]. According to the MarkeTrack VIII survey, some consumers with mild to moderate hearing loss said that they would adopt HAs if the price did not exceed a certain level or if they were covered by insurance [8].

To overcome this problem, over-the-counter (OTC) amplification devices including personal sound amplification products (PSAPs) have been introduced. PSAPs are defined by the US Food and Drug Administration as wearable electronic products for customers with hearing loss to amplify sounds in certain environments. In general, PSAPs are less expensive and simpler sound amplification devices with fewer features and less functionality than digital HAs. However, some studies have suggested some kinds of PSAPs as alternative devices for those with mild to moderate hearing loss [9]. In addition, we reported that wearable augmented reality devices (WARDs) with a broad spectrum of "hearable" have the potential to be beneficial for individuals with hearing loss. WARDs are a combination of smartphone apps and earbuds, providing a personalized listening experience. For example, the Samsung Galaxy Buds Pro has its own smartphone app called Galaxy Wearable. Users can take advantage of a feature called ambient sound. Similar to PSAPs, individuals can manage the level of sounds in their surroundings such as crowded restaurant or sidewalk with many cars. They can also reduce background noise and listen to music on the street or subway. WARDs helped people with mild to moderate hearing loss to understand conversations in quiet environments [10].

Objectives

Although most previous studies evaluated clinical effectiveness of PSAPs compared to conventional HAs, there are insufficient data on the WARD's ability to help people with hearing loss. To the best of our knowledge, no clinical field trial assessing the effectiveness of WARDs in the daily lives of hearing-impaired people has been conducted. Thus, the aims of this study were to investigate the hearing outcomes in patients with mild to moderate hearing loss aided with WARDs and to quantify the patient's subjective outcomes using the Korean version of the abbreviated profile of hearing aid benefit (K-APHAB) and the Korean adaptation of the international outcome inventory for hearing aids (K-IOI-HA) questionnaires in 2-week field trials. Furthermore, we attempted to assess the correlation between personal satisfaction and audiologic performance with WARDs.

Methods

Participants

The sample size was determined on the basis of previous research determining the effect of a web-based intervention program on positive changes in hearing aid use [11]. The resulting sample size was 21, using G*Power 3.1.9.7 for power set at 0.95 and α set at .05. A total of 40 individuals with mild to moderate hearing loss were enrolled in the study. A prospective study was conducted with subjects who visited the outpatient clinic of the department of otolaryngology for hearing loss from February to May 2021. The subjects who met the following appropriateness criteria were included: patients between 18 and 70 years of age who had bilateral mild to moderate hearing loss (26-55 dB hearing level [HL]; pure tone average 500-4000 Hz) and who were determined to have no abnormalities in the eardrum on otoscopy. The exclusion criteria were difficulty of communication or inspection and inability to handle the device.

Ethical Considerations

This study was approved by the institutional review board of Samsung Medical Center in Seoul, South Korea (2020-05-052, 2020-10-163), and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants.

Intervention

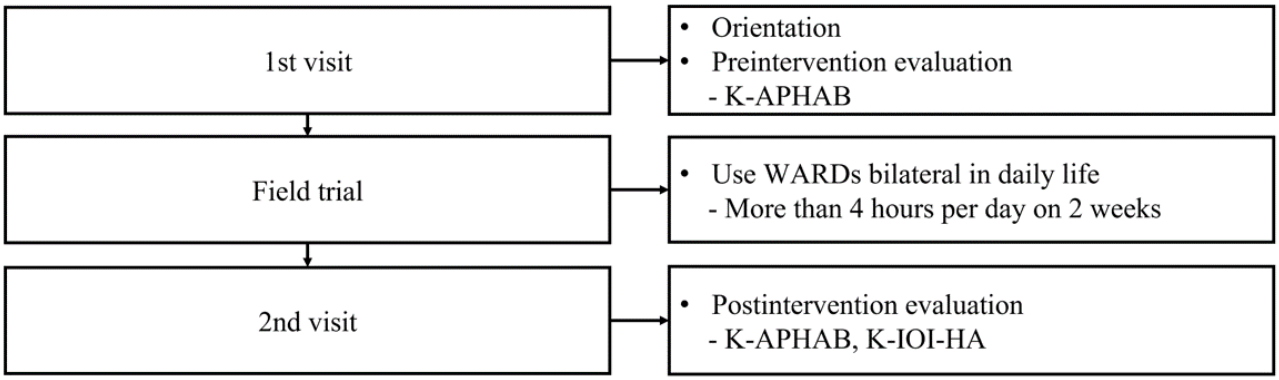
Galaxy Buds Pro (SM-R175, Samsung Electronics) was used for hearing rehabilitation in this study. Galaxy Buds Pro has its own smartphone app, Galaxy Wearable. Users can use the Ambient Sound feature with the app for sound amplification. Similar to PSAPs or HAs, users can control the level of sound in their surroundings using Galaxy Buds Pro. The Galaxy wearable device consists of 4 levels, of which only level 4 provided sound amplification. Therefore, the level was set at 4 in this study.

Each participant was provided with a pair of Galaxy Buds Pro for this field trial and taught how to use the device. Participants were required to use the device when having difficulties in communication or listening and for more than 4 hours a day for 2 weeks. They wore the device during daily activity such as conversation, TV watching, or driving.

We also recommended to stop wearing the devices when if the participants feel any pain or have any other troubles in their ears. Before the beginning of the field trial, participants filled out the K-APHAB questionnaire as a baseline [12].

After 2 weeks, participants returned the device and filled out the K-APHAB and the K-IOI-HA questionnaires to assess the benefit of using the device [13]. The APHAB is one of the most commonly used questionnaires to assess the benefit of HAs and often is cited for its ease of understanding and delivery [14]. The K-APHAB consists of 24 questions divided into 4 subscales that measure hearing loss in everyday situations. The ease of communication (EC) subscale examines basic hearing situations without ambient noise in a quiet environment, the background noise (BN) subscale examines hearing situations with background noise, the reverberation (RV) subscale investigates hearing situations in large spaces with echoes, and the aversiveness (AV) subscale measures the perception of loud sound events [15]. Global scores are calculated as the average of the EC, BN, and RV subscale scores [16]. Higher scores reflect a greater rate of problems. In general, HA benefit as indicated by K-APHAB is calculated as unaided scores minus aided scores and is represented by a positive value. We utilized the K-APHAB for evaluating Galaxy Buds Pro benefits. The IOI-HA was designed to formulate a standardized and internationally useful self-report measurement. A self-report measurement is necessary to acquire quantifiable data on the effects of HAs in users' daily lives [17]. Similar to the APHAB, the IOI-HA has been utilized to investigate an aspect of the personal impact of hearing rehabilitation devices [18]. The K-IOI-HA contains seven questions used to subjectively evaluate HA performance using these parameters: (1) duration of HA use (USE), (2) benefit (BEN), (3) residual limitation in daily life activities (RAL), (4) satisfaction (SAT), (5) residual restrictions to participation (RPR), (6) impact on other people (IO), and (7) quality of life (QOL). Patients select one of five responses. Therefore, each question can be scored from 1 to 5 points, and the total score ranges from 7 to 35 points, with a high score indicating a positive HA effect. Furthermore, we divided two subscales (Factors 1 and 2) within the K-IOI-HA when performing a principal component analysis. Factor 1 included USE, BEN, SAT, and QOL; Factor 2 included RAL, RPR, and IO. Factor 1 described the overall benefit with WARD, and Factor 2 described the residual limitations after WARD fitting [19]. We utilized the K-IOI-HA for evaluating an outcome with Galaxy Buds Pro. A flowchart of the 2-week field study is presented in Figure 1.

Figure 1. Flowchart of the 2-week field study. Subjects were required to use the WARD more than 4 hours a day for 2 weeks. At the second visit, questionnaires were completed by all subjects. K-APHAB: Korean version of abbreviated profile of hearing aid benefit, K-IOI-HA: Korean adaptation of the international outcome inventory for hearing aids, WARD: wearable augmented reality device.



Audiologic Evaluation

Three audiologic test batteries were administered in this study: (1) sound field audiometry, (2) sound-field word recognition score (WRS), and (3) the Korean version of hearing in noise test (K-HINT). The associated measurements were conducted with 14 subjects who agreed to participate in accordance with institutional review board approval.

Unaided and aided thresholds were obtained in sound field audiometry. Warble tones of 0.25, 0.5, 1, 2, 3, 4, and 6 kHz were presented through a loudspeaker located 1 m from the participant. The participant wore the WARD in both ears to mimic how the device would be used in his/her daily life. Speech perception with and without the WARD was evaluated through sound-field WRS. In total, 25 monosyllabic words from the Korean standard monosyllabic word list (KS-MWL-A) were presented at 50 dB HL through a loudspeaker located 1 m from the participant [20]. The participant was asked to repeat the word back to the tester. The percentage of correct scores was calculated. Last, K-HINT was performed to assess speech recognition in the presence of noise. The participant sat on a chair in the center of the sound field, facing a loudspeaker that was located approximately 1 m away at the 0˚ azimuth. The target sentences in the K-HINT and speech-shaped noise were presented by the loudspeaker at a fixed level of 65 dBA. The presentation level of the target speech was adjusted to measure a signal-to-noise ratio (SNR) at which the participant recognized the sentences 50% of the time.

Statistical Analysis

All statistical analyses were completed using SPSS (version 26; IBM Corp). The paired 2-tailed *t* test was conducted to compare the scores of questionnaires before and after intervention. The paired *t* test also was used to compare the variables of audiological measurements between unaided and aided conditions. In addition, Pearson correlation coefficients were calculated to further investigate relationships between scores on the questionnaires and laboratory assessments. A significance level of P=.05 was applied to determine statistical significance.

Results

Demographics

A total of 40 participants (18 male and 22 female; mean age 55.4, SD 10.73, range 28-67 years) with mild to moderate hearing loss were enrolled. The mean hearing threshold in pure-tone average was 40.75 (SD 6.63) dB on the right side and 41.16 (SD 7.93) dB on the left side (Figure 2).

The demographic characteristics of the enrolled patients are summarized in Table 1. If previous usage durations of HAs or PSAPs in each ear were different, the average usage duration of the 2 ears was calculated.

Figure 2. Audiogram representing the mean hearing threshold of all participants (N=40). Hearing threshold is the minimum volume of sound that participant can hear at a specific frequency, indicated by red "O" (right ear) or blue "X" (left ear). HL: hearing level.

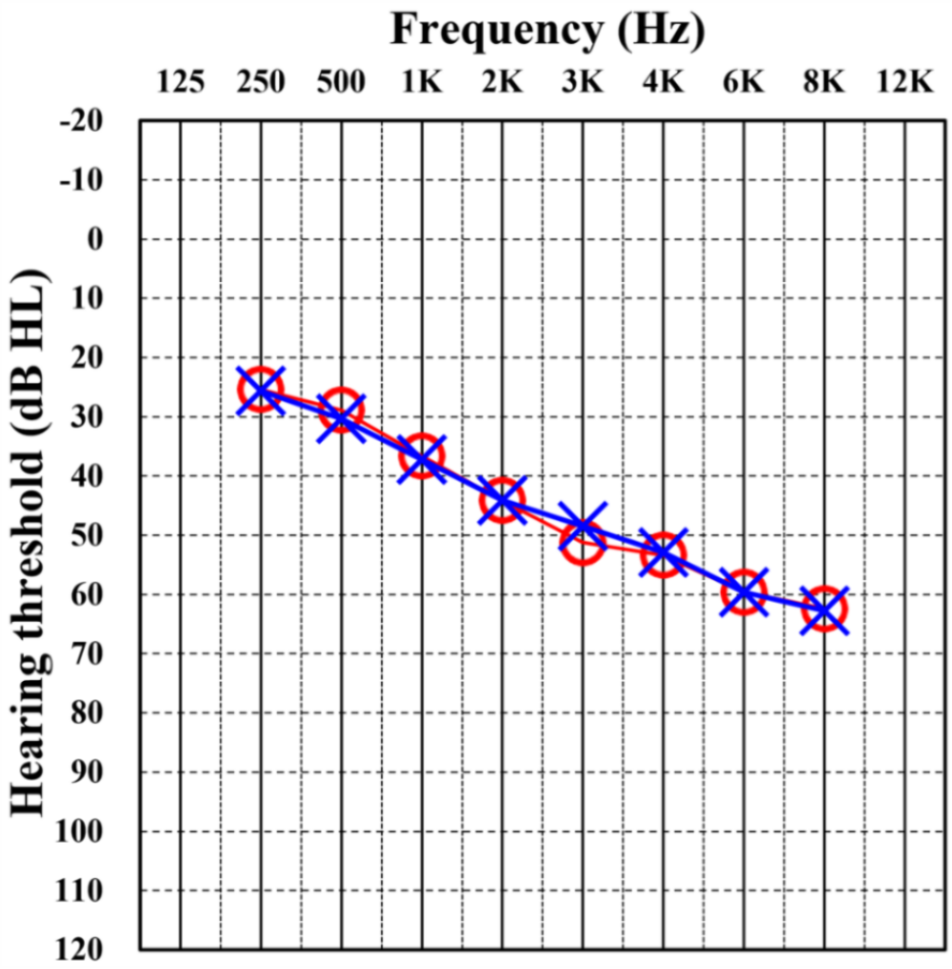


Table 1. Demographic information of all participants (N=40).

Variable	Value
Age (years), mean (SD)	55.4 (10.73)
Sex, n (%)	
Male	18 (45)
Female	22 (55)
Hearing threshold (dB hearing level), mean (SD)	
Right	40.75 (6.63)
Left	41.16 (7.93)
Previous hearing aid users	
Direction (Unilateral:Bilateral), n	3:4
Type (invisible in canal:receiver in canal:in the canal:receiver in ear:complete in canal), n	2:2:1:1:1
Usage duration (months), mean (SD)	8.57 (9.07)
Previous personal sound amplification product user, n	
Direction (Unilateral:Bilateral), n	2:0
Type (complete in canal:receiver in canal), n	1:1
Usage duration (months), mean (SD)	2.25 (1.77)

Questionnaires

The K-APHAB results for all 40 subjects under unaided and aided conditions are shown in Figure 3. The EC subscale under unaided (Pre EC) and aided conditions (Post EC) was 42.30 points (SD 21.04) and 20.15 points (SD 14.56), respectively. The EC subscale was significantly decreased with the WARD (P<.001). The RV subscale before using the WARD (Pre RV) was 51.58 points (SD 20.49). Under the aided condition, the RV subscale (Post RV) was 27.35 (SD 13.97) points, which was significantly improved (P<.001). There was a significant difference between the BN subscale under unaided (Pre BN) and aided conditions (Post BN) (P<.001). Pre BN score was 48.28 (SD 18.75) points, while the Post BN score was 35.17 (SD 18.18) points. Additionally, the global score (average of EC, RV, and BN subscales) showed significant improvement under the aided condition compared with the unaided condition (P<.001). In contrast, the AV subscale score in the aided condition (Post AV) was worse than that in the unaided condition (Pre AV). There was a significant difference between Pre AV points and Post AV points (P<.001).

The resulting scores on the K-IOI-HA for the 40 subjects were 3.3 (SD 0.5) points for daily USE, 3.0 (SD 0.9) points for BEN, 3.6 (SD 0.9) points for RAL, 3.0 (SD 1.0) points for SAT with devices and services, 3.7 (SD 0.9) points for RPR, 4.2 (SD 0.9) points for IO, and 3.0 (SD 0.7) points for QOL. In addition, the mean K-IOI-HA Factor 2 score was significantly higher than the Factor 1 score (P<.001; Figure 4).

Figure 3. The results of the K-APHAB. There were significant reductions in the subscales of EC, RV, and BN (*P*<.001). In contrast, significant increase in the subscale of AV was observed in the aided condition as compared to the unaided condition (*P*<.001). AV: aversiveness, BN: background noise, EC: ease of communication, K-APHAB: Korean version of abbreviated profile of hearing aid benefit, Post: aided, Pre: unaided, RV: reverberation.

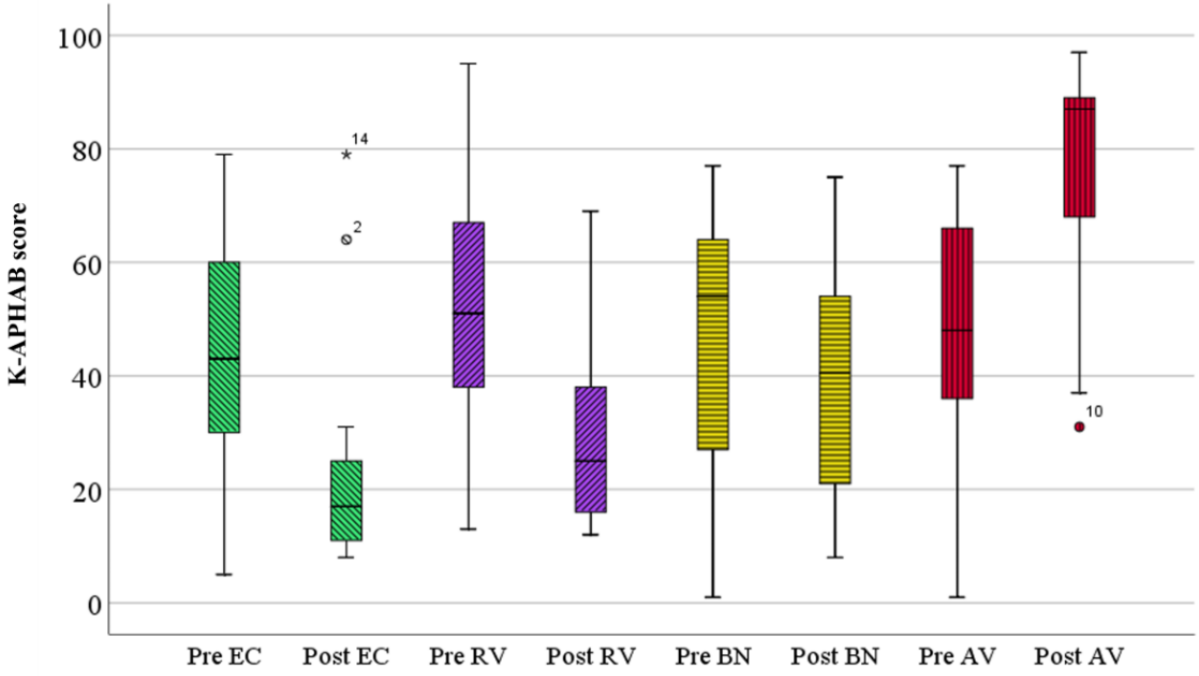
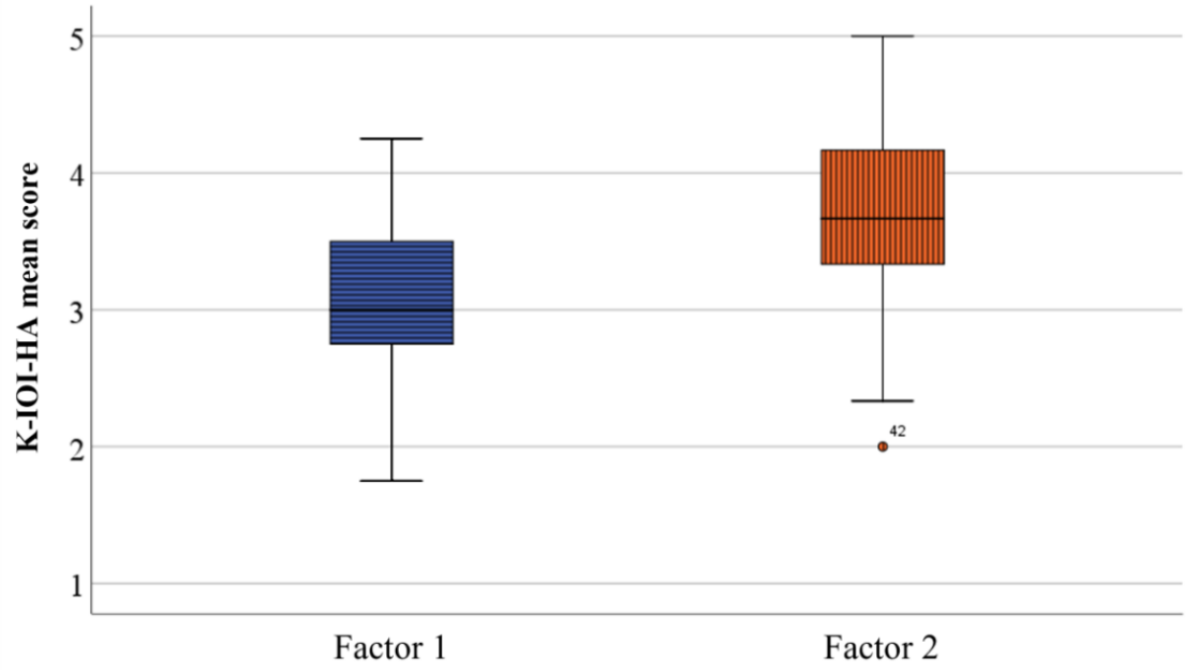


Figure 4. The comparison of the mean score on K-IOI-HA Factor 1 and Factor 2. There was a significant difference between the mean score of Factor 1 (total score of USE, BEN, SAT, and QOL) and the mean score of Factor 2 (total score of RAL, RPR, and IO) (*P*<.001). BEN: benefit, Factor 1: sum score of the USE (daily use), Factor 2: sum score of the RAL (residual activity limitations), IO: impact on others, K-IOI-HA: Korean adaptation of the international outcome inventory for hearing aids; QOL: quality of life, RPR: residual participation restriction, SAT: satisfaction (with the device and services).



Audiologic Measurements

For 14 participants, the average threshold of sound field audiometry under the unaided condition was 43.45 (SD 6.62) dB HL; this was significantly decreased under the aided condition to 40.48 (SD 6.99) dB HL (*P*=.001). Sound field WRS significantly improved in the aided condition from 55.43% (SD 21.45%; responded to ~14 of 25 test questions) to 67.71% (SD 16.11%; responded to ~17 of 25 test questions) (*P*=.002). However, there was no significant differences in the K-HINT score between the unaided and aided conditions (Table 2).

Table 2. Results of audiologic measurements (N=14).

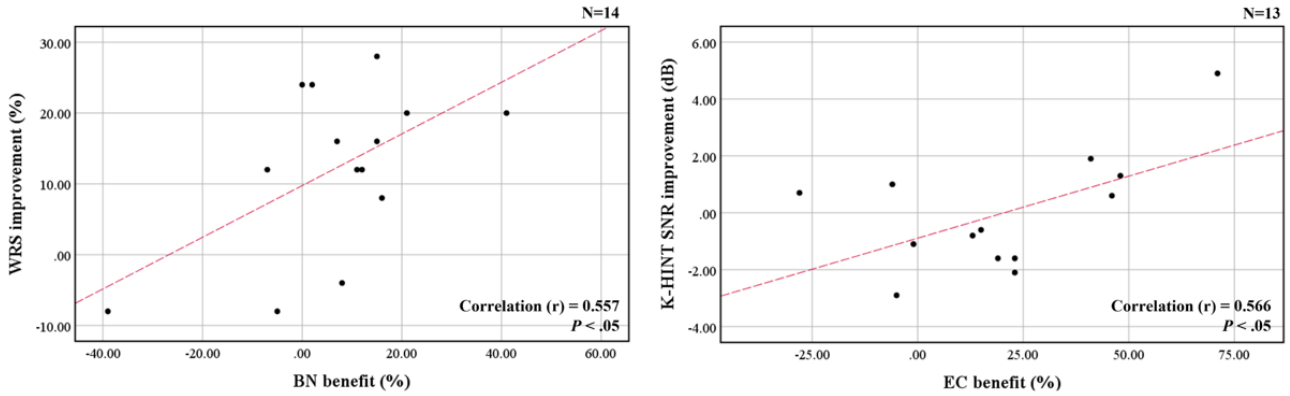
Audiologic measurements	Unaided condition	Using the Galaxy Buds Pro wearable augmented reality device	*P* value
Sound field audiometry threshold[a] (dB hearing level), mean (SD)	43.45 (6.62)	40.48 (6.99)	.001
Word recognition score (%), mean (SD)	55.43 (21.45)	67.71 (16.11)	.002
Korean version of hearing in noise test (dB signal-to-noise ratio), mean (SD)	–0.65 (1.77)	–0.68 (2.10)	.97

[a]Average of hearing thresholds at six frequencies: 0.5, 1, 1, 2, 2, and 4 kHz.

Correlations Between Audiologic Measurements and Patient-Reported Outcome Measurements

Correlation analysis between scores from questionnaire and audiologic evaluation was performed (Figure 5). The EC benefit on the K-APHAB questionnaire showed a significant correlation with dB SNR improvement on the K-HINT (*P*<.05). BN benefit also showed a significant correlation with WRS improvement (*P*<.05). However, EC benefit did not show a significant correlation with WRS improvement (*P*=.99), and BN benefit did not show a significant correlation with dB SNR improvement on the K-HINT (*P*=.20). No additional correlations were found between the parameters.

Figure 5. Correlation between audiologic parameters and personal satisfaction. BN and EC benefit on the K-APHAB each a showed significant correlation with WRS improvement and SNR improvement on the K-HINT. WRS improvement: aided WRS – unaided WRS. BN benefit: aided BN – unaided BN. EC benefit: aided EC – unaided EC. BN: background noise, EC: ease of communication, K-APHAB: Korean version of abbreviated profile of hearing aid benefit, K-HINT: Korean version of hearing in noise test, SNR: signal-to-noise ratio, WRS: word recognition score.



Discussion

Principal Findings

This study investigated personal satisfaction with the WARD among patients with mild to moderate hearing loss. Furthermore, the performance of the WARD was evaluated through audiologic tests including sound field audiometry, WRS, and the K-HINT. The results revealed significantly increased subjective satisfaction with the WARD. Furthermore, sound field audiometry and WRS also showed significant improvement under the aided condition.

In this study, significant improvements were shown in EC, RV, and BN subscales under aid with a WARD. These results indicated that if patients use a WARD, their difficulties in listening were improved in various situations including quiet or easy conversation, loud sounds or sound at a distance, and in the presence of ambient noise. However, the AV subscale was significantly increased under the aided condition: loud and noisy sounds were amplified with a WARD with a resulting increase in discomfort.

Worse scores on the AV subscale occur commonly, even in HA users. Johnson et al [21] reported that AV benefit decreased under both conditions, aided with linear processing type HAs and wide dynamic processing compression HAs. In another study, the AV subscale score was increased under the aided condition with all types of HAs, including those completely in the canal, in the canal, and behind the ear [22]. Ideally, patients with hearing impairment would show increased EC, RV, BN, and AV benefit when wearing HAs. However, various studies have shown that AV benefit is difficult to achieve even with high-quality HAs [23]. In addition to amplifying environmental noise, lack of acclimatization or individual fine fitting can be reasons for this unsatisfactory result [24]. WARDs cannot be tuned individually, which is a limitation of the device. Therefore, further technical development such as individual fitting or advanced artificial intelligence system would be needed to overcome those problems.

In this study, the average of all question scores was 3.0 points or higher. In particular, the IO score was higher than 4 points, indicating decreased inconvenience to others owing to hearing difficulty when using the WARD. Compared with previously published IOI-HA norms, the distribution characteristics of the mean scores were consistent with the normative data [25]. The comparison also indicated that our scores were slightly higher than the normative data for IO score.

The average threshold of sound field audiometry was significantly improved with WARDs in this study. Sound field WRS was increased while wearing a WARD. Thus, hearing-impaired people can receive listening benefit from WARDs under quiet conditions. However, speech recognition in a noisy environment was worse under the aided condition. Reed at al [26] reported that PSAPs improved speech recognition in a noisy environment for those with mild to moderate hearing loss. However, the results of this study were insufficient to demonstrate the benefit of a WARD for improvement of speech intelligibility in a noisy environment. Clarifying the reason for the difference of these results is difficult, and the effectiveness of WARDs in the noisy environment remains uncertain.

Interpreting the correlation between objective and subjective results requires caution. We analyzed the correlation between benefits from K-APHAB scores and parameters from audiologic measurements in this study. These were not significantly correlated, but there were some weak correlations between personal satisfaction and audiologic performance. We confirmed that the improvement in dB SNR on the K-APHAB and the EC benefit on the K-APHAB and that between the BN benefit and the WRS on the K-APHAB had significant correlations (P<.05). We expected the WRS improvement and EC benefit to have a significant correlation because both were derived in a quiet environment. We also expected that dB SNR improvement and BN benefit would have a significant correlation because both results were derived in a noisy environment. However, some results in this study did not agree with our expectations. The group with high satisfaction in a quiet environment showed dB SNR improvement, and the group with high satisfaction in a noisy environment had WRS improvement.

The reason for this correlation analysis can be considered to be due to the difference in the test conditions. First, as mentioned in previous studies, the laboratory-based evaluation method is a static, limited, and 1-way communication method, whereas the evaluation method experienced by patients in everyday life is a dynamic and expanded interactive communication method [27]. Second, the difference between sound and noise should be considered. In the test room, subjects heard the voice of one speaker at a certain volume of speech spectrum noise. However, in everyday life, subjects hear the voices of one or more speakers with a variety of noises. Third, in everyday life, various visual stimuli can be referenced for sound recognition; this was not the case under our test conditions. As such, difficulty arises in considering various factors that can affect the evaluation results in a laboratory environment. Therefore, a number of studies is being conducted to consider various conditions in daily life via VR technology [27,28].

Overall satisfaction with a WARD in daily life was high in this study, and we speculate that this high satisfaction might be influenced by the current COVID-19 pandemic. All Koreans must wear face masks when outdoors. Personal protective equipment such as the facial mask creates difficulty in understanding and communication in hearing impaired patients; these patients cannot read the speaker's lips, and sound clarity is decreased [29]. In this pandemic situation, WARDs could foster conversations.

Limitations

This study has some limitations. The first limitation is the lack of generalizability to the WARDs market. Only one kind of WARD, Galaxy Buds Pro, was used in this study. Even though availability of hearing devices is limited, WARDs vary considerably in style, quality, and technology. Therefore, further research using a variety of hearable devices is needed to generalize the feasibility of WARDs. Second, we did not take into consideration certain subject demographics including the duration of hearing loss, education, income, and perceived social support, which play an important role in determining personal satisfaction. Future large-scale research should consider the contribution of these factors to audiologic outcomes and personal satisfaction related to WARD benefit. Finally, we did not take into account the shape and size of the ear canal. Since the shape and size of human ear canal are very diverse, it might be possible that there were some participants who had WARDs that were either too large or too small [30]. Even though there was no participant who complained about it, the shape and size of WARDs would also be a problem worth investigating.

Conclusions

Our results indicate that WARDs could be helpful for individuals with mild to moderate hearing loss, especially under quiet conditions. Owing to high price and poor accessibility of HAs, OTC hearing devices such as WARDs could be an alternative partial solution for hearing loss. In the near future, WARDs will have greater potential as technology develops and government regulation changes. Further large-scale comparative research regarding the clinical effectiveness of WARDs is necessary.

Acknowledgments

The authors would like to thank to Samsung Electronics for their contributions to this work. The authors rented Samsung Galaxy Buds Pro from Samsung Electronics for this study and received funding included not only renting devices but also financial support for research, authorship, and publication of this article from Samsung Electronics. The study did not receive further funding.

Conflicts of Interest

None declared.

References

1. Dalton DS, Cruickshanks KJ, Klein BEK, Klein R, Wiley TL, Nondahl DM. The impact of hearing loss on quality of life in older adults. Gerontologist 2003 Oct;43(5):661-668. [doi: 10.1093/geront/43.5.661] [Medline: 14570962]
2. WHO: 1 in 4 people projected to have hearing problems by 2050. World Health Organization. 2021 Mar 02. URL: https://www.who.int/news/item/02-03-2021-who-1-in-4-people-projected-to-have-hearing-problems-by-2050 [accessed 2021-05-02]
3. Li LYJ, Wang S, Wu C, Tsai C, Wu T, Lin Y. Screening for hearing impairment in older adults by smartphone-based audiometry, self-perception, HHIE screening questionnaire, and free-field voice test: comparative evaluation of the screening accuracy with standard pure-tone audiometry. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e17213 [FREE Full text] [doi: 10.2196/17213] [Medline: 33107828]
4. Hughes ME, Nkyekyer J, Innes-Brown H, Rossell SL, Sly D, Bhar S, et al. Hearing aid use in older adults with postlingual sensorineural hearing loss: protocol for a prospective cohort study. JMIR Res Protoc 2018 Oct 26;7(10):e174 [FREE Full text] [doi: 10.2196/resprot.9916] [Medline: 30368434]
5. Nkyekyer J, Meyer D, Blamey PJ, Pipingas A, Bhar S. Investigating the impact of hearing aid use and auditory training on cognition, depressive symptoms, and social interaction in adults with hearing loss: protocol for a crossover trial. JMIR Res Protoc 2018 Mar 23;7(3):e85 [FREE Full text] [doi: 10.2196/resprot.8936] [Medline: 29572201]
6. Choi JE, Ahn J, Park HW, Baek S, Kim S, Moon IJ. Prevalence of minimal hearing loss in South Korea. PLoS One 2017;12(2):e0171635 [FREE Full text] [doi: 10.1371/journal.pone.0171635] [Medline: 28196098]
7. What to Know About Current Hearing Aid Costs and Pricing. Healthline. URL: https://www.healthline.com/health/cost-of-hearing-aids [accessed 2021-05-02]
8. Kochkin S. MarkeTrak V: Consumer satisfaction revisited. Hear J 2000;53(1):38. [doi: 10.1097/00025572-200001000-00005]
9. Cho YS, Park SY, Seol HY, Lim JH, Cho Y, Hong SH, et al. Clinical performance evaluation of a personal sound amplification product vs a basic hearing aid and a premium hearing aid. JAMA Otolaryngol Head Neck Surg 2019 Jun 01;145(6):516-522 [FREE Full text] [doi: 10.1001/jamaoto.2019.0667] [Medline: 31095263]
10. Seol HY, Kim G, Kang S, Jo M, Han UG, Cho YS, et al. Clinical comparison of a hearing aid, a personal sound amplification product, and a wearable augmented reality device. Clin Exp Otorhinolaryngol 2021 Aug;14(3):359-361 [FREE Full text] [doi: 10.21053/ceo.2021.00297] [Medline: 33735561]
11. Thorén ES, Oberg M, Wänström G, Andersson G, Lunner T. A randomized controlled trial evaluating the effects of online rehabilitative intervention for adult hearing-aid users. Int J Audiol 2014 Jul;53(7):452-461. [doi: 10.3109/14992027.2014.892643] [Medline: 24749664]
12. Lim HJ, Park MK, Cho Y, Han GC, Choi J, An Y, et al. Validation of the Korean version of the abbreviated profile of hearing aid benefit. Korean J Otorhinolaryngol-Head Neck Surg 2017 Apr;60(4):164-173. [doi: 10.3342/kjorl-hns.2016.17398]
13. Chu H, Cho Y, Park S, Byun JY, Shin JE, Han GC, et al. Standardization for a Korean adaptation of the International Outcome Inventory for Hearing Aids: study of validity and reliability. Korean J Otorhinolaryngol-Head Neck Surg 2012;55(1):20. [doi: 10.3342/kjorl-hns.2012.55.1.20]
14. Cox R, Hyde M, Gatehouse S, Noble W, Dillon H, Bentler R, et al. Optimal outcome measures, research priorities, and international cooperation. Ear Hear 2000 Aug;21(4 Suppl):106S-115S. [doi: 10.1097/00003446-200008001-00014] [Medline: 10981601]
15. Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. Ear Hear 1995 Apr;16(2):176-186. [doi: 10.1097/00003446-199504000-00005] [Medline: 7789669]
16. Dornhoffer J, Meyer T, Dubno J, McRackan T. Assessment of hearing aid benefit using patient-reported outcomes and audiologic measures. Audiol Neurootol 2020;25(4):215-223 [FREE Full text] [doi: 10.1159/000506666] [Medline: 32241007]
17. Cox RM, Alexander GC. The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. Int J Audiol 2002 Jan;41(1):30-35. [doi: 10.3109/14992020209101309] [Medline: 12467367]
18. Atas A, Tutar H, Gunduz B, Bayazıt YA. Vibrant SoundBridge application to middle ear windows versus conventional hearing aids: a comparative study based on international outcome inventory for hearing aids. Eur Arch Otorhinolaryngol 2014 Jan;271(1):35-40. [doi: 10.1007/s00405-013-2387-2] [Medline: 23400404]
19. Cox RM, Stephens D, Kramer SE. Translations of the International Outcome inventory for Hearing Aids (IOI-HA). Int J Audiol 2002 Jan 07;41(1):3-26. [doi: 10.3109/14992020209101307] [Medline: 12467365]
20. Kim J, Lee J, Lee KW, Bahng J, Lee JH, Choi C, et al. Test-retest reliability of word recognition score using Korean standard monosyllabic word lists for adults as a function of the number of test words. J Audiol Otol 2015 Sep;19(2):68-73 [FREE Full text] [doi: 10.7874/jao.2015.19.2.68] [Medline: 26413571]
21. Johnson JA, Cox RM, Alexander GC. Development of APHAB norms for WDRC hearing aids and comparisons with original norms. Ear Hear 2010 Feb;31(1):47-55. [doi: 10.1097/AUD.0b013e3181b8397c] [Medline: 19692903]
22. Quintino CA, Mondelli MFCG, Ferrari DV. Directivity and noise reduction in hearing aids: speech perception and benefit. Braz J Otorhinolaryngol 2010;76(5):630-638 [FREE Full text] [Medline: 20963348]
23. Skarzynski H, Lorens A, Piotrowska A, Anderson I. Partial deafness cochlear implantation provides benefit to a new population of individuals with hearing loss. Acta Otolaryngol 2006 Sep;126(9):934-940. [doi: 10.1080/00016480600606632] [Medline: 16864490]

24. Penteado SP, Bento RF, Battistella LR, Silva SM, Sooful P. Use of the satisfaction with amplification in daily life questionnaire to assess patient satisfaction following remote hearing aid adjustments (telefitting). JMIR Med Inform 2014 Sep 02;2(2):e18 [FREE Full text] [doi: 10.2196/medinform.2769] [Medline: 25599909]
25. Cox RM, Alexander GC, Beyer CM. Norms for the international outcome inventory for hearing aids. J Am Acad Audiol 2003 Oct;14(8):403-413. [Medline: 14655953]
26. Reed NS, Betz J, Kendig N, Korczak M, Lin FR. Personal sound amplification products vs a conventional hearing aid for speech snderstanding in noise. JAMA 2017 Jul 04;318(1):89-90 [FREE Full text] [doi: 10.1001/jama.2017.6905] [Medline: 28672306]
27. Hohmann V, Paluch R, Krueger M, Meis M, Grimm G. The virtual reality lab: realization and application of virtual sound environments. Ear Hear 2020;41 Suppl 1:31S-38S [FREE Full text] [doi: 10.1097/AUD.0000000000000945] [Medline: 33105257]
28. Seol HY, Kang S, Lim J, Hong SH, Moon IJ. Feasibility of virtual reality audiological testing: prospective study. JMIR Serious Games 2021 Aug 31;9(3):e26976 [FREE Full text] [doi: 10.2196/26976] [Medline: 34463624]
29. Trecca EM, Gelardi M, Cassano M. COVID-19 and hearing difficulties. Am J Otolaryngol 2020;41(4):102496 [FREE Full text] [doi: 10.1016/j.amjoto.2020.102496] [Medline: 32327217]
30. Kates JM. A computer simulation of hearing aid response and the effects of ear canal size. J Acoust Soc Am 1988 May;83(5):1952-1963. [doi: 10.1121/1.396481] [Medline: 3403807]

Abbreviations

AV: aversiveness
BEN: benefit
BN: background noise
EC: ease of communication
HA: hearing aid
HL: hearing level
IO: impact on other people
K-APHAB: Korean version of the abbreviated profile of hearing aid benefit
K-HINT: Korean version of hearing in noise test
K-IOI-HA: Korean adaptation of the international outcome inventory for hearing aids
OTC: over the counter
Post AV: aversiveness subscale under the aided condition
Post BN: background noise subscale under the aided condition
Post EC: ease of communication subscale under aided condition
Post RV: reverberation subscale under the aided condition
Pre AV: aversiveness subscale under the unaided condition
Pre BN: background noise subscale under the unaided condition
Pre EC: ease of communication subscale under the unaided condition
Pre RV: reverberation subscale before using the wearable augmented reality device
PSAP: personal sound amplification product
QOL: quality of life
RAL: residual limitation in daily life activities
RPR: residual restrictions to participation
RV: reverberation
SAT: satisfaction
USE: duration of HA use
WARD: wearable augmented reality device
WRS: word recognition score

Edited by L Buis; submitted 08.09.21; peer-reviewed by J Jung, R Lundin; comments to author 15.11.21; revised version received 26.11.21; accepted 07.01.22; published 23.03.22.

Please cite as:
Han UG, Lee JY, Kim GY, Jo M, Lee J, Bang KH, Cho YS, Hong SH, Moon IJ
Real-World Effectiveness of Wearable Augmented Reality Device for Patients With Hearing Loss: Prospective Study
JMIR Mhealth Uhealth 2022;10(3):e33476
URL: https://mhealth.jmir.org/2022/3/e33476
doi:10.2196/33476
PMID:35320113

Original Paper

Impact of Masticatory Behaviors Measured With Wearable Device on Metabolic Syndrome: Cross-sectional Study

Fumiko Uehara[1], DDS, PhD; Kazuhiro Hori[1], DDS, PhD; Yoko Hasegawa[1], DDS, PhD; Shogo Yoshimura[1], DDS, PhD; Shoko Hori[1], DDS; Mari Kitamura[2], PhD; Kohei Akazawa[3], MD, PhD; Takahiro Ono[1], DDS, PhD

[1]Division of Comprehensive Prosthodontics, Faculty of Dentistry and Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan

[2]School of Food Sciences and Nutrition, Mukogawa Women's University, Nishinomiya, Japan

[3]Department of Medical Informatics, Niigata University Medical and Dental Hospital, Niigata, Japan

Corresponding Author:
Kazuhiro Hori, DDS, PhD
Division of Comprehensive Prosthodontics
Faculty of Dentistry and Graduate School of Medical and Dental Sciences
Niigata University
2-5274 Gakkocho-dori
Niigata, 9518514
Japan
Phone: 81 25 227 2891
Fax: 81 25 229 3454
Email: hori@dent.niigata-u.ac.jp

Abstract

Background: It has been widely recognized that mastication behaviors are related to the health of the whole body and to lifestyle-related diseases. However, many studies were based on subjective questionnaires or were limited to small-scale research in the laboratory due to the lack of a device for measuring mastication behaviors during the daily meal objectively. Recently, a small wearable masticatory counter device, called bitescan (Sharp Co), for measuring masticatory behavior was developed. This wearable device is designed to assess objective masticatory behavior by being worn on the ear in daily life.

Objective: This study aimed to investigate the relation between mastication behaviors in the laboratory and in daily meals and to clarify the difference in mastication behaviors between those with metabolic syndrome (MetS) and those without (non-MetS) measured using a wearable device.

Methods: A total of 99 healthy volunteers (50 men and 49 women, mean age 36.4 [SD 11.7] years) participated in this study. The mastication behaviors (ie, number of chews and bites, number of chews per bite, and chewing rate) were measured using a wearable ear-hung device. Mastication behaviors while eating a rice ball (100 g) in the laboratory and during usual meals for an entire day were monitored, and the daily energy intake was calculated. Participants' abdominal circumference, fasting glucose concentration, blood pressure, and serum lipids were also measured. Mastication behaviors in the laboratory and during meals for 1 entire day were compared. The participants were divided into 2 groups using the Japanese criteria for MetS (positive/negative for MetS or each MetS component), and mastication behaviors were compared.

Results: Mastication behaviors in the laboratory and during daily meals were significantly correlated (number of chews r=0.36; P<.001; number of bites r=0.49; P<.001; number of chews per bite r=0.33; P=.001; and chewing rate r=0.51; P<.001). Although a positive correlation was observed between the number of chews during the 1-day meals and energy intake (r=0.26, P=.009), the number of chews per calorie ingested was negatively correlated with energy intake (r=–0.32, P=.002). Of the 99 participants, 8 fit the criteria for MetS and 14 for pre-MetS. The number of chews and bites for a rice ball in the pre-MetS(+) group was significantly lower than the pre-MetS(–) group (P=.02 and P=.04, respectively). Additionally, scores for the positive abdominal circumference and hypertension subgroups were also less than the counterpart groups (P=.004 and P=.01 for chews, P=.006 and P=.02 for bites, respectively). The number of chews and bites for an entire day in the hypertension subgroup were significantly lower than in the other groups (P=.02 and P=.006). Furthermore, the positive abdominal circumference and hypertension subgroups showed lower numbers of chews per calorie ingested for 1-day meals (P=.03 and P=.02, respectively).

Conclusions: These results suggest a relationship between masticatory behaviors in the laboratory and those during daily meals and that masticatory behaviors are associated with MetS and MetS components.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry R000034453; https://tinyurl.com/mwzrhrua

(JMIR Mhealth Uhealth 2022;10(3):e30789) doi:10.2196/30789

KEYWORDS

metabolic syndrome; mastication behaviors; wearable device; daily meal; energy intake; chew; internet of things

Introduction

Recent studies in various fields have revealed that mastication affects various functions in the body, such as obesity prevention [1,2], immune cell differentiation [3,4], and dementia prevention [5]. Among these health effects of mastication, research attention has focused particularly on obesity [6] and diabetes [7,8]. Furthermore, many studies have reported a relationship between chewing frequency, eating speed, and metabolic syndrome (MetS) [9-13]. In these studies, masticatory behavior was also thought to be associated with energy intake. Indeed, Borvornparadorn et al [14] reported that increased number of chews per bite could reduce energy intake.

The relationship between eating behavior, especially fast eating, and MetS or obesity has long been a focus of attention. Some large-scale epidemiological studies also suggested a link between eating speed and obesity and MetS [1,2]. In particular, studies have shown that slower eating speed can reduce energy intake [10,14-18] and that prolonged chewing of food can reduce meal intake [19]. However, most studies were limited to self-administered questionnaires. It is difficult to exactly recognize our own eating behavior (ie, eating fast or slow and less or more chewing). Additionally, assessing mastication behaviors lacks objectivity because of the lack of devices for measuring mastication behaviors during daily meals.

In previous studies, chewing frequency was assessed using dedicated jaw movement measuring devices [20], muscular activity measuring devices (electromyography) [21], videorecordings [22-24], or wearable chewing counting devices [25-27]. However, most of these devices were not developed past the prototype stage, and participants might not eat naturally with these large devices since the requirement for special devices often hinders the evaluation of masticatory movement. Additionally, masticatory behaviors when eating meals while being videorecorded might be different from those during a usual meal.

A simple and accurate mastication measuring device called bitescan (Sharp Co) [28] was developed. The bitescan is a wearable ear hook–type device that monitors daily mastication behavior without disturbing the wearer. Furthermore, this revolutionary device can accurately monitor mastication behavior, namely the number of chews, number of bites, and chewing rate, making it possible to obtain these measurements in the laboratory as well as at home. We have previously confirmed the validity of this mastication measuring device [28].

The aim of this study was to investigate whether masticatory behavior in laboratory had a relationship to that in daily meals and to clarify the relationship between masticatory behavior and MetS by more objectively and accurately monitoring masticatory behavior using the bitescan. We hypothesized that masticatory behaviors in different environments had a relationship and that the presence or absence of MetS was related to masticatory behaviors.

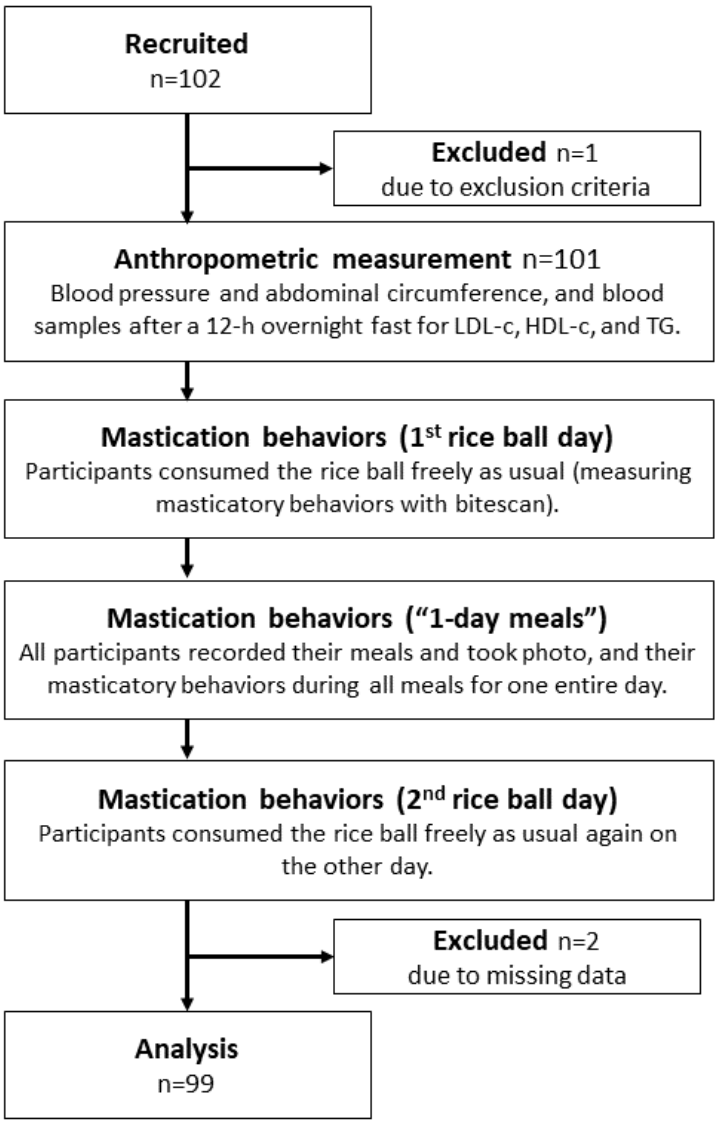
Methods

Participants

The participants were 50 healthy men and 49 healthy women, with a mean age of 36.4 (SD 11.7) years, who volunteered for the study. Participants were recruited for the study using advertisements, flyers, and personal communications. We excluded participants who had dysphagia, dental pain, periodontal problems, temporomandibular joint dysfunction syndrome, and those who were receiving dental treatment or medication for lifestyle-related diseases, such as diabetes, hypertension, and hyperlipidemia. These exclusion criteria were confirmed verbally. We explained the purpose of our study, and those who agreed to participate provided written informed consent, which was documented before they entered the study.

Sample size was estimated based on the number of chews while consuming a rice ball in MetS and non-MetS groups. Results from our preliminarily study (MetS group: mean 167, SD 67; non-MetS group: mean 222, SD 79) provided the effect size as 0.75. The prevalence of metabolic syndrome in Japanese was approximately 20% [29]. Therefore, 94 participants were required for 80% power with a 2-sided =.05 for Mann-Whitney *U* test (G*Power 3.1.9.7, Heinrich-Heine-Universität). A total of 102 people applied to this study, 1 was excluded using the exclusion criteria, and 2 people with missing data were excluded; 99 participants were included in the final analysis (Figure 1).

Figure 1. Flow diagram of participant assessment through the trial.



Ethics Approval

The study conformed to the standards of the Declaration of Helsinki and was approved by the institutional review board of Niigata University (approval number: 2017-0230). The study was registered at the University Hospital Medical Information Network Clinical Trials Registry [R000034453].

Bitescan Device to Measure Masticatory Behaviors

The number of chews, number of chews per bite, number of bites, and chewing rate were measured with the bitescan device [28] (Figure 2). This wearable device has an infrared distance sensor and accelerometer and scans the morphological change in the skin surface at 20 Hz on the posterior side of the pinna during mastication. This device is designed to be worn on the right side and has an adjustable ear hook. Three different ear hook sizes (small, medium, and large) were prepared; therefore, we could adjust the device and use the ear hook best suited to each participant's pinna (Figure 1). Before the measurement, we fit the participants with the bitescan to ensure that the sensor was correctly located on the back of the pinna. The bitescan was then connected to a smartphone (SHM05, Sharp Co) via Bluetooth, and the data were collected with a dedicated smartphone app (Figure 3).

Figure 2. The bitescan: (A) main body and (B) in position.

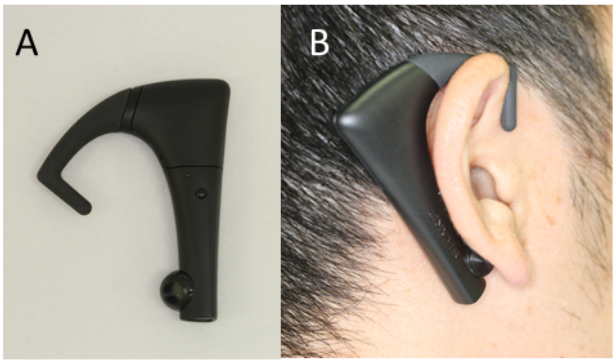


Figure 3. Screenshot of bitescan app: (A) during measurement and (B) showing results of mastication behavior.

Data Collection

Mastication Behaviors

A rice ball (100 g, Maho Cold Foods Co Ltd) was prepared as the prescribed test food. Each participant was instructed to sit in a chair and relax while the properly sized bitescan was adjusted and we confirmed that it worked properly. Participants were then instructed to consume the rice ball freely, as usual. There were no limitations and no special instructions for participants when consuming the test food regarding the number of chews, chewing rate, or the timing of swallowing; they were asked only to eat the test food as they normally would.

The measurements of the participant chewing the rice ball were taken twice. All participants were also asked to record their masticatory behaviors during all meals for 1 entire day. The 2 rice ball measurements and the 1-day measurement were performed on different days.

Energy Intake

All participants completed dietary records accompanied by photographs during the 1-day measurement. Using the photographs, a registered dietitian identified foods and estimated their portion sizes. Energy intake was calculated in accordance with the Standard Tables of Food Composition in Japan 2015 (Seventh Revised Edition) [30].

Anthropometric Measurement

The anthropometric measurements were blood pressure (BP) and abdominal circumference (AC), which were measured by the medical staff on the first visit day before measuring masticatory behavior. Blood pressure was measured using a sphygmomanometer (HBP-9020, Omron Corp) with the right arm with participant in the supine position. Abdominal circumference was measured at the midpoint between the iliac crest and rib cage on the midaxillary line using a tape measure. All measurements were performed with participants dressed in light clothing and barefoot.

We also collected blood samples after a 12-hour overnight fast. Fasting glucose concentration (GLU) and serum lipids (SL; low-density lipoprotein cholesterol, high-density lipoprotein cholesterol [HDL-c], and triglycerides) were measured using an automated analyzer.

Diagnosis of MetS

We used the Japanese diagnosis of MetS, which is made according to the following criteria [31]:

1. AC: men: ≥0.85 m; women: ≥0.90 m
2. SL: hypertriglyceridemia: ≥1.69 mmol/L or low HDL-c <1.04 mmol/L
3. BP: systolic pressure ≥130 mm Hg or diastolic pressure ≥85 mm Hg
4. GLU: ≥6.1 mmol/L

We divided the participants into the following groups:

- MetS(+) versus MetS(–) groups: criterion 1 and more than two of criteria 2, 3, and 4
- pre-MetS(+) versus pre-MetS(–) groups: criterion 1 and more than one of criteria 2, 3, and 4
- possible MetS(+) versus possible MetS(–) groups: more than one of criteria 1, 2, 3, and 4

Furthermore, in accordance with the above diagnostic criteria for MetS, we also divided participants into the following groups:

- AC(+) versus AC(–) groups according to criterion 1
- SL(+) versus SL(–) groups according to criterion 2
- BP(+) versus BP(–) groups according to criterion 3
- GLU(+) versus GLU(–) groups according to criterion 4

Statistical Analysis

We analyzed the number of chews, number of bites, number of chews per bite, and chewing rate while participants consumed a rice ball and during the 1-day meals. When recording the data during ingestion of the rice balls, we used the average of the 2 measurements as the representative value. We then calculated the number of chews per calorie ingested for the 1-day measurements.

To compare masticatory behaviors in different environments, the relationship of masticatory behaviors between consuming a rice ball and ingesting all meals for the 1-day measurement was calculated using a Spearman correlation. The masticatory behaviors between the 2 groups were compared using the Mann-Whitney *U* test. Statistical analyses were conducted using SPSS (version 23.0 for Windows, IBM Corp), and the level of significance was set at *P*=.05.

Results

Masticatory Behaviors Under Different Conditions

The Spearman correlation coefficients between consuming a rice ball and ingesting the 1-day meals were 0.36 (*P*<.001), 0.49 (*P*<.001), 0.33 (*P*=.001), and 0.51 (*P*<.001) for the number of chews, number of bites, number of chews per bite, and chewing rate, respectively, and significant correlations were observed (Table 1). A positive correlation was observed between the number of chews during the 1-day meals and energy intake (*r*=0.26, *P*=.009). However, the number of chews per calorie ingested was negatively correlated with energy intake for the 1-day meals (*r*=–0.32, *P*=.002). Furthermore, the number of chews per calorie in the 1-day meals showed significant positive correlations with the number of chews, number of chews per bite, number of bites, and the chewing rate while eating a rice ball (*r*=0.48, *P*<.001; *r*=0.20; *P*=.04; *r*=0.32; *P*=.001; and *r*=0.24; *P*=.02, respectively).

The mean number of chews for a rice ball and the 1-day meals were 215 (SD 85) and 2306 (SD 1123), respectively (Table 2 and Multimedia Appendix 1). The number of bites for a rice ball and the 1-day meals were 19.5 (SD 8.0) and 210 (SD 135), respectively (Table 2 and Multimedia Appendix 1). The number of chews per bite (11.7, SD 4.3) and the chewing rate (70.8, SD 7.1) when eating a rice ball were similar to the number of chews per bite (12.4, SD 5.7) and the chewing rate (71.4, SD 7.6) during the 1-day meals.

Female participants had higher numbers of chews (*P*=.009), numbers of bites (*P*<.001) for a rice ball, and numbers of chews per calorie (*P*=.045), a smaller number of chews per bite (*P*=.01 for rice ball, *P*=.02 for 1-day meals), and a slower chewing rate (*P*=.008 for rice ball, *P*=.02 for 1-day meals).

Table 1. Relationship between masticatory behaviors in the laboratory and during meals ingested for 1 entire day (n=99).

	Eating a rice ball in the laboratory				Caloric intake for 1 entire day, *r* (*P* value)
	Number of chews, *r* (*P* value)	Number of chews per bite, *r* (*P* value)	Number of bites, *r* (*P* value)	Chewing rate, *r* (*P* value)	
Number of chews	.360 (<.001)	.282 (.005)	.126 (.21)	.328 (.001)	.262 (.009)
Number of chews per bite	.122 (.23)	.493 (<.001)	–.281 (.005)	.448 (<.001)	.062 (.54)
Number of bites	.205 (.04)	–.104 (.30)	.334 (.001)	–.038 (.71)	.151 (.14)
Chewing rate	.185 (.07)	.519 (<.001)	–.222 (.03)	.512 (<.001)	.047 (.65)
Caloric intake for 1 entire day	–.172 (.09)	.164 (.11)	–.357 (<.001)	.179 (.08)	—[a]
Number of chews per calorie ingested	.484 (<.001)	.203 (.04)	.324 (.001)	.235 (.02)	–.315 (.002)

[a]Not applicable.

Masticatory Behaviors and MetS

Of the participants, 8% (8/99) fulfilled the diagnostic criteria for MetS. There was no significant difference in mastication behaviors between the MetS(+) group and MetS(–) groups.

A total of 14% (14/99) of participants fulfilled the diagnostic criteria for pre-MetS. The numbers of chews (*P*=.02) and bites (*P*=.04) of a rice ball for the pre-MetS(+) group were significantly lower than for those in the pre-MetS(–) group. However, the 1-day mastication behaviors were not significantly different between the pre-MetS(+) and pre-MetS(–) groups.

A total of 53% (52/99) met one or more criteria for a MetS diagnosis. The possible MetS(+) group showed significantly lower numbers of chews and bites for both the rice ball (*P*<.001, *P*=.001) and 1-day meals (*P*=.007, *P*=.008) and a lower number of chews per calorie ingested (*P*=.005) than the possible MetS(–) group.

Table 2. Masticatory behavior while eating a rice ball in the laboratory.

	n	Number of chews		Number of chews per bite		Number of bites		Chewing rate (/min)	
		median (IQR)	*P* value	median (IQR)	*P* value	median (IQR)	*P* value	median (IQR)	*P* value
All	99	192 (106)	—[a]	10.7 (5.1)	—	18.0 (7.5)	—	70.7 (8.6)	—
Sex	—	—	.009	—	.01	—	<.001	—	.008
Men	50	172 (106)	—	11.7 (7.6)	—	15.0 (7.5)	—	72.3 (10.2)	—
Women	49	208 (106)	—	10.1 (3.6)	—	21.0 (9.0)	—	70.1 (9.3)	—
MetS[b]	—	—	.24	—	.83	—	.39	—	.25
Yes	8	171 (67)	—	11.6 (7.7)	—	16.5 (5.8)	—	72.9 (12.9)	—
No	91	199 (110)	—	10.7 (5.1)	—	18.5 (8.0)	—	70.4 (9.2)	—
pre-MetS	—	—	.02	—	.92	—	.04	—	.39
Yes	14	162 (56)	—	11.6 (7.2)	—	15.0 (8.5)	—	71.7 (10.7)	—
No	85	202 (114)	—	10.7 (5.0)	—	19.0 (9.0)	—	70.4 (8.8)	—
possible MetS	—	—	<.001	—	.74	—	.001	—	.92
Yes	52	174 (84)	—	10.6 (5.8)	—	16.3 (7.3)	—	71.4 (9.9)	—
No	47	214 (114)	—	10.8 (4.6)	—	20.5 (9.5)	—	70.3 (8.6)	—
Abdominal circumference (m)	—	—	.004	—	.90	—	.006	—	.86
AC[c](+)	24	166 (80)	—	10.9 (7.1)	—	15.0 (7.5)	—	71.2 (10.5)	—
AC(–)	75	206 (118)	—	10.7 (4.8)	—	20.0 (10.5)	—	70.4 (8.5)	—
Serum lipid (mmol/L)	—	—	.04	—	.87	—	.15	—	.73
SL[d](+)	13	161 (48)	—	11.7 (7.9)	—	17.0 (9.5)	—	71.3 (14.1)	—
SL(–)	86	203 (113)	—	10.7 (4.9)	—	18.5 (8.6)	—	70.5 (8.4)	—
Blood pressure (mm Hg)	—	—	.01	—	.91	—	.02	—	.91
BP[e](+)	36	168 (84)	—	10.9 (5.6)	—	16.3 (6.4)	—	71.4 (9.9)	—
BP(–)	63	206 (113)	—	10.7 (4.9)	—	20.0 (10.0)	—	70.4 (8.5)	—
Fasting glucose (mmol/L)	—	—	.77	—	.76	—	.65	—	.40
GLU[f](+)	7	208 (90)	—	11.5 (5.2)	—	17.0 (11.0)	—	73.4 (6.5)	—
GLU(–)	92	191 (110)	—	10.7 (5.2)	—	18.3 (7.9)	—	70.4 (9.4)	—

[a]Not applicable.

[b]MetS: metabolic syndrome.

[c]AC: abdominal circumference (men ≥0.85 m, women ≥0.90 m).

[d]SL: serum lipids (triglyceride ≥1.69 mmol/L or HDL cholesterol <1.04 mmol/L).

[e]BP: blood pressure (systolic pressure ≥130 mm Hg or diastolic pressure ≥85 mm Hg).

[f]GLU: fasting glucose concentration (≥6.1 mmol/L).

For abdominal circumference, 24% (24/99) of participants exceeded the criteria. The AC(+) group showed significantly lower numbers of chews (P=.004) and bites (P=.006) when ingesting a rice ball and a lower number of chews per calorie (P=.03) ingested than the AC(–) group.

For serum lipids, 13% (13/99) of participants exceeded the criteria. The number of chews of a rice ball in the SL(+) group was significantly lower than in the SL(–) group (P=.04).

For blood pressure, 36% (36/99) of participants exceeded the criteria. The BP(+) group showed significantly lower numbers of chews (P=.01, P=.02) and bites (P=.02, P=.006) for both a rice ball and 1-day meals and a significantly lower number of chews per calorie (P=.02) ingested than the BP(–) group.

For blood glucose concentration, 7% (7/99) of participants exceeded the criteria. There was no significant difference in the masticatory behaviors between the GLU(+) and GLU(–) groups except for the number of bites for the 1-day meals.

Discussion

Principal Findings

In this study using a wearable mastication monitoring device, we identified baseline mastication behaviors, which have not been determined previously. To our knowledge, there are no studies investigating the relationship between mastication behaviors eating usual meals and eating in the laboratory. This was also the first time that this many participants were objectively evaluated regarding their eating behavior, including the numbers of chews and bites, and the first time a relationship between masticatory behaviors and MetS has been examined. Therefore, we believe that this study monitoring daily dietary mastication behaviors using wearing device can be considered innovative research.

Mastication Behaviors and Environment

In most studies evaluating the number of chews, participants consumed only a prescribed meal in the laboratory, which was considered different from usual meals or eating behaviors. Additionally, masticatory behaviors when eating meals while being videorecorded might be different from those during a usual meal.

Although Zhang et al [32] measured mastication behavior in the lab and home environment for assessment of detection of mastication, they did not investigate the relationship. Petty et al [33] found that self-reported eating rate aligned with the eating rate measured in laboratory but not with free-living meals. For these reasons, we compared the masticatory behavior in the laboratory environment with that in the normal environment. We found the number of chews per bite (rice ball: 11.7, SD 4.3; 1-day meals: 12.4, SD 5.7) and chewing rate (rice ball: 70.8, SD 7.1; 1-day meals: 71.4, SD 7.6) had significant correlation and were almost the same. In our study, the contents of daily meals were not regulated, and the participants ingested as usual. The amount and physical characteristics of food affect mastication behavior, and mastication behavior varies greatly among individuals. However, even when comparing the results for each individual, the correlation between the number of chews per bite and the chewing rate between the laboratory environment and the usual daily environment was approximately 0.50, a moderate-intensity correlation.

Accordingly, the masticatory behavior in the laboratory had significant correlation with the participants' usual masticatory behaviors, suggesting that the chewing behavior in daily life can be inferred from the chewing behavior of rice balls in the laboratory.

We confirmed that our ear hook–type device hangs only on the pinna and has little effect on eating behavior; however, it is difficult to make comparisons of eating behavior with no device. Additionally, participants' awareness of masticatory behavior monitoring might affect outcomes. To avoid this, long-term measurements should be performed to allow participants to become accustomed to the device. However, similarly, the effects of attaching electrodes to the masseter and temporalis muscles and eating while being videorecorded have not been investigated.

In our study, the mean number of chews per day was 2306 (SD 1123). To our knowledge, few studies have measured and reported the number of chews during usual meals or within 1 entire day. This result is an indicator of the masticatory behaviors of Japanese people eating their daily diet.

The correlation between the number of chews while eating a rice ball and the 1-day meals was 0.36, and the correlation coefficient of the number of bites was 0.33, which indicated a weak correlation. In other words, generally, the number of chews and the number of bites when eating usual meals did not change even when eating a rice ball in the laboratory. However, we should also consider the amount of dietary intake.

As a result of investigating nutritional intake, we found those who chewed more tended to have a large energy intake. In other words, we expected that people who ate a large amount of food in a day would chew a lot. Borvornparadorn et al [14] measured energy intake using regulated food in a laboratory and investigated the relationship between the amount of mastication and calorie intake. However, as far as we know, no studies have measured both the energy intake of daily meals and mastication behaviors. The amount of food was considered to have a great influence on masticatory behavior. However, it was difficult to measure the volume and weight of meals in this study. Petty et al [33] reported eating rate calculated in calories per minute but not volume or weight. Due to differences in the amount of food people eat, we decided to calculate the number of chews per calorie ingested. As a result, a moderate correlation coefficient of 0.48 was obtained for the number of chews of a rice ball. Furthermore, the number of chews per calorie ingested showed a significant correlation with the other items measured when participants ingested rice balls. Therefore, the investigation of the number of chews per calorie intake appears to be meaningful.

In this research, we chose rice balls as the test food after consideration that preference and tableware [34] might affect masticatory behaviors. Rice balls have long been popular as a convenient food for Japanese people, although they may not be a familiar food internationally. We also chose rice balls because these were easily accepted by the Japanese and could be used

as a prescribed amount of test food. Asian people generally use chopsticks for eating, but rice balls are usually ingested using the hands.

Relationship Between MetS and Masticatory Behavior

A significant relationship between MetS and eating speed [1,2,35,36] has been reported previously. However, most of these studies were limited to participant self-reported assessment of eating behavior. Regarding self-reported eating behavior, Woodward et al [37] reported a discrepancy between self-reporting and objective observation on eating rate. Furthermore, masticatory performance was reported to be associated with the prevalence of MetS [38,39]. These previous reports suggested that mastication was associated with MetS through nutrition and feeling full. On the other hand, our results showed no significant difference in the number of chews between the MetS(+) and MetS(–) groups. One of the reasons is that the percentage of participants in our study with MetS (8%) was much lower than the Japanese prevalence of MetS [29]. However, we found a difference in the numbers of chews and bites of a rice ball between the pre-MetS(+) group and pre-MetS(–) group. Furthermore, the possible MetS(+) group showed significantly lower numbers of chews and bites for both the rice ball and 1-day meals and a lower number of chews per calorie ingested than the possible MetS(–) group. MetS arises from a combination of factors that result in obesity, hypertension, serological abnormalities, and abdominal obesity. It is impossible to explain such complicated pathological conditions with one component (chewing), but we suspect that masticatory behavior may have an effect on lifestyle-related diseases.

Abdominal circumference is an indicator associated with obesity and is the most important factor in MetS. One study reported that it was difficult to obtain a feeling of fullness during meals if the number of chews was low or when participants ate faster [14]. As a result, daily energy intake increased significantly, which was thought to cause obesity [40]. Some studies reported that the amount of food and snack intake decreased significantly with an increase in the number of chews and the chewing time [19,41]. In our study, the number of chews while consuming a rice ball was significantly lower in the AC(+) group. In addition, although no significant difference was observed, the number of chews per day was less in the AC(+) group and the daily energy intake was higher, which may support the above consideration. Differences in the number of chews of rice balls and the number of chews per calorie ingested in the 1-day meals suggested that the number of chews per unit amount might be associated with obesity.

In contrast, no significant difference in the number of chews per bite was observed in this study. In Japan, enlightenment activities with an emphasis on the number of chews per bite such as “chew 30 times per bite” have been recommended. However, considering the results of our study, the number of chews per bite may not be directly related to obesity. In addition, from our results, the number of bites in the AC(+) group was significantly lower. This suggested that the AC(+) group ingested large bites (ie, we considered that the number of chews decreased as a result of taking a large amount into the oral cavity and eating it with a modest number of chews). From these findings, the instruction not only to increase the number of chews but also to reduce the amount taken in one bite might be effective. According to a report by Fukuda et al [40], the number of chews per bite did not change even if the amount per bite increased; therefore, if the amount per bite was large, the total mealtime duration would be shorter and the total number of chews would be lower when eating the same amount of food. In addition, some studies reported that a reduction in the size of the bite helped prevent obesity [42,43].

In our study, there was no difference between the GLU(+) group and the GLU(–) group except for the number of bites in the 1-day meals. In our study, the GLU(+) group constituted only 7 participants (7%). However, some reports suggested a relationship between diabetes and mastication. Masticatory movement promotes glucagon-like peptide-1 secretion, which leads to rapid insulin secretion and may also affect dietary sugar absorption [44,45]. In addition, Read et al [46] suggested that the easiest way to avoid raising the blood sugar level even after eating was to swallow without chewing. Although it was impossible to swallow without chewing for all of the usual meals, in the study by Ranawana et al [47], which compared 15 and 30 chews during a rice meal, fewer chews resulted in a lower postprandial total blood and glycemic index. In contrast, Sato et al [48] reported that frequent chews suppressed the rise in postprandial blood glucose concentration and promoted insulin secretion. As described above, there are many conflicting reports regarding mastication and blood glucose concentration. Furthermore, obesity, diabetes, and postprandial blood glucose responses vary greatly depending on race and sex [49]. Because many complex factors might be involved, further research is required.

In our study, blood pressure was the second factor most associated with mastication after obesity. Studies report that slower and more thorough chewing and eating could increase eating-induced heat production, which could increase systemic metabolism [50]. In our study, the BP(+) group showed significantly lower numbers of chews and bites when ingesting rice balls, lower numbers of chews and bites per day, and a lower number of chews per calorie ingested.

A lower number of chews could affect taste perception because the food is not sufficiently crushed and exposure time of the food in the oral cavity is short. Bolhuis et al [51] reported a relationship between saltiness and appetite using two types of soup with different salinity and density. According to the study, longer exposure times to the oral sensation led to lower soup intake. The authors also reported that the exposure time to the oral sensation had a greater effect than salinity. Increasing the exposure to taste buds per food unit might be effective in reducing food intake. Therefore, it is possible that a participant who chewed less and had a short exposure time in the oral cavity regarding the taste stimulus of the food might have ingested more food. One of the major reasons for increased blood pressure is excessive salt intake; therefore, low masticatory behavior might be related to blood pressure.

Diabetes, hyperlipidemia, and hypertension may have multiple forms related to obesity and arteriosclerosis. Furthermore, MetS

is related to lifestyle issues, such as diet, exercise, drinking, smoking, and stress. It is known that negative emotions such as anger, fear, and sadness have been associated with increased impulsive eating and consumption of unhealthy foods [52]. In this study, we investigated only masticatory behaviors and dietary energy intake, and multivariate analysis including these factors is needed. Nevertheless, mastication behavior is a lifestyle issue, and it should receive attention not only regarding the amount and kind of food eaten, but also regarding the way of eating. Our results suggested that masticatory behavior when eating a test meal in the laboratory as well as in the participants' usual daily diet was associated with MetS. This was a cross-sectional study, and a longitudinal study is necessary to examine whether people with poor masticatory behaviors are more likely to develop MetS or whether better masticatory behaviors are useful for preventing MetS. Furthermore, it is necessary to consider whether masticatory behavior can be transformed by certain interventions.

Although this was a cross-sectional study and other MetS-related factors should be considered, our results suggested that mastication behavior might be related to MetS and MetS components. These results help clarify the relationship between masticatory behaviors, metabolic syndrome, and energy intake. We will perform multivariate analysis for each item related to MetS in the future. The wearable chewing counter that we developed was useful for monitoring masticatory behavior in daily meals. In addition, the smartphone app connected to this device can be equipped with a masticatory behavior change algorithm. Masticatory behavior targets selected for each participant and wearable devices that are easy to wear should lower the hurdles for intervention studies and could contribute to the evidence of a relationship between masticatory behavior and health and the effects of masticatory behavior change. Furthermore, the mastication data measured by this wearable device can send the user's information to medical personnel by using the internet environment. Medical personnel could provide support for improving daily life, such as sending advice based on this information. The usefulness of mastication could be examined using the big data collected in this way. We believe that this device and app could be used in the future as an approach to monitor and change daily masticatory habits.

Conclusions

In this study, we investigated mastication behaviors (ie, number of chews and bites, number of chews per bite, and chewing rate) measured using a wearable ear-hung device. We found a significant correlation between mastication behaviors in the laboratory and in daily meals, which are different environments. Furthermore, a significant correlation was observed between the number of chews during the 1-day meals and energy intake and between the number of chews per calorie ingested and energy intake. Neither the pre-MetS obesity nor the hypertension group had a lower number of chews, bites, and chews per calorie. These results suggest that masticatory behaviors are related to MetS and MetS components.

Acknowledgments

We deeply thank and send our best regards to Ms Yuna Kinoshita for supporting the data collection. We thank Jane Charbonneau, DVM, from Edanz for editing a draft of this manuscript.

Authors' Contributions

This study was designed by KH. The data were collected by FU, YY, YH, SY, and SH. Nutritional analysis was performed by MK. Statistical analysis was performed by KA. The manuscript was drafted by FU and KH and edited by TO.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Masticatory behavior during meals ingested for one entire day.
[DOCX File , 25 KB - mhealth_v10i3e30789_app1.docx]

References

1. Otsuka R, Tamakoshi K, Yatsuya H, Murata C, Sekiya A, Wada K, et al. Eating fast leads to obesity: findings based on self-administered questionnaires among middle-aged Japanese men and women. J Epidemiol 2006 May;16(3):117-124 [FREE Full text] [doi: 10.2188/jea.16.117] [Medline: 16710080]
2. Maruyama K, Sato S, Ohira T, Maeda K, Noda H, Kubota Y, et al. The joint impact on being overweight of self reported behaviours of eating quickly and eating until full: cross sectional survey. BMJ 2008 Oct 21;337:a2002 [FREE Full text] [doi: 10.1136/bmj.a2002] [Medline: 18940848]
3. Dutzan N, Abusleme L, Bridgeman H, Greenwell-Wild T, Zangerle-Murray T, Fife ME, et al. On-going mechanical damage from mastication drives homeostatic Th17 cell responses at the oral barrier. Immunity 2017 Jan 17;46(1):133-147 [FREE Full text] [doi: 10.1016/j.immuni.2016.12.010] [Medline: 28087239]
4. Proctor GB, Carpenter GH. Chewing stimulates secretion of human salivary secretory immunoglobulin A. J Dent Res 2001 Mar;80(3):909-913. [doi: 10.1177/00220345010800031201] [Medline: 11379894]

5. Kim E, Lee SK, Choi Y, Tanaka M, Hirotsu K, Kim HC, et al. Relationship between chewing ability and cognitive impairment in the rural elderly. Arch Gerontol Geriatr 2017;70:209-213. [doi: 10.1016/j.archger.2017.02.006] [Medline: 28214402]
6. Tanihara S, Imatoh T, Miyazaki M, Babazono A, Momose Y, Baba M, et al. Retrospective longitudinal study on the relationship between 8-year weight change and current eating speed. Appetite 2011 Aug;57(1):179-183. [doi: 10.1016/j.appet.2011.04.017] [Medline: 21565235]
7. Sakurai M, Nakamura K, Miura K, Takamura T, Yoshita K, Nagasawa S, et al. Self-reported speed of eating and 7-year risk of type 2 diabetes mellitus in middle-aged Japanese men. Metabolism 2012 Nov;61(11):1566-1571. [doi: 10.1016/j.metabol.2012.04.005] [Medline: 22560127]
8. Yamazaki T, Yamori M, Asai K, Nakano-Araki I, Yamaguchi A, Takahashi K, Nagahama Study Collaboration Group. Mastication and risk for diabetes in a Japanese population: a cross-sectional study. PLoS One 2013;8(6):e64113 [FREE Full text] [doi: 10.1371/journal.pone.0064113] [Medline: 23755114]
9. Hattori T, Konno S, Munakata M. Gender differences in lifestyle factors associated with metabolic syndrome and preliminary metabolic syndrome in the general population: the Watari Study. Intern Med 2017 Sep 01;56(17):2253-2259 [FREE Full text] [doi: 10.2169/internalmedicine.8578-16] [Medline: 28794374]
10. Zhu Y, Hollis JH. Increasing the number of chews before swallowing reduces meal size in normal-weight, overweight, and obese adults. J Acad Nutr Diet 2014 Jun;114(6):926-931. [doi: 10.1016/j.jand.2013.08.020] [Medline: 24215801]
11. Lee M, Huang Y, Wahlqvist ML. Chewing ability in conjunction with food intake and energy status in later life affects survival in Taiwanese with the metabolic syndrome. J Am Geriatr Soc 2010 Jun;58(6):1072-1080. [doi: 10.1111/j.1532-5415.2010.02870.x] [Medline: 20487084]
12. Fukui N, Shimazaki Y, Shinagawa T, Yamashita Y. Periodontal status and metabolic syndrome in middle-aged Japanese. J Periodontol 2012 Nov;83(11):1363-1371. [doi: 10.1902/jop.2012.110605] [Medline: 22248226]
13. Kral JG, Buckley MC, Kissileff HR, Schaffner F. Metabolic correlates of eating behavior in severe obesity. Int J Obes Relat Metab Disord 2001 Feb;25(2):258-264. [doi: 10.1038/sj.ijo.0801469] [Medline: 11410829]
14. Borvornparadorn M, Sapampai V, Champakerdsap C, Kurupakorn W, Sapwarobol S. Increased chewing reduces energy intake, but not postprandial glucose and insulin, in healthy weight and overweight young adults. Nutr Diet 2019 Feb;76(1):89-94. [doi: 10.1111/1747-0080.12433] [Medline: 29767425]
15. Li J, Zhang N, Hu L, Li Z, Li R, Li C, et al. Improvement in chewing activity reduces energy intake in one meal and modulates plasma gut hormone concentrations in obese and lean young Chinese men. Am J Clin Nutr 2011 Sep;94(3):709-716. [doi: 10.3945/ajcn.111.015164] [Medline: 21775556]
16. Scisco JL, Muth ER, Dong Y, Hoover AW. Slowing bite-rate reduces energy intake: an application of the bite counter device. J Am Diet Assoc 2011 Aug;111(8):1231-1235. [doi: 10.1016/j.jada.2011.05.005] [Medline: 21802572]
17. Shah M, Copeland J, Dart L, Adams-Huet B, James A, Rhea D. Slower eating speed lowers energy intake in normal-weight but not overweight/obese subjects. J Acad Nutr Diet 2014 Mar;114(3):393-402. [doi: 10.1016/j.jand.2013.11.002] [Medline: 24388483]
18. Teo PS, van Dam RM, Whitton C, Tan LWL, Forde CG. Association between self-reported eating rate, energy intake, and cardiovascular risk factors in a multi-ethnic Asian population. Nutrients 2020 Apr 13;12(4):1080 [FREE Full text] [doi: 10.3390/nu12041080] [Medline: 32295057]
19. Higgs S, Jones A. Prolonged chewing at lunch decreases later snack intake. Appetite 2013 Mar;62:91-95. [doi: 10.1016/j.appet.2012.11.019] [Medline: 23207188]
20. Johansson AS, Westberg K, Edin BB. Task-dependent control of the jaw during food splitting in humans. J Neurophysiol 2014 Jun 15;111(12):2614-2623 [FREE Full text] [doi: 10.1152/jn.00797.2013] [Medline: 24671539]
21. Park S, Shin W. Differences in eating behaviors and masticatory performances by gender and obesity status. Physiol Behav 2015 Jan;138:69-74. [doi: 10.1016/j.physbeh.2014.10.001] [Medline: 25447481]
22. Bellisle F, Guy-Grand B, Le Magnen J. Chewing and swallowing as indices of the stimulation to eat during meals in humans: effects revealed by the edogram method and video recordings. Neurosci Biobehav Rev 2000 Mar;24(2):223-228. [doi: 10.1016/s0149-7634(99)00075-5] [Medline: 10714385]
23. Ioakimidis I, Zandian M, Eriksson-Marklund L, Bergh C, Grigoriadis A, Södersten P. Description of chewing and food intake over the course of a meal. Physiol Behav 2011 Oct 24;104(5):761-769. [doi: 10.1016/j.physbeh.2011.07.021] [Medline: 21807012]
24. Farooq M, Doulah A, Parton J, McCrory MA, Higgins JA, Sazonov E. Validation of sensor-based food intake detection by multicamera video observation in an unconstrained environment. Nutrients 2019 Mar 13;11(3):609 [FREE Full text] [doi: 10.3390/nu11030609] [Medline: 30871173]
25. Taniguchi K, Kondo H, Tanaka T, Nishikawa A. Earable RCC: development of an earphone-type reliable chewing-count measurement device. J Healthc Eng 2018;2018:6161525 [FREE Full text] [doi: 10.1155/2018/6161525] [Medline: 29796232]
26. Yamaguchi T, Mikami S, Saito M, Okada K, Gotouda A. A newly developed ultraminiature wearable electromyogram system useful for analyses of masseteric activity during the whole day. J Prosthodont Res 2018 Jan;62(1):110-115. [doi: 10.1016/j.jpor.2017.04.001] [Medline: 28566138]

27. Sazonov E, Schuckers S, Lopez-Meyer P, Makeyev O, Sazonova N, Melanson EL, et al. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. Physiol Meas 2008 May;29(5):525-541 [FREE Full text] [doi: 10.1088/0967-3334/29/5/001] [Medline: 18427161]
28. Hori K, Uehara F, Yamaga Y, Yoshimura S, Okawa J, Tanimura M, et al. Reliability of a novel wearable device to measure chewing frequency. J Prosthodont Res 2021 Aug 21;65(3):340-345 [FREE Full text] [doi: 10.2186/jpr.JPR_D_20_00032] [Medline: 33441503]
29. The National Health and Nutrition Survey in Japan, 2018. Ministry of Health, Labour and Welfare, Japan. URL: https://www.e-stat.go.jp/dbview?sid=0003223783 [accessed 2022-03-07]
30. Standard tables of food composition in Japan, seventh revised version. Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. 2015. URL: https://www.mext.go.jp/en/policy/science_technology/policy/title01/detail01/1374030.htm [accessed 2015-03-07]
31. Takahara M, Katakami N, Kaneto H, Noguchi M, Shimomura I. Prediction of the presence of insulin resistance using general health checkup data in Japanese employees with metabolic risk factors. J Atheroscler Thromb 2014;21(1):38-48 [FREE Full text] [doi: 10.5551/jat.18622] [Medline: 24025703]
32. Zhang R, Amft O. Monitoring chewing and eating in free-living using smart eyeglasses. IEEE J Biomed Health Inform 2018 Jan;22(1):23-32. [doi: 10.1109/JBHI.2017.2698523] [Medline: 28463209]
33. Petty AJ, Melanson KJ, Greene GW. Self-reported eating rate aligns with laboratory measured eating rate but not with free-living meals. Appetite 2013 Apr;63:36-41. [doi: 10.1016/j.appet.2012.12.014] [Medline: 23266516]
34. Sun L, Ranawana DV, Tan WJK, Quek YCR, Henry CJ. The impact of eating methods on eating rate and glycemic response in healthy adults. Physiol Behav 2015 Feb;139:505-510. [doi: 10.1016/j.physbeh.2014.12.014] [Medline: 25484351]
35. Zhu B, Haruyama Y, Muto T, Yamazaki T. Association between eating speed and metabolic syndrome in a three-year population-based cohort study. J Epidemiol 2015;25(4):332-336. [doi: 10.2188/jea.je20140131]
36. Nanri A, Miyaji N, Kochi T, Eguchi M, Kabe I, Mizoue T. Eating speed and risk of metabolic syndrome among Japanese workers: the Furukawa Nutrition and Health Study. Nutrition 2020 Oct;78:110962. [doi: 10.1016/j.nut.2020.110962] [Medline: 32866899]
37. Woodward E, Haszard J, Worsfold A, Venn B. Comparison of self-reported speed of eating with an objective measure of eating rate. Nutrients 2020 Feb 26;12(3):599 [FREE Full text] [doi: 10.3390/nu12030599] [Medline: 32110855]
38. Kikui M, Ono T, Kokubo Y, Kida M, Kosaka T, Yamamoto M, et al. Relationship between metabolic syndrome and objective masticatory performance in a Japanese general population: the Suita study. J Dent 2017 Jan;56:53-57. [doi: 10.1016/j.jdent.2016.10.014] [Medline: 27793706]
39. Pérez-Sayáns M, González-Juanatey JR, Lorenzo-Pouso AI, Chamorro-Petronacci CM, Blanco-Carrión A, Marichalar-Mendía X, et al. Metabolic syndrome and masticatory hypofunction: a cross-sectional study. Odontology 2021 Jul;109(3):574-584. [doi: 10.1007/s10266-020-00577-x] [Medline: 33428015]
40. Fukuda H, Saito T, Mizuta M, Moromugi S, Ishimatsu T, Nishikado S, et al. Chewing number is related to incremental increases in body weight from 20 years of age in Japanese middle-aged adults. Gerodontology 2013 Sep;30(3):214-219. [doi: 10.1111/j.1741-2358.2012.00666.x] [Medline: 22607548]
41. Smit HJ, Kemsley EK, Tapp HS, Henry CJK. Does prolonged chewing reduce food intake? Fletcherism revisited. Appetite 2011 Aug;57(1):295-298. [doi: 10.1016/j.appet.2011.02.003] [Medline: 21316411]
42. Walden HM, Martin CK, Ortego LE, Ryan DH, Williamson DA. A new dental approach for reducing food intake. Obes Res 2004 Nov;12(11):1773-1780 [FREE Full text] [doi: 10.1038/oby.2004.220] [Medline: 15601972]
43. Bolhuis DP, Lakemond CMM, de Wijk RA, Luning PA, de Graaf C. Consumption with large sip sizes increases food intake and leads to underestimation of the amount consumed. PLoS One 2013;8(1):e53288 [FREE Full text] [doi: 10.1371/journal.pone.0053288] [Medline: 23372657]
44. Xu J, Xiao X, Li Y, Zheng J, Li W, Zhang Q, et al. The effect of gum chewing on blood GLP-1 concentration in fasted, healthy, non-obese men. Endocrine 2015 Sep;50(1):93-98 [FREE Full text] [doi: 10.1007/s12020-015-0566-1] [Medline: 25758865]
45. Sonoki K, Iwase M, Takata Y, Nakamoto T, Masaki C, Hosokawa R, et al. Effects of thirty-times chewing per bite on secretion of glucagon-like peptide-1 in healthy volunteers and type 2 diabetic patients. Endocr J 2013;60(3):311-319 [FREE Full text] [doi: 10.1507/endocrj.ej12-0310] [Medline: 23138354]
46. Read NW, Welch IM, Austen CJ, Barnish C, Bartlett CE, Baxter AJ, et al. Swallowing food without chewing; a simple way to reduce postprandial glycaemia. Br J Nutr 1986 Jan;55(1):43-47. [doi: 10.1079/bjn19860008] [Medline: 3311145]
47. Ranawana V, Leow MK, Henry CJK. Mastication effects on the glycaemic index: impact on variability and practical implications. Eur J Clin Nutr 2014 Jan;68(1):137-139. [doi: 10.1038/ejcn.2013.231] [Medline: 24219890]
48. Sato A, Ohtsuka Y, Yamanaka Y. Morning mastication enhances postprandial glucose metabolism in healthy young subjects. Tohoku J Exp Med 2019 Nov;249(3):193-201 [FREE Full text] [doi: 10.1620/tjem.249.193] [Medline: 31761819]
49. Kataoka M, Venn BJ, Williams SM, Te Morenga LA, Heemels IM, Mann JI. Glycaemic responses to glucose and rice in people of Chinese and European ethnicity. Diabet Med 2013 Mar;30(3):e101-e107. [doi: 10.1111/dme.12080] [Medline: 23181689]

50. Hamada Y, Miyaji A, Hayashi N. Effect of postprandial gum chewing on diet-induced thermogenesis. Obesity (Silver Spring) 2016 Apr;24(4):878-885 [FREE Full text] [doi: 10.1002/oby.21421] [Medline: 26887536]
51. Bolhuis DP, Lakemond CMM, de Wijk RA, Luning PA, Graaf CD. Both longer oral sensory exposure to and higher intensity of saltiness decrease ad libitum food intake in healthy normal-weight men. J Nutr 2011 Dec;141(12):2242-2248. [doi: 10.3945/jn.111.143867] [Medline: 22049294]
52. Macht M. How emotions affect eating: a five-way model. Appetite 2008 Jan;50(1):1-11. [doi: 10.1016/j.appet.2007.07.002] [Medline: 17707947]

Abbreviations

AC: abdominal circumference
BP: blood pressure
GLU: fasting glucose concentration
HDL-c: high-density lipoprotein cholesterol
MetS: metabolic syndrome
SL: serum lipids

Edited by L Buis; submitted 25.06.21; peer-reviewed by M Simione, YL Leung; comments to author 23.07.21; revised version received 16.10.21; accepted 18.02.22; published 24.03.22.

Please cite as:
Uehara F, Hori K, Hasegawa Y, Yoshimura S, Hori S, Kitamura M, Akazawa K, Ono T
Impact of Masticatory Behaviors Measured With Wearable Device on Metabolic Syndrome: Cross-sectional Study
JMIR Mhealth Uhealth 2022;10(3):e30789
URL: https://mhealth.jmir.org/2022/3/e30789
doi:10.2196/30789
PMID:35184033

Corrigenda and Addenda

Correction: Social, Organizational, and Technological Factors Impacting Clinicians' Adoption of Mobile Health Tools: Systematic Literature Review

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

[1]Anglia Ruskin University, Cambridge, United Kingdom

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

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

Corresponding Author:
Christine Jacob, BSc, MSc, PhD
Anglia Ruskin University
Cambridge Campus
East Rd
Cambridge, CB1 1PT
United Kingdom
Phone: 44 798769376
Email: christine.k.jacob@gmail.com

Related Article:

Correction of: https://mhealth.jmir.org/2020/2/e15935/

(JMIR Mhealth Uhealth 2022;10(3):e37747) doi:10.2196/37747

In "Social, Organizational, and Technological Factors Impacting Clinicians' Adoption of Mobile Health Tools: Systematic Literature Review" (JMIR Mhealth Uhealth 2020;8(2):e15935), one error was noted.

In the originally published manuscript, Table 4 was incorrectly displayed as an identical copy of Table 3. The complete, corrected version of Table 4 has now been included in the corrected version of the manuscript.

The full corrected table is included below.

The correction will appear in the online version of the paper on the JMIR Publications website on March 10, 2022, 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.

Table 4. Patient-related factors and their occurrence, with references.

Factor	Subthemes	References
Quality and efficiency of care (n=77)	Examples: treatment outcomes, clinical delivery, patient monitoring, and treatment compliance	[1-3,5,9,10,34,36-40,42-44,47-49,52-57,61-64,66,69,72,75,78,80,83,86,88,89,92,93, 100,103,107-109,114,115,120,122,125,127,130,131,137,140,146,157,159, 160,162,165,166,168,169,174,175,180,182-186,188,190,194,196]
Provider-patient communication (n=53)	Quality and ease of communication between patients and the care team	[3,5,9,10,34,36,41,43,46,47,49,50,52,55,58,59,66,67,72,73,75,77,78,80-82, 84-86,88-91,103,115,122,127,128,150,156,160,163,172,174,175,180,182,185,189, 190,194,195]
Access to care (n=41)	Enhancing patients' access to care and reaching the underserved	[1-5,8,34,39,43,47,54,55,60,62,64,67,73,75,78,81,82,91,93,96,99,102, 123,129,130,156,159,162,169,174,178,180,182,184,187,189,190, 194,195]
Patient consent, comfort, and preference (n=30)	Comfort with technology, personal preferences, and the ease of getting an informed consent from the patients	[1-3,9,39,43,49,52,53,60,61,66,79,86,90,104,122,123,132,153,163, 174,176,180,184,185,191,192,194,195]
Applicability and appropriateness (n=22)	The suitability of patients on the basis of their needs and characteristics	[9,41,57,61,62,66,68,71,73,75,79,81,86,88-91,103,104,108,162,182,185,189]
Empowerment and engagement (n=21)	Opportunity to empower and reassure patients and increase their engagement in managing their condition	[5,34,41,48,62,70,71,73,77,78,88,89,115,120,128,157,162,166,175,180,190]
Safety (n=19)	Patient safety and the safety of clinical practice	[10,44,63,66,72,78,81,86,103,109,140,147,155,176,179,188-190,195]
Digital divide (n=15)	Age, living standard, and access to technology	[49,53,55,60,62,73,75,79,81,84,120,123,138,185,191]
Education (n=12)	Better patient education and awareness	[52,53,60,75,81,86,88,162,164,168,183,190]
Service abuse, overreliance (n=8)	Patient overdependence on practitioner support	[3,41,62,77,78,156,169,182]
Data and surveillance–related anxiety (n=6)	Worries and anxiety related to the understanding and interpretation of data, or the feeling of being observed	[46,62,78,128,140,190]
Sustainability (n=3)	Long-term commitment and use	[5,167,169]
Gate keeping by clinicians (n=2)	Protective or paternalistic attitudes of the care team	[97,180]

Submitted 04.03.22; this is a non–peer-reviewed article;accepted 04.03.22; published 10.03.22.

Please cite as:
Jacob C, Sanchez-Vazquez A, Ivory C
Correction: Social, Organizational, and Technological Factors Impacting Clinicians' Adoption of Mobile Health Tools: Systematic Literature Review
JMIR Mhealth Uhealth 2022;10(3):e37747
URL: https://mhealth.jmir.org/2022/3/e37747
doi:10.2196/37747
PMID:35271457

Original Paper

Using Momentary Assessment and Machine Learning to Identify Barriers to Self-management in Type 1 Diabetes: Observational Study

Peng Zhang[1,2], PhD; Christopher Fonnesbeck[3], PhD; Douglas C Schmidt[1,2], PhD; Jules White[1], PhD; Samantha Kleinberg[4], PhD; Shelagh A Mulvaney[5,6,7], PhD

[1]Department of Computer Science, School of Engineering, Vanderbilt University, Nashville, TN, United States
[2]Data Science Institute, Vanderbilt University, Nashville, TN, United States
[3]Department of Biostatistics, Vanderbilt University, Nashville, TN, United States
[4]Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ, United States
[5]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States
[6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[7]School of Nursing, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Peng Zhang, PhD
Department of Computer Science
School of Engineering
Vanderbilt University
1400 18th Ave S
Rm 2023
Nashville, TN, 37212
United States
Phone: 1 615 343 8630
Email: peng.zhang@vanderbilt.edu

Abstract

Background: For adolescents living with type 1 diabetes (T1D), completion of multiple daily self-management tasks, such as monitoring blood glucose and administering insulin, can be challenging because of psychosocial and contextual barriers. These barriers are hard to assess accurately and specifically by using traditional retrospective recall. Ecological momentary assessment (EMA) uses mobile technologies to assess the contexts, subjective experiences, and psychosocial processes that surround self-management decision-making in daily life. However, the rich data generated via EMA have not been frequently examined in T1D or integrated with machine learning analytic approaches.

Objective: The goal of this study is to develop a machine learning algorithm to predict the risk of missed self-management in young adults with T1D. To achieve this goal, we train and compare a number of machine learning models through a learned filtering architecture to explore the extent to which EMA data were associated with the completion of two self-management behaviors: mealtime self-monitoring of blood glucose (SMBG) and insulin administration.

Methods: We analyzed data from a randomized controlled pilot study using machine learning–based filtering architecture to investigate whether novel information related to contextual, psychosocial, and time-related factors (ie, time of day) relate to self-management. We combined EMA-collected contextual and insulin variables via the MyDay mobile app with Bluetooth blood glucose data to construct machine learning classifiers that predicted the 2 self-management behaviors of interest.

Results: With 1231 day-level SMBG frequency counts for 45 participants, demographic variables and time-related variables were able to predict whether daily SMBG was below the clinical threshold of 4 times a day. Using the 1869 data points derived from app-based EMA data of 31 participants, our learned filtering architecture method was able to infer nonadherence events with high accuracy and precision. Although the recall score is low, there is high confidence that the nonadherence events identified by the model are truly nonadherent.

Conclusions: Combining EMA data with machine learning methods showed promise in the relationship with risk for nonadherence. The next steps include collecting larger data sets that would more effectively power a classifier that can be deployed to infer

individual behavior. Improvements in individual self-management insights, behavioral risk predictions, enhanced clinical decision-making, and just-in-time patient support in diabetes could result from this type of approach.

(JMIR Mhealth Uhealth 2022;10(3):e21959) doi:10.2196/21959

KEYWORDS

machine learning; type 1 diabetes; psychosocial; self-management; adolescents; behavioral medicine; ecological momentary assessment; informatics; mobile phone

Introduction

Background

Type 1 diabetes (T1D) is a prevalent chronic illness, with increasing incidence rates reported worldwide [1,2]. It is an autoimmune disorder in which the body does not produce insulin and requires patients to perform critical self-management tasks multiple times per day [3]. Two key self-management tasks in T1D are frequent monitoring of blood glucose (BG) and administration of insulin. These tasks help manage glycemic control to avoid or delay serious short- and long-term consequences, such as retinopathy, neuropathy, and mortality [4-6]. Mealtime is a critical time for diabetes self-management.

Adolescents and young adults have the worst glycemic control of any age group [4]. For young people with diabetes, living successfully with T1D is particularly hard because of many potential psychosocial and contextual barriers to self-management [7-9]. A recommended approach to improve self-management involves promoting and supporting problem-solving skills to reduce barriers [10]. To identify problems related to self-management, patients, caregivers, and clinicians must rely on BG and insulin administration data from devices along with a patient recall of behavioral, emotional, and contextual events that could pose barriers to self-management. However, using retrospective memory or recall for events that are days or weeks in the past has been identified as generally unreliable and potentially biased in nature [11]. Unreliable recall of events in diabetes problem-solving could result in incorrect modifications to the insulin regimen.

To address the limitations of human recall and bias in health behavior research, ecological momentary assessment (EMA) methods have been developed and successfully used in a range of health conditions. In contrast to traditional assessment methods, EMA uses more frequent and in vivo ambulatory assessments of factors that affect health behaviors and decision-making. EMA methods provide a more proximal, and often more accurate, technology-mediated method to monitor and assess the contexts, subjective experiences, and processes that surround health decisions in daily life [12,13]. In particular, EMA methods provide more relevant and frequent observations per person and generate rich data to assess correlates of health behavior more accurately and identify novel correlates for intervention [14].

Many studies in the EMA literature typically use mixed effects or hierarchical linear modeling [15,16]. This analytic approach does not provide a means to automate analyses or use learning algorithms that improve and integrate incoming data over time. A promising approach for identifying such a model involves integrating EMA with techniques and tools associated with machine learning, which is a data analysis method that automates statistical model building by identifying patterns and making decisions with minimal human intervention [17,18]. Machine learning has been used with wearable sensor data and may also be useful in analyzing intensive self-report data, such as EMA. Machine learning techniques provide a viable means of examining both big and small data by providing automated classification and prediction for more feasible behavioral interventions.

Objective

The objective of our study is to develop a machine learning algorithm to predict the risk of missed self-management. We seek to identify the momentary psychosocial and contextual factors that have an impact on T1D self-management, as assessed by EMA. To achieve these objectives, we train and compare a number of machine learning models through a learned filtering architecture (LFA) to explore the extent to which EMA data could predict the completion of two self-management behaviors: insulin administration and self-monitoring of blood glucose (SMBG). By integrating these two strategies (EMA and machine learning), we aim to provide researchers with not only a better understanding of what may hinder or promote adolescents' adherence to their T1D regimen from a behavioral perspective but also an efficient and adaptive analytic computational method.

Methods

Study Design and Setting

These subanalyses analyzed data from a feasibility trial of the mobile EMA and feedback app called MyDay, which is a self-management feedback and problem-solving tool designed for adolescent patients with T1D [19]. Youth from the Vanderbilt Eskind Pediatrics Diabetes Clinic were invited to participate in a 30-day assessment period if (1) they were aged between 13 and 19 years, (2) had been diagnosed with T1D for at least 6 months, (3) owned either an Android or iPhone smartphone, (4) understood and spoke English, and (5) were willing to use a Bluetooth BG meter during the study [1]. The study was reviewed and approved by the Vanderbilt University institutional review board (IRB #150685). All parents provided consent before the adolescents provided assent. Both consent and assent were obtained before the study procedures commenced.

Participants

A total of 48 participants were recruited for the pilot study. Of the 48 participants, 3 (6%) dropped out of the study, noting

competing demands, leaving 45 (94%) for our analyses. Participants were randomized in a 2:1 ratio to the MyDay app + Bluetooth BG (meter group 31/45, 69%) and a control group (14/45, 31%). The control group provided SMBG data only using Bluetooth BG meters but did not use the MyDay app. Design processes, engagement, and momentary relationship results for MyDay have been published previously [19-21].

Momentary Assessments and Glucose Meter Data

All SMBG data were objectively assessed using iHealth [22] glucometers. The iHealth glucometers are commercially available Bluetooth low-energy meters that can upload data automatically to the iHealth secure cloud server via their open application programming interface. Of the 45 participants, 31(69%) participants were instructed to use the MyDay app at each mealtime and bedtime to answer questions that focused on factors likely to affect diabetes self-management.

MyDay provided notifications to complete the EMA assessment personalized to typical mealtimes identified by participants. Time stamps were associated with all data entries. Only mealtime EMA was used in analyses. Variables analyzed in relation to self-management outcomes were organized into subsets. The first two domains of variables were collected for all participants: (1) *demographics* obtained at baseline (ie, gender, age, father's education, mother's education, family income, and race) and (2) *time variables* that were coded using the original time stamps of the collected data entries (eg, weekday, weekend, and mealtime [breakfast, lunch, and dinner]).

The next three domains of EMA data were available only for the 31 participants using the MyDay app: (3) *social context* related to who was with the youth at the time of self-management (ie, parent, sibling, alone, casual friend, close friend, other family, other person, strangers, and boyfriend or girlfriend) and where the youth was at the time of self-management (ie, home, school, work, restaurant, friends' house, or on the road); (4) *stress*, *fatigue*, and *mood* levels at the reported self-management event, scored as 0 to 100, with higher scores indicating greater stress, more fatigue, and worse negative mood; and (5) selected situational *barriers* at the time of self-management event (ie, participant was rushing, feeling sick, on the road, hungry, wanting privacy, busy, without supplies, or having fun). Details of the EMA data collection process can be found in the study by Zhang et al [20].

Outcomes

We examined three self-management behavioral outcomes:

1. Daily SMBG frequency of *<4* or *≥4* times a day; 4 glucose checks per day are generally considered as the minimum recommended [23]
2. Missed SMBG at mealtimes
3. Insulin administration at mealtimes

Data from all 45 participants were available for analyses examining the daily number of SMBG from meters. The data available for all participants were demographic and time variables. Analyses for outcomes 2 and 3 examined data from participants who used the MyDay EMA app (31/45, 69%), which obtained mealtimes.

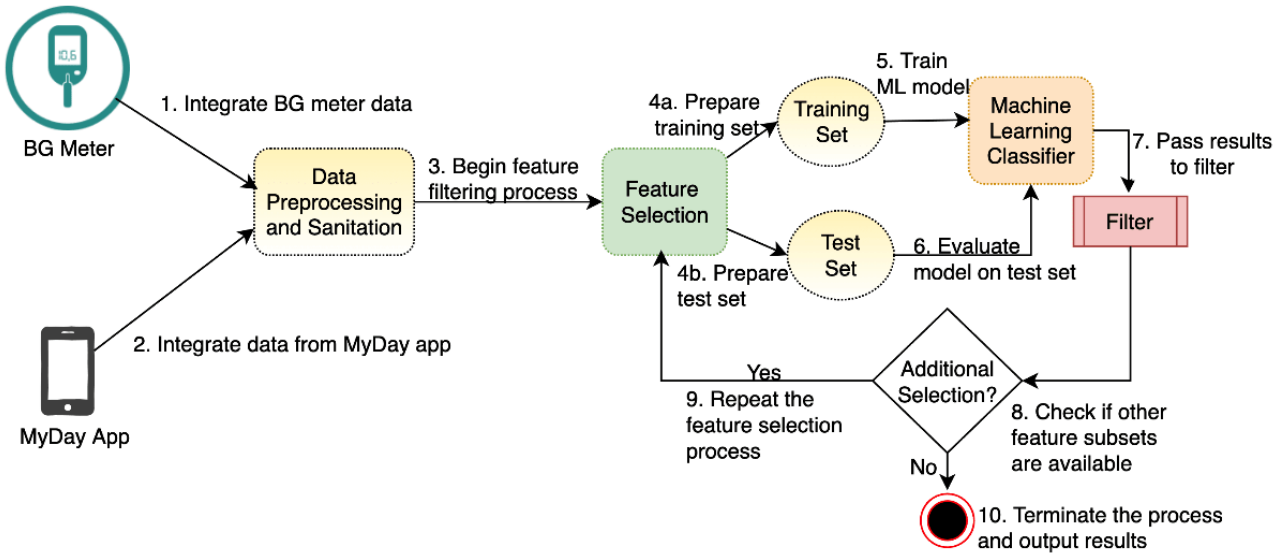
LFA Approach

To extract domains of variables to predict insulin administration and SMBG self-management behaviors via the training of a series of models, an LFA was created in this study as a byproduct, and a similar process was used in the study by Zhang et al [24] but not formally constructed. For this study, the LFA created and compared four machine learning models: k-nearest neighbors (KNN), logistic regression, random forest (RF), and support vector machines. These models performed binary classification for each behavioral outcome observed in this study.

KNN classifies each sample by finding its K-most similar instances in the training set and chooses the class to which most neighboring instances belong [25]. The value of k is determined by running KNN models with varying k values iteratively and selecting the k value that produces the most optimal model. Logistic regression is a statistical model that classifies a sample by predicting the probability of an output using the maximum likelihood estimation method and using a probability threshold (P=.50 was used in our study as the threshold such that an output with a probability of P≥.50 was classified as true and false otherwise) to separate the 2 classes [26]. RF is a popular ensemble learning method that trains multiple decision trees on different parts of the data set and then averages the results to improve classification accuracy [27]. The number of trees, or *estimators*, is determined by running a number of RF models with varying estimator values, such as 10, 50, and 100, and selecting the value that produced the most performant model. Support vector machines work by finding an optimal hyperplane in the feature space that optimally separates the data points into different classes [28].

Figure 1 presents the workflow of this LFA and shows that the SMBG data and EMA data collected from the MyDay app were integrated as a complete data set fed into the LFA (steps 1 and 2). The LFA then performed specified data preprocessing, such as normalizing numeric values, removing entries that were empty or had many missing features, and one-hot encoding based on the type of each column (step 3). After step 3, a data filtering process began, where subsets of variables were extracted from the cleaned data either based on configurable user input, such as the names of columns that would be grouped to create a clinically meaningful, or to-be-observed, feature subset. The features were grouped as described above to create multiple data subsets. Owing to the small sample size of the data available, the data subsets were each split further for evaluating each classification model using cross-validation (steps 4a and 4b).

Figure 1. Iterative process of the learned filtering architecture. BG: blood glucose; ML: machine learning.



The LFA calculates the distribution of the target variables of each data set. If the data set is balanced, it evaluates each model using k-fold cross-validation that further splits the data into training and validation sets *k* times and produces mean values of the performance metrics. Otherwise, if the classes are unevenly distributed, it uses the stratified k-fold cross-validation to create *k* (*k*=7) splits, with each split of training and validation sets maintaining the original class distributions. The performance metrics are averaged across the results from the *k* different splits. The process is then repeated for each of the specified machine learning models (step 6).

Specifically, we used the following metrics to assess the models: (1) accuracy, which is the percentage of correct predictions; (2) precision, which is the ratio of true positives and all predicted positives that evaluates what proportion of predicted positives was actually correct; (3) recall, which is the ratio of true positives and all actual positives and calculates what proportion of actual positives was predicted correctly; (4) F1 score, which evenly weighs precision and recall; and (5) for imbalanced classification tasks, the Brier score, which is a continuous scoring loss function that evaluates the goodness of predicted probabilities in a classification task—a lower number corresponds to a stronger model and vice versa.

The classification results were then used by the filter component to compare them across all feature subsets (step 7). The filter component had a configurable tolerance value that was used to select feature subsets with relatively good classification results compared with the best-performing models. Next, the LFA checked whether additional feature groups remained to be processed (step 8). If so, feature selection was repeated to create the next data subset (step 9). Otherwise, the filtering process would terminate and output the filtered results; that is, variable groups with relatively strong predictive power of the outcomes (step 10).

The classification results were filtered to extract the best predictor groups for the target class variable. For example, if the overall performance metrics exceeded the specified threshold values (such as 15% compared with the performance metrics of the model trained with all features together), the predictor group was added to the final output queue. When all variable groups were evaluated, LFA returned the final insights obtained from the input; that is, feature groups that had significant predictive power for the outcomes observed in this study.

Although the number of observations per participant was substantial (average number of observations 60), the overall number of participants was relatively small (n=45). Thus, the collected data had some imbalance in the distribution of the outcomes, with missed mealtime insulin being a relatively less frequent event. Classification models constructed using imbalanced data sets may result in the minority class being neglected [29]. Techniques such as Synthetic Minority Oversampling Technique [30] and Tomek link [31] have been used in the literature for training imbalanced data, especially for small data sets [32-35]. However, given the small size of the population in this study, using such sampling methods would risk introducing bias and misleading results. Therefore, in this study, we used a stratified K-fold (*k*=7) cross-validation [36] evaluation method instead of random oversampling or introducing synthetic samples based on the existing data.

In stratified K-fold cross-validation, the original data set was randomly split into *k* folds. Each fold was further split into separate training and testing sets that are used to generate the evaluation metrics of a model. The distributions of the majority and minority classes within each training and testing set follow the distribution of the majority and minority classes in the original data set. After the model was trained and tested against all *k* folds, the results were averaged to represent the overall classification performance.

In addition to the machine learning methods previously described, we also used a Bayesian hierarchical regression model for the entire EMA data set that has a large number of features but a small sample size. This approach was applied to confirm the inferential power of the collected EMA data rather than focusing on which specific category was the most predictive of the outcomes.

Hierarchical modeling can capture the similarities of multiple participants within a data set while allowing estimations of individual parameters for data containing multiple participants. With the Bayesian approach, the entire data set is considered known information that is used to derive the distributions of unknown parameters of the model. It is a probabilistic model that intends to estimate the expected values or density.

In our analysis, we applied Markov chain Monte Carlo methods [37] to assist with the model formation and sampling process. Monte Carlo is a method for randomly sampling a probability distribution to approximate the desired target function. Markov chain is a sampling technique that can generate a sequence of random samples where the current sample is drawn based on the prior sample. The goal of the Markov chain Monte Carlo is to construct a Markov chain that eventually stabilizes on the desired quantity to be inferred. Specifically, we created a noncentered Bayesian hierarchical model to estimate the likelihoods of SMBG and insulin administration.

Results

Overview

This section first reports findings from our initial statistical analysis and then analyzes the results obtained from the LFA constructed in accordance with the methods described in the previous sections. Table 1 shows the characteristics of the sample.

Table 1. Characteristics of the sample (N=45).

Variable	Values
Age (years), mean (SD)	13.3 (1.7)
Female, n (%)	24 (53)
Race or ethnicity, n (%)	
White	38 (84)
African American	4 (10)
Asian	1 (2)
Hispanic	1 (2)
Other	0 (0)
Father's education, n (%)	
Less than high school	1 (2)
High school or GED[a]	13 (29)
2-year college	7 (16)
4-year college	15 (33)
Graduate degree	5 (11)
N/A[b]	4 (9)
Mother's education, n (%)	
Less than high school	0 (0)
High school or GED	10 (22)
2-year college	12 (27)
4-year college	17 (38)
Graduate degree	2 (4)
N/A	12 (27)
Household income (US $), n (%)	
<25,000	2 (4)
25,001-35,000	3 (7)
35,001-75,000	7 (16)
75,001-100,000	14 (31)
>100,000	3 (7)
N/A	4 (9)
Duration of diabetes (years), mean (SD)	5.5 (3.7)
HbA_{1c}[c], mean (SD)	9.0 (1.9)
Use insulin pump (yes), n (%)	26 (58)

[a]GED: General Educational Development.

[b]N/A: missing values.

[c]HbA_{1c}: hemoglobin A_{1c}.

Statistical Analysis

The data set was preprocessed using statistical approaches. First, it was observed that the data set contained missing values in demographic features: 9% (5/45) missing for both father's education and household income categories and 27% (12/45) missing for mother's education category (the percentage of missing values in each category is denoted as "N/A" entry in our report). In this study, the missing values of a feature were imputed using the mode value for features of mother's education and father's education and the median value for the feature of family income. Ordinal categorical variables whose order of the values were significant, such as parent education and family income level, were each transformed into a single feature with numeric values, whereas nominal variables whose significance could be assumed, such as participant race and day of the week,

were converted to numeric values using one-hot encoding. Each feature was normalized using the minimum–maximum scaler such that all the final values of that feature were between 0 and 1. The source code for data preprocessing is included in Multimedia Appendix 1.

Tables 2-4 display the summary statistics of features that have $P<.05$ (ranked in ascending order) for the target feature (or dependent variable) of daily SMBG frequency, missed glucose, and insulin not administered categories, respectively. *P* value is an initial indicator that the corresponding features are statistically significant in our analysis: (1) for daily SMBG frequency, most features reported in Table 2 belong to the demographic group; (2) for SMBG, variables from the demographics, social context, barriers, and stress or mood or energy feature groups are reported in Table 3; (3) for insulin administration, variables from groups of demographics, time variables, stress or mood or energy, and barriers are reported in Table 4.

Table 2. Summary statistics of features with statistical significance on daily self-monitoring of blood glucose frequency.

Feature	Coefficient	SE	*P* value
Mother's education	0.5221	0.062	<.001
Age	−0.2494	0.057	<.001
Male	0.2721	0.032	<.001
Father's education	−0.1691	0.066	.01

Table 3. Summary statistics of features with statistical significance on self-monitoring of blood glucose.

Feature	Coefficient	SE	*P* value
Busy	0.1706	0.041	<.001
No supplies	0.7417	0.089	<.001
Other family	0.1436	0.038	<.001
Gender	−0.1543	0.019	<.001
Mother's education	−0.1835	0.033	<.001
Income	−0.2569	0.039	<.001
Parent	−0.0785	0.026	<.001
Black race	−0.1064	0.038	.01
Casual	−0.084	0.031	.01
Father's education	0.0906	0.035	.01
With sibling	0.0522	0.02	.01
In restaurant	−0.2582	0.106	.02
Hungry	−0.0436	0.021	.04
Other place	−0.2177	0.108	.045
Stress+energy	0.9274	0.466	.047

Table 4. Summary statistics of features with statistical significance on insulin administration.

Feature	Coefficient	SE	*P* value
Hungry	−0.0958	0.021	<.001
No supplies	0.3703	0.091	<.001
Breakfast	0.1134	0.021	<.001
Mother's education	−0.145	0.034	<.001
Black race	−0.1637	0.039	<.001
Diabetes burnout	0.1495	0.047	<.001
Third day of week	−0.2369	0.077	<.001
Lunch	0.0695	0.022	<.001
Busy	0.1219	0.043	<.001
Second day of week	−0.216	0.077	.01
Fourth day of week	−0.2146	0.077	.01
Weekend	−0.1999	0.078	.01
Fatigue	0.0508	0.02	.01
Fifth day of week	−0.1765	0.077	.02
Low blood glucose	0.0849	0.039	.03
Gender	−0.0425	0.02	.03
Mood	−0.0919	0.043	.03
Sixth day of week	−0.1602	0.077	.04

Daily SMBG Frequency

The average age of all participants was 13 (SD 1.7) years; 53% (24/45) were female, 84% (37/45) were White, 58% (26/45) used an insulin pump, and participants had a mean hemoglobin A_{1c} (indicating overall glycemic control) of 9.03% (SD 1.91). Additional characteristics of the sample are summarized in Table 4.

A total of 4475 BG measurements were obtained from the iHealth Bluetooth meters used by all participants (n=45). For this analysis, the demographic and time variables were studied to identify if they had any impact on the outcome of SMBG frequency per day. The measurements were aggregated on a daily basis to obtain a new data set of 1231 entries, with each entry per participant being the total number of measurements an individual had each day during the study period. SMBG frequency ranged from 1 to 12 measurements per day. If a participant did not report an entry on a particular day, the entry for that day was not assumed to have an SMBG daily frequency of 0, and hence, the entry for the participant on that day was not created.

Several distributions of daily SMBG frequencies were observed. There were 591 entries with *<4* frequency and 640 entries with *≥4 or*. Of all the classifiers trained with the same training data, RF was the best performing model based on the overall classification metrics using the same test data. The mean and SD values of the evaluation results from the best-performing RF model are shown in Table 5 for SMBG frequencies *<4* (the source code comparing the performance of all machine learning models is included in Multimedia Appendix 1). The filter then compared the benchmark value with the outcome classification results obtained from each variable group. A tolerance value of 15% was configured for the filter to select subsets with significant predictive power. As shown in Table 5, the demographic variable group for SMBG frequency resulted in a better performance than time variables and all variables.

Table 5. Self-monitoring of blood glucose <4 classification results.

Feature group	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	F1 score, mean (SD)
Demographics	75% (0.04)	75% (0.08)	72% (0.07)	74% (0.06)
Time variables	49% (0.04)	46% (0.06)	21% (0.14)	28% (0.12)
All	68% (0.03)	67% (0.06)	68% (0.06)	67% (0.03)

Missed Mealtime SMBG and Insulin Administration

From the app group (31/45, 69%), a total of 1869 entries were associated with breakfast, lunch, or dinner and used to analyze factors that could affect SMBG and insulin administration. *Missed insulin administration* had a distribution of 1:5.72 for true (missed) versus false (administered) outcomes. In contrast, the outcome *missed SMBG* had a class distribution of 1:5.44 for true (missed) versus false (checked). LFA created classification models for each variable group (ie, demographic,

time, social context, and psychosocial) using the stratified K-fold approach, as discussed previously. Similar to the previous experiment, the RF model resulted in the best classification performance in all metrics compared with other models (the source code comparing the performance of all machine learning models is included in Multimedia Appendix 2).

Tables 6 and 7 present the classification results of missed SMBG and missed insulin administration, respectively. The results showed mixed sentiments on the predictive power of individual groups of indicators on self-management behavior; however, their combined effect can be used to infer when the lack of SMBG or insulin administration occurred with high accuracy and high precision.

Table 6. Missing mealtime blood glucose measurement classification results.

Feature group	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Brier test (%)
Demographics	78	38	62	47	22
Time variables	50	13	42	20	51
Social context	61	21	55	30	25
Stress, fatigue, and mood	74	22	29	25	33
Barriers	73	33	44	33	25
All	88	78	35	48	12
All (MCMC[a])	87	78	25	38	13

[a]MCMC: Markov chain Monte Carlo.

Table 7. Missing mealtime insulin administration classification results.

Feature group	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Brier test (%)
Demographics	65	25	65	36	36
Time variables	59	21	64	32	41
Social context	49	16	59	25	51
Stress, fatigue, and mood	74	22	28	25	32
Barriers	73	26	44	32	27
All	86	61	14	23	14
All (MCMC[a])	85	54	15	24	15

[a]MCMC: Markov chain Monte Carlo.

Discussion

Principal Findings

To better understand the factors affecting the self-management behavior of adolescents with T1D, this study applied machine learning analyses to construct an LFA using demographic, BG, and momentary psychosocial and self-management data. The relative association of the 5 domains of variables for the predictability of self-management behaviors was compared using all the variables collectively as the benchmark.

For the demographic data, the results indicated that demographics were most associated with average daily SMBG frequency. These results highlight the value of social determinants of health, as defined by demographics. Although demographic factors are generally not modifiable, social determinants of health are increasingly used to adapt care to those who are most vulnerable and may not receive the full benefit of current approaches to health care [36,37].

The EMA data were able to infer nonadherence to SMBG and insulin with high accuracy and precision. Although the recall score was low, there was high confidence that the nonadherence events identified by the model are truly nonadherent. A reason for the lower recall score has to do with the small data sets that have disparities in the frequencies of observed classes or outcomes. Nonetheless, this study shows promise in the collection of larger data sets that would more effectively power a classifier that is deployable in the real world. These results also concord with our reported results from the initial statistical analysis in that (1) demographic features are correlated with daily SMBG frequencies; (2) features from each group, except for time points, have a statistically significant impact on SMBG; and (3) features from each group, except for social context, have statistically significant inferential power on insulin administration.

These results support the feasibility and value of integrating EMA and machine learning to improve behavioral assessment and automate behavioral pattern recognition in health care [18,38]. Our learned models show promise in quantifying the impact of psychosocial factors on self-management. In diabetes, stress and mood are modifiable factors that may be positively influenced by coping and problem-solving interventions [39,40]. The use of machine learning and EMA was also seen in a recent study on tinnitus (the phantom perception of sounds), where an

RF classifier was applied on EMA data collected from the TrackYourTinnitus mobile app across devices to predict the mobile operating system used [41].

Social context also provided a framework for understanding risk and may be modified by interventions focused on social competence and problem solving [39]. In previous studies [42,43], behavioral observations were used to identify patterns of hand hygiene compliance monitoring, from which we obtained useful initial insights into which domains of variables had the most impact on compliance behavior.

Moving forward, the use of primarily intensive self-reported and passive psychosocial and behavioral data streams combined with machine learning could provide the basis for population-based monitoring systems to help guide automated pattern detection for clinical risk management. For example, experimental unobtrusive indicators of mealtimes are in development [44], and insulin administration is available via pumps [44]. If successful, additional passive data streams would greatly improve our methodological rigor and reach [45].

The LFA machine learning methods used here should be applied to a large, diverse sample of patients to confirm and expand the results reported in this paper. Although passive methods are increasingly used to infer behavior and psychosocial status [46,47], there are important subjective experiences, such as mood, which may continue to require self-reporting. For the foreseeable future, both self-reported real-time data and passive data, such as social networking [48], may be integrated to optimize insights for health care.

Prior research using traditional retrospective questionnaire methods has focused largely on identifying psychosocial correlates and predictors of self-management in chronic illness in general and specifically in diabetes [9]. With a few exceptions, little research using EMA has been conducted on diabetes. The few studies conducted have uniquely identified time-based factors, such as time of day and momentary negative mood, as related to self-management behaviors [49-51].

Machine learning analyses have been applied in various studies, focusing largely on the improvement of diabetes management and control. Earlier studies have constructed and fine-tuned different machine learning models to predict future BG levels based on historical physiological data [52-54], detect incorrect BG measurements [55], predict hypoglycemia [56,57], and manage insulin dosing [58] and applied it to provide lifestyle support integrating food recognition and energy expenditure [59,60]. The study results reported here advance the assessment and analysis of factors previously associated with self-management, including stress [49], mood [61,62], stigma [9,63], and social contexts [8,12]. Our study also uniquely assesses novel factors not previously studied in the T1D population, such as fatigue [64], location [65], social contexts [8], and contextual factors, such as rushing and traveling. The collected EMA data have a promising ability to infer the 2 diabetes self-management behaviors under study.

Limitations

This study had several limitations. First, although intensive assessment resulted in a substantial number of observations per participant, the number of participants was relatively small. Although the inferential ability of this data was identified during our empirical analysis, a larger sample size in future iterations will help produce higher quality results. Second, some of the data collected here using momentary self-report, such as stress, may eventually become available as feasible passive data streams. This could reduce the burden of momentary assessment for participants and enhance the accuracy and reliability of the data. Consideration of burden should influence behavioral sampling strategies and research designs using momentary assessment. Finally, this study used a self-report of insulin administration. Moving forward, integration of insulin pumps or automated insulin administration systems will be necessary to infer insulin dosing and timing accurately.

Conclusions

On the basis of the current findings, psychosocial context may be successfully assessed using momentary assessment, combined with physiological data, and analyzed using machine learning to optimize, and ultimately automate, health behavior insights. Similar experiments are needed with larger samples to prioritize multiple potential domains of influence on health behaviors and advance the assessment and analytic approaches used here. Future work validating self-reporting with sensor data will enhance our ability to use passive indicators of health-related behaviors. For example, experimental unobtrusive indicators of mealtimes are in development and, if successful, would greatly enhance our methodological approach [45]. The LFA machine learning methods used here will be applied to a large, diverse sample of patients to confirm and expand the results reported in this paper.

Acknowledgments

This work was supported by a grant to author SAM from the National Institutes of Health (United States); National Institute for Diabetes, Digestive, and Kidney Disease DP3 DK097706; and the National Center for Advancing Translational Sciences (United States) UL1 TR000445.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Source code for comparing models of daily self-monitoring of blood glucose frequency.

[DOCX File , 154 KB - mhealth_v10i3e21959_app1.docx]

Multimedia Appendix 2
Source code for comparing models of self-monitoring of blood glucose and insulin administration.
[DOCX File , 8530 KB - mhealth_v10i3e21959_app2.docx]

References

1. Borchers AT, Uibo R, Gershwin M. The geoepidemiology of type 1 diabetes. Autoimmun Rev 2010 Mar;9(5):A355-A365. [doi: 10.1016/j.autrev.2009.12.003] [Medline: 19969107]
2. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, SEARCH for Diabetes in Youth Study. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. JAMA 2014 May 07;311(17):1778-1786 [FREE Full text] [doi: 10.1001/jama.2014.3201] [Medline: 24794371]
3. Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, et al. Innate immunity and intestinal microbiota in the development of type 1 diabetes. Nature 2008 Oct 23;455(7216):1109-1113 [FREE Full text] [doi: 10.1038/nature07336] [Medline: 18806780]
4. Wood JR, Miller KM, Maahs DM, Beck RW, DiMeglio LA, Libman IM, T1D Exchange Clinic Network. Most youth with type 1 diabetes in the T1D Exchange Clinic Registry do not meet American Diabetes Association or International Society for Pediatric and Adolescent Diabetes clinical guidelines. Diabetes Care 2013 Jul;36(7):2035-2037 [FREE Full text] [doi: 10.2337/dc12-1959] [Medline: 23340893]
5. White NH, Cleary PA, Dahms W, Goldstein D, Malone J, Tamborlane WV, Diabetes ControlComplications Trial (DCCT)/Epidemiology of Diabetes InterventionsComplications (EDIC) Research Group. Beneficial effects of intensive therapy of diabetes during adolescence: outcomes after the conclusion of the Diabetes Control and Complications Trial (DCCT). J Pediatr 2001 Dec;139(6):804-812. [doi: 10.1067/mpd.2001.118887] [Medline: 11743505]
6. Diabetes ControlComplications Trial Research Group, Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
7. Hilliard ME, De Wit M, Wasserman RM, Butler AM, Evans M, Weissberg-Benchell J, et al. Screening and support for emotional burdens of youth with type 1 diabetes: strategies for diabetes care providers. Pediatr Diabetes 2018 May;19(3):534-543 [FREE Full text] [doi: 10.1111/pedi.12575] [Medline: 28940936]
8. Wiebe DJ, Helgeson V, Berg CA. The social context of managing diabetes across the life span. Am Psychol 2016 Oct;71(7):526-538 [FREE Full text] [doi: 10.1037/a0040355] [Medline: 27690482]
9. Mulvaney SA, Hood KK, Schlundt DG, Osborn CY, Johnson KB, Rothman RL, et al. Development and initial validation of the barriers to diabetes adherence measure for adolescents. Diabetes Res Clin Pract 2011 Oct;94(1):77-83 [FREE Full text] [doi: 10.1016/j.diabres.2011.06.010] [Medline: 21737172]
10. Fitzpatrick SL, Schumann KP, Hill-Briggs F. Problem solving interventions for diabetes self-management and control: a systematic review of the literature. Diabetes Res Clin Pract 2013 May;100(2):145-161 [FREE Full text] [doi: 10.1016/j.diabres.2012.12.016] [Medline: 23312614]
11. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
12. Dunton G, Dzubur E, Li M, Huh J, Intille S, McConnell R. Momentary assessment of psychosocial stressors, context, and asthma symptoms in hispanic adolescents. Behav Modif 2016 Jan;40(1-2):257-280 [FREE Full text] [doi: 10.1177/0145445515608145] [Medline: 26438664]
13. Linas BS, Genz A, Westergaard RP, Chang LW, Bollinger RC, Latkin C, et al. Ecological momentary assessment of illicit drug use compared to biological and self-reported methods. JMIR Mhealth Uhealth 2016 Mar 15;4(1):e27 [FREE Full text] [doi: 10.2196/mhealth.4470] [Medline: 26980400]
14. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med 2016 Dec;6(4):558-565 [FREE Full text] [doi: 10.1007/s13142-016-0442-4] [Medline: 27678501]
15. Myers TC, Wonderlich SA, Crosby R, Mitchell JE, Steffen KJ, Smyth J, et al. Is multi-impulsive bulimia a distinct type of bulimia nervosa: psychopathology and EMA findings. Int J Eat Disord 2006 Dec;39(8):655-661. [doi: 10.1002/eat.20324] [Medline: 16927382]
16. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Stat Med 2012 Nov 30;31(27):3328-3336 [FREE Full text] [doi: 10.1002/sim.5338] [Medline: 22419604]
17. Bishop C. Pattern Recognition and Machine Learning: All "Just the Facts 101" Material. India: Springer; 2013.
18. 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]

19. Mulvaney SA, Vaala S, Hood KK, Lybarger C, Carroll R, Williams L, et al. Mobile momentary assessment and biobehavioral feedback for adolescents with type 1 diabetes: feasibility and engagement patterns. Diabetes Technol Ther 2018 Jul;20(7):465-474 [FREE Full text] [doi: 10.1089/dia.2018.0064] [Medline: 29882677]
20. Zhang P, Schmidt D, White J, Mulvaney S. Towards precision behavioral medicine with the internet of things (IoT): iterative design and optimization of a self-management tool for type 1 diabetes. In: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI). 2018 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); Jun 4-7, 2018; New York, USA. [doi: 10.1109/ichi.2018.00015]
21. Mulvaney SA, Vaala SE, Carroll RB, Williams LK, Lybarger CK, Schmidt DC, et al. A mobile app identifies momentary psychosocial and contextual factors related to mealtime self-management in adolescents with type 1 diabetes. J Am Med Inform Assoc 2019 Dec 01;26(12):1627-1631 [FREE Full text] [doi: 10.1093/jamia/ocz147] [Medline: 31529065]
22. iHealth COVID-19 antigen rapid test. iHealth. URL: https://ihealthlabs.com [accessed 2022-01-28]
23. Miller KM, Foster NC, Beck RW, Bergenstal RM, DuBose SN, DiMeglio LA, T1D Exchange Clinic Network. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. Diabetes Care 2015 Jun;38(6):971-978. [doi: 10.2337/dc15-0078] [Medline: 25998289]
24. Zhang P, White J, Schmidt D. Architectures and patterns for leveraging high-frequency, low-fidelity data in the healthcare domain. In: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI). 2018 Presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); Jun 4-7, 2018; New York, USA. [doi: 10.1109/ichi.2018.00104]
25. Dudani SA. The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst Man Cybern 1976 Apr;SMC-6(4):325-327. [doi: 10.1109/TSMC.1976.5408784]
26. Seber G, Lee A. Linear Regression Analysis, Second Edition. Hoboken, New Jersey, United States: Wiley; 2003.
27. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002 Dec;2/3:18-22.
28. Suykens J, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293-300. [doi: 10.1023/A:1018628609742]
29. Chawla NV, Japkowicz N, Kotcz A. Editorial: special issue on learning from imbalanced data sets. ACM SIGKDD Explorations Newsletter 2004 Jun;6(1):1-6. [doi: 10.1145/1007730.1007733]
30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
31. -. An experiment with the edited nearest-neighbor rule. IEEE Trans Syst Man Cybern 1976 Jun;SMC-6(6):448-452. [doi: 10.1109/tsmc.1976.4309523]
32. Blagus R, Lusa L. Evaluation of SMOTE for high-dimensional class-imbalanced microarray data. In: Proceedings of the 2012 11th International Conference on Machine Learning and Applications. 2012 Presented at: 2012 11th International Conference on Machine Learning and Applications; Dec 12-15, 2012; Boca Raton, FL, USA. [doi: 10.1109/icmla.2012.183]
33. Elhassan AT, Aljourf M, Al-Mohanna F, Shoukri M. Classification of Imbalance Data using Tomek Link (T-Link) combined with Random Under-sampling (RUS) as a data reduction method. Global J Technol Optim 2016;01(S1). [doi: 10.4172/2229-8711.s1111]
34. Thai-Nghe N, Schmidt-Thieme L. Learning optimal threshold on resampling data to deal with class imbalance. In: Proceedings of the IEEE RIVF International Conference on Computing and Telecommunication Technologies. 2010 Presented at: IEEE RIVF International Conference on Computing and Telecommunication Technologies; Nov 1-4, 2010; Hanoi, Vietnam.
35. Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: a review. GESTS Int Transactions Comput Sci Eng 2006;30(1):36 [FREE Full text]
36. Zeng X, Martinez TR. Distribution-balanced stratified cross-validation for accuracy estimation. J Exper Theor Artificial Intell 2000 Jan;12(1):1-12. [doi: 10.1080/095281300146272]
37. Qian SS, Stow CA, Borsuk ME. On Monte Carlo methods for Bayesian inference. Ecological Modelling 2003 Jan;159(2-3):269-277. [doi: 10.1016/s0304-3800(02)00299-5]
38. Diez Roux AV, Katz M, Crews DC, Ross D, Adler N. Social and behavioral information in electronic health records: new opportunities for medicine and public health. Am J Prev Med 2015 Dec;49(6):980-983. [doi: 10.1016/j.amepre.2015.08.027] [Medline: 26590943]
39. Whittemore R, Jaser SS, Jeon S, Liberti L, Delamater A, Murphy K, et al. An internet coping skills training program for youth with type 1 diabetes. Nursing Res 2012;61(6):395-404. [doi: 10.1097/nnr.0b013e3182690a29]
40. Kumah-Crystal YA, Hood KK, Ho Y, Lybarger CK, O'Connor BH, Rothman RL, et al. Technology use for diabetes problem solving in adolescents with type 1 diabetes: relationship to glycemic control. Diabetes Technol Ther 2015 Jul;17(7):449-454 [FREE Full text] [doi: 10.1089/dia.2014.0422] [Medline: 25826706]
41. Probst T, Pryss R, Langguth B, Schlee W. Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: results from the "TrackYourTinnitus" application. Sci Rep 2016 Feb 08;6:20382 [FREE Full text] [doi: 10.1038/srep20382] [Medline: 26853815]
42. Zhang P, Rodriguez-Cancio M, Schmidt D, White J, Dennis T. Discussions of a preliminary hand hygiene compliance monitoring application-as-a-service. In: Proceedings of the HEALTHINF. 2017 Presented at: HEALTHINF; Feb 21-23, 2017; Porto, Portugal. [doi: 10.5220/0006293705370544]

43. Zhang P, White J, Schmidt D, Dennis T. Applying machine learning methods to predict hand hygiene compliance characteristics. In: Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2017 Presented at: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); Feb 16-19, 2017; Orlando, FL, USA. [doi: 10.1109/bhi.2017.7897278]
44. Farooq M, Sazonov E. Accelerometer-based detection of food intake in free-living individuals. IEEE Sens J 2018 May 01;18(9):3752-3758 [FREE Full text] [doi: 10.1109/JSEN.2018.2813996] [Medline: 30364677]
45. Samadi S, Rashid M, Turksoy K, Feng J, Hajizadeh I, Hobbs N, et al. Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. Diabetes Technol Ther 2018 Mar;20(3):235-246 [FREE Full text] [doi: 10.1089/dia.2017.0364] [Medline: 29406789]
46. Gimpel H, Regal C, Schmidt M. myStress: unobtrusive smartphone-based stress detection. In: Proceedings of the European Conference on Information Systems. 2015 Presented at: European Conference on Information Systems; May 26-29, 2015; Münster, Germany.
47. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
48. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
49. Merwin RM, Dmitrieva NO, Honeycutt LK, Moskovich AA, Lane JD, Zucker NL, et al. Momentary predictors of insulin restriction among adults with type 1 diabetes and eating disorder symptomatology. Diabetes Care 2015 Nov;38(11):2025-2032 [FREE Full text] [doi: 10.2337/dc15-0753] [Medline: 26384389]
50. Mulvaney SA, Rothman RL, Dietrich MS, Wallston KA, Grove E, Elasy TA, et al. Using mobile phones to measure adolescent diabetes adherence. Health Psychol 2012 Jan;31(1):43-50 [FREE Full text] [doi: 10.1037/a0025543] [Medline: 21967662]
51. Merwin RM, Moskovich AA, Honeycutt LK, Lane JD, Feinglos M, Surwit RS, et al. Time of day when type 1 diabetes patients with eating disorder symptoms most commonly restrict insulin. Psychosom Med 2018;80(2):222-229. [doi: 10.1097/psy.0000000000000550]
52. Georga E, Protopappas V, Fotiadis D. Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques. In: Knowledge-Oriented Applications in Data Mining. London, United Kingdom: IntechOpen; 2011.
53. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther 2013 Oct;15(10):792-801 [FREE Full text] [doi: 10.1089/dia.2013.0104] [Medline: 23883406]
54. Ståhl F. Diabetes mellitus glucose prediction by linear and Bayesian ensemble modeling. Lund University. 2012. URL: https://portal.research.lu.se/en/publications/diabetes-mellitus-glucose-prediction-by-linear-and-bayesian-ensem [accessed 2022-01-31]
55. Bondia J, Tarín C, García-Gabin W, Esteve E, Fernández-Real JM, Ricart W, et al. Using support vector machines to detect therapeutically incorrect measurements by the MiniMed CGMS. J Diabetes Sci Technol 2008 Jul;2(4):622-629 [FREE Full text] [doi: 10.1177/193229680800200413] [Medline: 19885238]
56. Sudharsan B, Peeples M, Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. J Diabetes Sci Technol 2015 Jan;9(1):86-90 [FREE Full text] [doi: 10.1177/1932296814554260] [Medline: 25316712]
57. Biester T, Kordonouri O, Holder M, Remus K, Kieninger-Baum D, Wadien T, et al. "Let the algorithm do the work": reduction of hypoglycemia using sensor-augmented pump therapy with predictive insulin suspension (SmartGuard) in pediatric type 1 diabetes patients. Diabetes Technol Ther 2017 Mar;19(3):173-182 [FREE Full text] [doi: 10.1089/dia.2016.0349] [Medline: 28099035]
58. Bastani M. Model-free intelligent diabetes management using machine learning. University of Alberta. URL: https://era.library.ualberta.ca/items/fee1e7a7-1993-43f6-8d93-1d93855f6275 [accessed 2022-01-31]
59. Kawano Y, Yanai K. Real-time mobile food recognition system. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013 Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; Jun 23 - 28, 2013; Portland, Oregon. [doi: 10.1109/cvprw.2013.5]
60. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
61. Lansing AH, Berg CA, Butner J, Wiebe DJ. Self-control, daily negative affect, and blood glucose control in adolescents with type 1 diabetes. Health Psychol 2016 Mar 25:10.1037/hea0000325 (forthcoming) [FREE Full text] [doi: 10.1037/hea0000325] [Medline: 26914647]
62. Pugach O, Hedeker D, Richmond M, Sokolovsky A, Mermelstein R. Modeling mood variation and covariation among adolescent smokers: application of a bivariate location-scale mixed-effects model. Nicotine Tob Res 2014 May;16 Suppl 2:S151-S158 [FREE Full text] [doi: 10.1093/ntr/ntt143] [Medline: 24052502]

63. Schabert J, Browne JL, Mosely K, Speight J. Social stigma in diabetes : a framework to understand a growing problem for an increasing epidemic. Patient 2013;6(1):1-10. [doi: 10.1007/s40271-012-0001-0] [Medline: 23322536]
64. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
65. Li T, Lin G. Examining the role of location-specific associations between ambient air pollutants and adult asthma in the United States. Health Place 2014 Jan;25:26-33. [doi: 10.1016/j.healthplace.2013.10.007] [Medline: 24216026]

Abbreviations

BG: blood glucose
EMA: ecologic momentary assessment
KNN: k-nearest neighbors
LFA: learned filtering architectureRF: random forestSMBG: self-monitoring of blood glucose
T1D: type 1 diabetes

Edited by L Buis; submitted 29.06.20; peer-reviewed by S Rostam Niakan Kalhori, H Suominen, J Chen; comments to author 21.10.20; revised version received 16.07.21; accepted 15.12.21; published 03.03.22.

Please cite as:
Zhang P, Fonnesbeck C, Schmidt DC, White J, Kleinberg S, Mulvaney SA
Using Momentary Assessment and Machine Learning to Identify Barriers to Self-management in Type 1 Diabetes: Observational Study
JMIR Mhealth Uhealth 2022;10(3):e21959
URL: https://mhealth.jmir.org/2022/3/e21959
doi:10.2196/21959
PMID:35238791

Viewpoint

Leveraging Polio Geographic Information System Platforms in the African Region for Mitigating COVID-19 Contact Tracing and Surveillance Challenges: Viewpoint

Godwin Ubong Akpan[1], BEng, MSc, PhD; Isah Mohammed Bello[1], BSc, MSc; Kebba Touray[1], BSc, MSc; Reuben Ngofa[1], BSc, MSc; Daniel Rasheed Oyaole[2], MSc; Sylvester Maleghemi[3], MD, MSc; Marie Babona[1], BSc, MSc; Chanda Chikwanda[1], BSc, MPH; Alain Poy[1], BSc, MSc; Franck Mboussou[1], MD, MPH; Opeayo Ogundiran[1], MD, MPH; Benido Impouma[1], MD, MPH; Richard Mihigo[1], MD, MPH; Nda Konan Michel Yao[1], MD, MPH; Johnson Muluh Ticha[1], MD, MPH; Jude Tuma[4], MSc, PhD; Hani Farouk A Mohamed[1], BEng, MSc, DBA; Kehinde Kanmodi[5], BDS; Nonso Ephraim Ejiofor[1], MD, MBA; John Kapoi Kipterer[1], BSc, MSc; Casimir Manengu[1], MD, MPH; Francis Kasolo[1], MD, PhD; Vincent Seaman[6], PhD; Pascal Mkanda[1], MD, MPH

[1]Regional Office of Africa, World Health Organization, Brazzaville, Congo

[2]World Health Organization, Abuja, Nigeria

[3]World Health Organization, Juba, South Sudan

[4]World Health Organization, Geneva, Switzerland

[5]School of Health and Life Sciences, Teesside University, Middlesbrough, United Kingdom

[6]Bill and Melinda Gates Foundation, Seattle, WA, United States

Corresponding Author:
Godwin Ubong Akpan, BEng, MSc, PhD
Regional Office of Africa
World Health Organization
World Health Organization,Regional Office for Africa
Cite Du Djoue
Brazzaville, 500101
Congo
Phone: 242 055736476
Email: akpang@who.int

Abstract

Background: The ongoing COVID-19 pandemic in Africa is an urgent public health crisis. Estimated models projected over 150,000 deaths and 4,600,000 hospitalizations in the first year of the disease in the absence of adequate interventions. Therefore, electronic contact tracing and surveillance have critical roles in decreasing COVID-19 transmission; yet, if not conducted properly, these methods can rapidly become a bottleneck for synchronized data collection, case detection, and case management. While the continent is currently reporting relatively low COVID-19 cases, digitized contact tracing mechanisms and surveillance reporting are necessary for standardizing real-time reporting of new chains of infection in order to quickly reverse growing trends and halt the pandemic.

Objective: This paper aims to describe a COVID-19 contact tracing smartphone app that includes health facility surveillance with a real-time visualization platform. The app was developed by the AFRO (African Regional Office) GIS (geographic information system) Center, in collaboration with the World Health Organization (WHO) emergency preparedness and response team. The app was developed through the expertise and experience gained from numerous digital apps that had been developed for polio surveillance and immunization via the WHO's polio program in the African region.

Methods: We repurposed the GIS infrastructures of the polio program and the database structure that relies on mobile data collection that is built on the Open Data Kit. We harnessed the technology for visualization of real-time COVID-19 data using dynamic dashboards built on Power BI, ArcGIS Online, and Tableau. The contact tracing app was developed with the pragmatic considerations of COVID-19 peculiarities. The app underwent testing by field surveillance colleagues to meet the requirements of linking contacts to cases and monitoring chains of transmission. The health facility surveillance app was developed from the

knowledge and assessment of models of surveillance at the health facility level for other diseases of public health importance. The Integrated Supportive Supervision app was added as an appendage to the pre-existing paper-based surveillance form. These two mobile apps collected information on cases and contact tracing, alongside alert information on COVID-19 reports at the health facility level; the information was linked to visualization platforms in order to enable actionable insights.

Results: The contact tracing app and platform were piloted between April and June 2020; they were then put to use in Zimbabwe, Benin, Cameroon, Uganda, Nigeria, and South Sudan, and their use has generated some palpable successes with respect to COVID-19 surveillance. However, the COVID-19 health facility–based surveillance app has been used more extensively, as it has been used in 27 countries in the region.

Conclusions: In light of the above information, this paper was written to give an overview of the app and visualization platform development, app and platform deployment, ease of replicability, and preliminary outcome evaluation of their use in the field. From a regional perspective, integration of contact tracing and surveillance data into one platform provides the AFRO with a more accurate method of monitoring countries' efforts in their response to COVID-19, while guiding public health decisions and the assessment of risk of COVID-19.

(JMIR Mhealth Uhealth 2022;10(3):e22544) doi:10.2196/22544

KEYWORDS

contact tracing; GIS; COVID-19; surveillance

Introduction

Since the confirmation of COVID-19 in Wuhan, China, in late December 2019 [1], the disease continues to spread globally. The African region has not been spared [2]. As of July 5, 2020, the continent has recorded 466,300 cases and 11,121 deaths across all countries. Currently, the African continent, which makes up about 16.7% of the world's population, accounts for 4% of the global cases [3]. Existing evidence shows that countries that have implemented public health measures, including rapid case identification, testing, isolation, contact tracing, and quarantine of contacts, at the onset of outbreaks have suppressed the spread of COVID-19 to low thresholds, ones that do not overwhelm existing health systems. In those countries, excess mortality has been prevented as they have been able to deliver quality clinical care and minimize secondary mortality due to other causes through the continuity of essential health services [4]. With no available vaccine or therapeutics, contact tracing, social distancing, and quarantine are the only available strategies for controlling the pandemic. In Africa, where testing capacity varies greatly and is very limited in most of its member states, the importance of contact tracing in stopping further progression of COVID-19 cannot be overemphasized.

Contact tracing is a process that involves early case recognition, isolation, and tracking of people who have been exposed to a disease [5,6]. It is an essential public health tool (Textbox 1 [7-11]) for breaking human chains of transmission and has been used extensively in the control of different types of infectious disease outbreaks. Recent studies on COVID-19 surveillance (Textbox 1) have used digital contact tracing combined with other measures, such as social distancing and quarantine, to demonstrate a greater effect on the reduction of new COVID-19 cases [12]. Pertinently, Wei et al [13] and Ng et al [14] demonstrated the public health importance of contact tracing on both asymptomatic and presymptomatic persons infected with COVID-19; these are considered silent drivers of COVID-19 infection.

The implementation of an efficient contact tracing system can vary depending on the place and the type of disease. In the case of COVID-19, for instance, controlling the epidemic with a comprehensive contact tracing system guides the type of intervention to be used (ie, self-monitored quarantine for presymptomatic and asymptomatic contacts for the duration of the incubation), whereas those with severe active disease may be hospitalized. Additionally, in most scenarios, simple traditional and manual contact tracing methods can be effective at the beginning of outbreaks when the numbers are low; however, large-scale epidemics and pandemics require newer digital contact tracing methods. During highly infectious pandemics, such as the current COVID-19 pandemic where very large populations are affected, it is pertinent to use digital solutions to effectively locate contacts who have been exposed to people with the disease as well as to monitor them consistently, especially those currently on home care and self-isolation.

Numerous types of contact tracing methods have been implemented during different pandemics and endemics, and their effectiveness largely depends on the tools used [15]. Geographic information system (GIS) technology and big data analytics are major tools used in contact tracing during large outbreaks. GIS technology had been used successfully in the past to identify the rate of transmission and corresponding incidences of notorious airborne diseases [16]. Also, big data analytics have been used successfully for real-time contact tracing of disease outbreaks in livestock, as well as with other highly contagious viral respiratory diseases, such as SARS and fibromuscular dysplasia [17,18].

Over the years, scientific data collated by the World Health Organization's (WHO) emergency preparedness and response (EPR) unit showed that most countries in Africa are struggling to implement efficient contact tracing methods.This has greatly increased the number of new community-transmitted COVID-19 cases; this is a very critical situation for the continent, as it has been predicted to be the next hot spot for the coronavirus [19]. To effectively curb the problem of inefficient contact tracing in Africa, the WHO has sought the support of the African

Regional Office (AFRO) GIS to build a GIS-enabled tool for contact tracing that is pertinent to the African continent. Since 2017, the AFRO GIS Center has put in place mobile-based solutions to collect health information in near real time and, thus, has access to health program implementation and can measure the effectiveness of interventions. The tools provided by the center contributed significantly to the successes recorded with polio certification efforts [20]. In the course of proferring a workable solution, the AFRO GIS Center developed an application model (Textbox 1)—a GIS-enabled tool—for contact tracing for COVID-19 surveillance. In this paper, we present a robust and efficient contact tracing app that can be used across Africa to effectively respond to the COVID-19 outbreak. This app builds on effective tools, such as the Open Data Kit (ODK) [21], to collect and manage data in a constrained environment, combined with the AFRO Polio GIS platform (Textbox 1) [22], in order to locate, identify, monitor, and track contacts during the COVID-19 pandemic or any other large-scale pandemic. However, before proceeding to the description of the app, it will be of relevance to give an overview of a very important COVID-19 surveillance problem in Africa: the problem of contact tracing, in which the quest for solving this particular problem led to the development, deployment, and use of the app in Africa.

Textbox 1. Definition of selected terms.

Application: a program or group of programs designed for end users. Examples of an application include a word processor, a spreadsheet, an accounting application, a web browser, an email client, a media player, a file viewer, a simulator, a console game, or a photo editor [7].

Platform: a computing platform or digital platform is the environment in which a piece of software is executed. It may be the hardware, the operating system, a web browser and its associated application programming interfaces, or other underlying software, as long as the program code is executed with it [8].

Module: a section of an application or app that focuses on a unique set of deliverables or assessments [9].

Surveillance: an epidemiological practice by which the spread of disease is monitored via data collection in order to establish patterns of progression. The main role of disease surveillance is to predict, observe, and minimize the harm caused by outbreak, epidemic, and pandemic situations, as well as to increase knowledge about which factors contribute to such circumstances. A key part of modern disease surveillance is the practice of disease case reporting [10].

Tool: a programming tool or software development tool is a computer program that software developers use to create, debug, maintain, or otherwise support other programs and applications [11].

Effective Contact Tracing: The African Problem Regarding COVID-19 Surveillance

Contact tracing is a very serious limitation in disease surveillance in Africa. As shown in Figure S1 in Multimedia Appendix 1, the problems associated with contact tracing in many African countries include an inadequate number of skilled personnel who are already overstretched, suboptimal technology and tools, underfunded health systems, and poor infrastructure with associated myths, rumors, and communication barriers.

WHO AFRO–Recommended Solutions for Contact Tracing in Africa: Key Points

Overview

The WHO's AFRO GIS Center considered different scenarios during preliminary discussions in March 2020, with the COVID-19 incident management team managing the regional response at WHO's regional office. During these preliminary discussions, it was highlighted that the best contact tracing solution would leverage existing GIS platforms and would be deployed to fill the current gaps known in COVID-19 surveillance and to help address existing challenges with contact tracing (Figure S1 in Multimedia Appendix 1).

It was noted that while traditional contact tracing could still work in places with few contacts, these approaches would be constrained in countries with many contacts, coupled with the impact of lockdowns on contact tracing [23]. Consequently, the technical team that was tasked to review different possible solutions to the problems associated with contact tracing in Africa recommended the following:

1. The rapid development of a GIS-enabled COVID-19 self-reporting contact tracing app.
2. The rapid deployment of a contact registration and follow-up app by COVID-19 surveillance teams in-country.
3. Visualization of field surveillance and reporting with interactive and near real-time dashboards.
4. COVID-19 surveillance at the health facility level to assess the preparedness and readiness of health systems to cope with COVID-19 at this reporting level.

Following the review, Benin and Zimbabwe were selected based on their indication of interest to immediately use some or all aspects of the platform.

The Rapid Development of a GIS-Enabled COVID-19 Self-reporting Contact Tracing App

Overview

At the very early stage of the COVID-19 outbreak in Africa, the AFRO GIS Center leveraged the success of its contribution in the eradication of polio in the African region. Coupled with its surveillance experience over the years [24] and in collaboration with the WHO EPR team, the AFRO GIS Center was able to develop some novel tools for immediate field deployment to collect real-time data and monitor the COVID-19 pandemic in the field by surveillance personnel. Different technologies were used to develop and integrate this rapid intervention (ie, the COVID-19 app used by surveillance personnel); these included the following: ODK technology [25]

and stacks via Ona, which is an open-source tool. These tools are housed and secured inside the WHO infrastructure, primarily for real-time data collection in the field; these include the use of external tools like ArcGIS Online [26], Power BI [27], application programming interfaces (APIs), KoBoToolbox [28], and DHIS2 (District Health Information Software 2) [29] for contact tracing and visualization processes. ArcGIS Online and Power BI are data visualization platforms that have different use cases for real-time data visualization.

Shortly after the development and deployment of the above-described application for COVID-19 surveillance personnel, the AFRO GIS Center went ahead to develop a GIS-enabled COVID-19 self-reporting contact tracing app for registering and following up contacts by the surveillance teams in-country, in accordance with the recommendations of the COVID-19 incident management team of the WHO AFRO. The developed app enables contacts in home-based care, self-isolation, and in quarantine centers to provide daily updates on their health condition. In addition, the app provides an opportunity for these contacts to be able to identify the nearest health facility to which they can quickly report if they develop any symptoms of the disease. Importantly, the traditional in-country contact tracing teams are also able to register and follow up contacts and cases with the app or any other app that the country may opt for. If the country opts for other apps, the solution allows for interoperability of the toolbox, which enables aggregating, analyzing, and visualizing of the data in the same regional dashboards.

Figure S2 in Multimedia Appendix 1 depicts the architecture of the COVID-19 self-reporting contact tracing app developed by the AFRO GIS Center, in collaboration with the WHO EPR team. The app has the following three components: (1) a data collection component, (2) an API component, and (3) a polio GIS toolbox and platform.

Data Collection Component

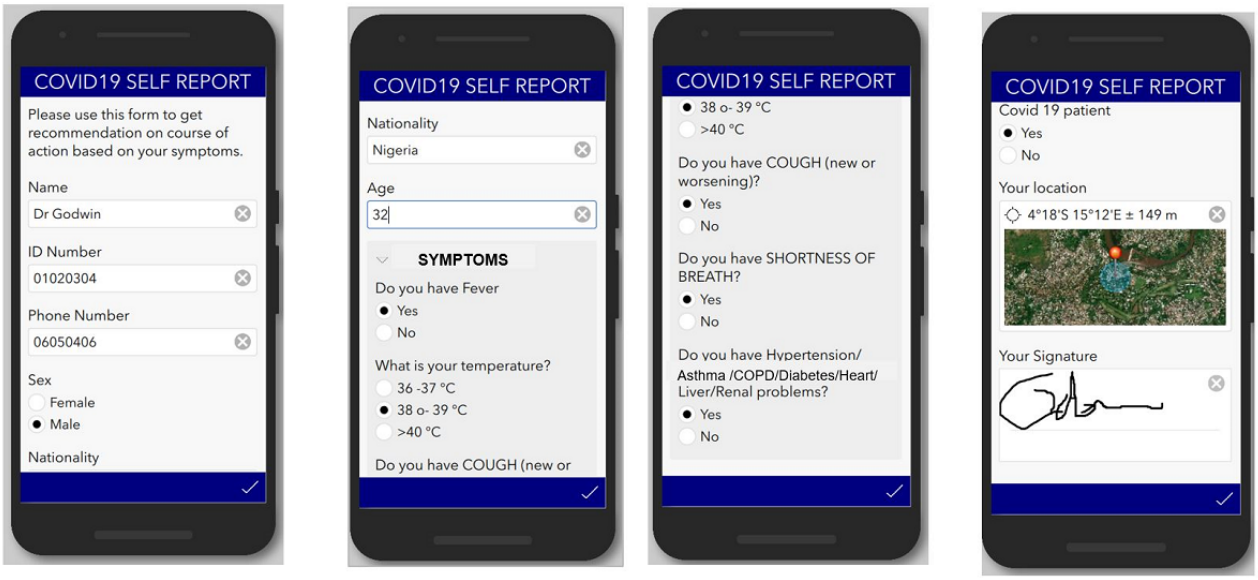
Overview

The architecture was structured to accommodate a wide range of data collection tools, giving countries the liberty to use the WHO data collection tool or any data collection tool that best suits them, such as KoBoToolbox and DHIS2. These data collection tools are then imported in real time into the polio GIS platform for analysis and visualization. These tools are summarized in the following sections.

COVID-19 Self-report Form

This form was structured to collect details and the daily status of a case or contact under quarantine or being followed up for the mandatory 14-day period. It enables the individual to self-report his or her daily status and submit to the central servers for monitoring and feedback (Figure 1).

Figure 1. An example of a self-report form for use with widespread contacts.



COVID-19 Case Investigation Form

This form was structured to collect the details and profiles of confirmed, suspected, and probable cases. The form can also collect the medical provider's information, patient information, clinical information, travel history, and final classification of the case. It is important to note that this form is linked to the contact listing form; as such, the case phone number or a concatenation of the country, province, district, and case number is used as a unique identifier for referencing and following up. This form is used by health personnel at the health facility level once a COVID-19 case is suspected.

Contact Listing and Follow-up Form

This form is divided into two components: the first component is used to collect the contacts of a confirmed case (ie, contact tracing form), and the second component is used to collect and record the temperature reading of a case contact for 14 days (ie, contact follow-up and self-reporting). This form references the actual case for those contacts that could be linked to a COVID-19 case.

Contact Registration Form

Every contact connected to a case is expected to be registered in the COVID-19 database. A prompt is shown at the beginning

of the form that allows for contact registration. The index case's ID number is required at this stage to link the contact to an already-existing case in the database (ie, a new contact cannot be recorded without a prerecorded index case in the database). This displays the records of the index case for confirmation and validation before registering the details of the new contact.

Contact Follow-up Form

After contact registration, the contacts are expected to have a 14-day record update of their temperature and medical condition. There is a prompt at the beginning of the form that allows for contact follow-up (ie, it prompts self-isolated contacts to submit daily results for 14 days).

Traveler Health Questionnaire and Follow-up Form

This form is divided into two components: the first component is used to collect general information about a traveler entering into the country, and the second component is used to collect and record the temperature reading of the traveler for 14 days (ie, contact follow-up). Note that the follow-up component is linked to the traveler health questionnaire and is referenced using the case phone number of the traveler as the unique identifier.

API Component

This component is responsible for the interoperability and interaction of data between platforms. The APIs link the various platforms together, facilitating data exchange in real time. This allows for seamless operations and data exchange between the data collection, hosting, analytics, and visualization tools within the polio GIS toolbox and platform.

Polio GIS Toolbox and Platform

This comprises all the resources developed and deployed within the WHO AFRO infrastructure. It comprises front-end data collection tools on mobile devices and web interfaces, database servers, data visualization tools, and platforms.

Rapid Deployment of the Self-reporting Contact Tracing App

The AFRO GIS Center, in collaboration with the WHO EPR team, went ahead to deploy the COVID-19 Emergency Deployment Toolkit, using the ODK platform to support contact tracing through self-reporting.

Contact Tracing Data Management

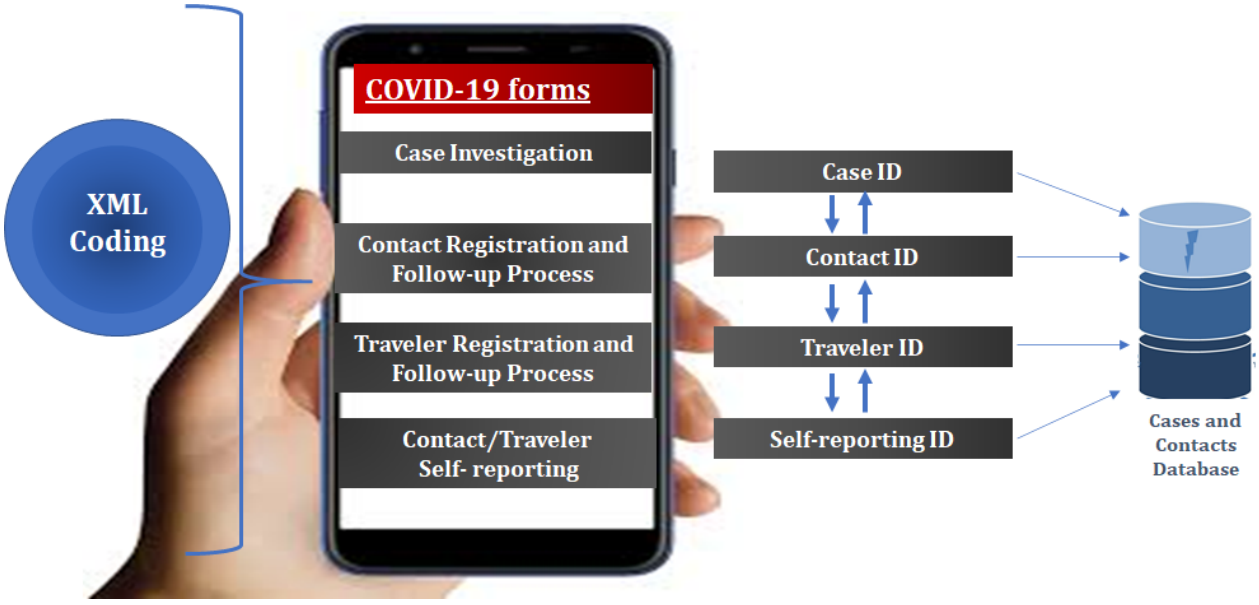
Overview

Data management is a very crucial aspect of surveillance, particularly when it comes to data of relevance to contact tracing. In the management of contact tracing data, database linkages, performance monitoring, visualization, and analysis are important factors to consider. How these factors were managed are discussed below.

Database Linkages

The biggest challenge of contact tracing data management is enrolling contacts and performance monitoring of the contact tracing processes [23]. This was overcome by ensuring that each case ID was matched to the contact ID and that the traveler ID was listed in the contact database during the quarantine period for ease of tracking if the traveler were to become a case. This section describes the linkage system we used to connect the identity management system of the cases to contacts and travelers, while maintaining confidentiality and integrity of the cases and our contacts database (Figure 2).

Figure 2. African Regional Office geographic information system architecture showing forms.



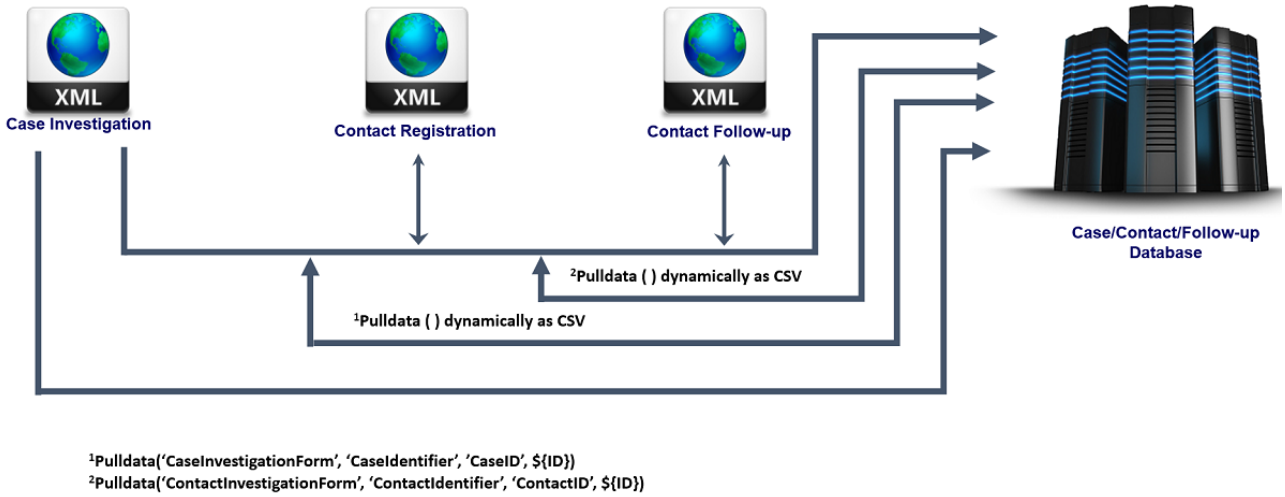
Linking the Dynamic Case and Contact Data via XML Form to Populate the Case and Contact Database

This process hinges around specifying form data as a media file for another form. Used in conjunction with the *pulldata* function from the ODK, this allows the developer to pull data from other dynamic data sets and surveys (ie, other forms in the Ona ODK system that are still active and accepting submissions) in the

same or different project, similar to pulling data from a preloaded CSV file (Figure 3).

On clicking the button "Link Dataset," the changes are saved automatically. The linked data set appears in forms on the Android device. After data set linkage to the form, data can be effectively pulled from the linked data set into the form using the *pulldata* function under the calculation column. The file name of the linked data set entered above will be the file name referenced in the *pulldata* function (ie, locations). See Figure 3 for an illustration of the *pulldata* function and considerations for replicating this in another situation that warrants the use of dynamically pulled data from one mobile form through a server to another mobile form.

Figure 3. Linking dynamic data via XML files using pull functions.



Performance Monitoring

Evidence-based performance monitoring of contact tracing, surveillance activities, and other health interventions by health workers was a key consideration in the repurposing of polio GIS platforms. A geotracking feature was added to our app to provide geographic evidence of a team's visits to specific contacts. This also ensures that health workers conduct surveillance and other health interventions with the knowledge that their activities are being monitored for accountability.

Visualization and Analysis

A one-stop dashboard that brings data from participating countries into a regional interface for COVID-19 activities was developed using connectors from the databases to ArcGIS and Power BI to develop the key performance indicators for the pillars of the COVID-19 response. The dashboard allows the incident management team to conduct an integrated and effective response to the pandemic. An example is seen in Figure S3 in Multimedia Appendix 1.

Health Facility–Based COVID-19 Surveillance

A module (Textbox 1) for reporting COVID-19 activities, infection prevention and control (IPC) readiness assessment, and recording of cases that fit the COVID-19 clinical definitions in facility-based registers was developed and deployed for field use by health workers that conduct surveillance at the facility level. This module leveraged an existing intervention called the Integrated Supportive Supervision (ISS) app. Thus, it was easy to deploy, as the development was appended on top of the existing app, which is in use by over 5000 health workers across the region with supportive supervision visits that average 150,000 annual visits across the African region.

The module focused on the use of the ODK-based app (ie, the ISS app) with the following COVID-19 assessment focuses:

1. Awareness of the existence of a COVID-19 surveillance system.
2. Display of COVID-19 case definition posters.
3. Presence of a COVID-19 surveillance focal point.
4. Number of health workers in the facility who know the definition of a suspect case.
5. Number of people who know the COVID-19 alerting number and notification channel.
6. Appraisal of IPC awareness in terms of handwashing, functioning of handwashing kits, availability of isolation room, use of personal protective equipment, and availability of a triage system in the health facility.
7. Recording of details of suspected cases that fit the clinical system of COVID-19 and facilitating testing.

Field Use and Outcome Evaluation

Contact Tracing Use

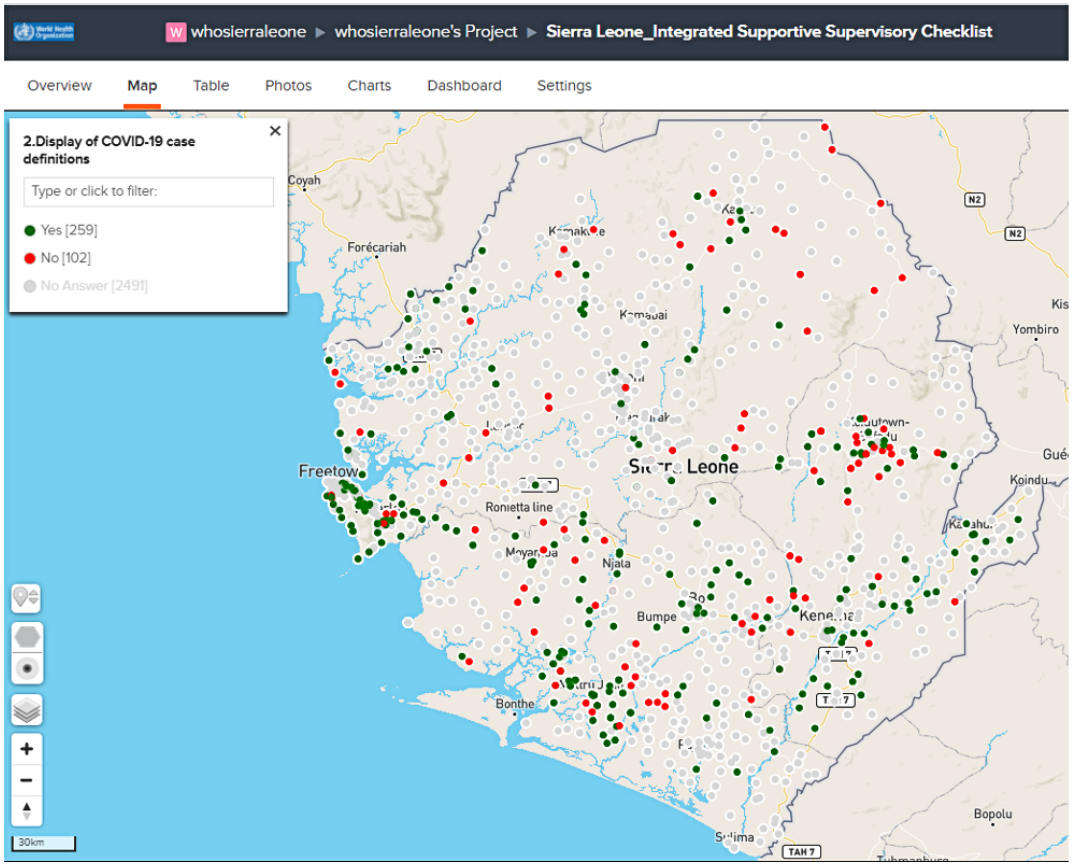
The adoption of the solutions developed by the AFRO GIS Center in collaboration with the WHO EPR team was piloted in Zimbabwe and Benin in April 2020; the solutions were then subsequently used in Nigeria, Uganda, Cameroon, and South Sudan to conduct contact tracing or support other contact tracing data collection expeditions. See Figure S4 in Multimedia Appendix 1 for the status of deployments. Since most countries in the region (98%) are familiar with the deployment of the ODK tool from its use in polio eradication, the COVID-19 contact tracing app is quite easy for countries to deploy and use without extensive training. In fact, several additional countries have demonstrated interest following a conference call with the AFRO GIS Center, in which 15 countries participated. Countries

already using similar tools for contact tracing were encouraged to share their APIs with AFRO to enable the GIS Center to pull data into the regional platform. However, countries with a high number of contacts have been encouraged to use the self-reporting module for daily contact self-reporting.

Health Facility–Based COVID-19 Surveillance

A total of 27 countries in the region adopted this module, which was developed to mitigate the gaps of COVID-19 surveillance at the health facility level. Event-based surveillance is highly optimized in low- and middle-income countries at the health facility level [30,31] and could also be key to improved COVID-19 surveillance. This module was easy to develop and deploy, as there was no training necessary since other priority disease surveillance was already ongoing at the health facility level using the ODK-based ISS forms. Real-time interactive visualization of the data from health facility–based surveillance is available at the country level for all the key variables. Also, geographic representation in maps for important accessories for IPC and triaging by the health facility is accessible to decision makers at all levels in-country. Figure 4 shows an example of a map of a health facility with COVID-19 posters displayed with definitions, which is an entry point for adequate sensitization at that level of reporting.

Figure 4. Sample status of COVID-19 sensitization posters and displays from the health facility–based surveillance module for Sierra Leone.



Discussion

Strengths of the Polio GIS Platform for Contact Tracing and Health Facility–Based Surveillance

The scope of leveraging the polio GIS platform in terms of apps and visualization tools addresses the mitigation plans for contact tracing and surveillance gaps in the region regarding the following aspects:

1. During the COVID-19 pandemic, the platform was able to support the identification of COVID-19 cases, contacts, and database consolidation gaps (eg, matching contacts to their index cases), helping to guide the response.
2. During outbreaks and routine contexts alike, the platform was able to identify community transmissions that were not detected by the traditional contact and case databases.
3. Health facility–based surveillance was built on an already-existing ISS module and was, thus, quite sustainable, as other active disease surveillance was already being conducted with the app.
4. The chain of transmission and surveillance gaps were easily seen on the interactive visualization made possible by the real-time connection of data being submitted from the apps into Power BI; thus, faster decision-making was possible at all levels.
5. The platform allowed ease of use of the data entry modules; there was little or no training, as surveillance teams at the country level were already used to similar technologies as a result of using mobile phones for other interventions.
6. The platform was developed with interoperability being given the highest consideration so that other countries' contact tracing efforts and facility-based surveillance systems, such as DHIS2, KoBoTool, and SORMAS (Surveillance Outbreak Response Management and Analysis System), could be easily connected to share data.

This platform offers an additional dimension to the existing platforms that are being used in supporting the COVID-19 response [32,33], with features that make it more adaptive to the local context. In the African region, 93% of the countries are already using this platform for other health projects. The AFRO GIS Center has been leveraging GIS technology to ensure equitable access to essential health services, ranging from ISS for the Expanded Programme on Immunization, support for microplanning, effective coordination during the Ebola response, and monitoring the cholera outbreak response, among others [34,35], thereby making it easy to adapt and implement. The tool has been shown to support polio vaccination activities in complex humanitarian settings, including within refugee camps and camps for internally displaced people [36].

The use of the GIS platform has been shown to provide adequate support and promote care of individuals across different thematic areas of health. Documented evidence exists that has demonstrated its capability to identify high-priority areas that require maternal care [37]. In addition, it has helped to identify health trends, including tracking the spread of infectious diseases, such as Ebola and measles. The use of the GIS platform has been shown to help identify the problem, identify where it exists, and equally support the provision and maintenance of care for individuals through a more efficient and coherent manner.

Trade-offs for Leveraging the Polio GIS Platform

Many questions with regard to information sensitivity and privacy regarding COVID-19 cases and contacts have been answered by the security features of the app and the platform. However, there still exists the possibility of intrusion, careless handling of passwords, and hacking that may lead to compromised information if countries do not adequately manage their access and control. Also, the app and platform deployment involves users having phones for data entry and visualization; thus, a lot of phones are required for implementation of all aspects of surveillance and contact tracing.

How to Implement the Platform

The solutions outlined can easily be implemented with the procedural steps that were enumerated and explained in the data collection and visualization sections. The use of technology stacks via ODK and Power BI to collect data and to visualize them in real time can easily be replicated by technical users who are familiar with these tools. However, it is important to note that early engagement and meetings with all stakeholders was crucial to ensure ownership, increase coordination, and gain a better understanding of the existing surveillance landscape [31]. Good linkage to the response was described as essential for all systems, as every verified case and contact is documented. In order to ensure timely reporting of cases found via health facility–based reporting, health workers at the facility level required training and supportive supervision by trained district-level teams.

Conclusions

Health facility–based surveillance and effective contact tracing management tools outlined here will provide valuable information that can strengthen the use of data by national surveillance systems during the pandemic. Evidence from the global pandemic thus far clearly presents three challenges in controlling COVID-19: its lasting pandemic potential, high fatality due to its infectivity, and its ability to disrupt health systems. Similarly, in the absence of vaccines and therapeutics, the only available tools for control include contact tracing, social distancing, and quarantine. The African region must, therefore, adopt a variety of methods to minimize the above challenges and adapt the response to the specific needs of each country as the outbreak evolves.

Digital contact tracing and surveillance at the facility level will be paramount at some point for every country to stay a step ahead of the virus. The applications and apps that have been developed for COVID-19 surveillance and contact tracing are not standalone interventions but should be implemented together with social distancing measures and quarantine, depending on the size of the outbreak.

In the African region, COVID-19 testing capacity varies widely. Countries with limited testing capacities and large outbreaks will need more advanced comprehensive contact tracing solutions, such as the one described herein, to suppress the virus to lower transmission rates. However, those with smaller outbreaks can use traditional methods with ODK or KoBoCollect platforms and simply share their APIs with the regional office to ensure that all data are available to the incident management system.

As countries begin to relax public health lockdowns, traditional and advanced contact tracing methods will be necessary to highlight areas of ongoing transmission; these data will be needed not only in their respective countries but also on a regional level to enable a better understanding of the pandemic and optimal decision-making in managing the risks and responses.

If widely adopted in the region, this innovation, alongside existing data collection tools (eg, KoBoCollect and DHIS2), will help countries respond effectively and efficiently to the pandemic. Other benefits to countries using this tool include real-time monitoring of the epidemic and their response, timeliness and completeness of contact tracing, and staff accountability. These factors are the cornerstone of surveillance during epidemics. During the Ebola outbreak in West Africa from 2014 to 2016, it was well-documented that a major contributor fueling the epidemic was the lack of standardized and synchronized contact tracing. Once adopted and implemented, a significant decrease in cases was observed across the affected countries, even though, at the time, the solutions were not as advanced as they are today.

With the high penetration rate of mobile phones across the African region, mobile-based monitoring of COVID-19, from traditional methods to voluntary self-reporting and remote follow-up of contacts, will greatly improve the identification of suspected cases and contacts; these are important resources to help in the region's fight against this debilitating disease. Additionally, the use of this tool should reduce the burden on health systems, allowing for the provision of essential health services and minimizing mortalities from COVID-19 and

neglected secondary diseases, which can result from a system overwhelmed by the pandemic.

From a regional perspective, integration of contact tracing and surveillance data into one platform provides the AFRO with a more accurate method of monitoring country efforts in their response to COVID-19, while guiding public health decisions and the assessment of risk for COVID-19.

Authors' Contributions

GUA, IMB, and KT conceived and designed the study process. GUA, IMB, KT, RN, DRO, and MB participated in platform implementation and iterations. SM, CC, AP, JT, FM, OO, JKK, NEE, HFAM, JMT, and BI contributed to the data quality control process and editing of the manuscript. KK, NKMY, CM, RM, VS, FK, and PM provided oversight functions, managed the deployments, and provided project guidance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Leveraging Polio GIS platforms in the African Region for mitigating Covid-19 contact tracing and Surveillance challenges.
[DOCX File , 587 KB - mhealth_v10i3e22544_app1.docx]

References

1. Coronavirus Disease (COVID-19): Situation Report – 133. Geneva, Switzerland: World Health Organization; 2020 Jun 01. URL: https://apps.who.int/iris/bitstream/handle/10665/332281/nCoVsitrep01Jun2020-eng.pdf?sequence=1&isAllowed=y [accessed 2022-02-14]
2. Coronavirus cases in Africa 2020. Worldometer. 2021 Mar 14. URL: https://www.worldometers.info/coronavirus/ [accessed 2021-03-14]
3. Novel Coronavirus (2019-nCoV): Situation Report - 1. Geneva, Switzerland: World Health Organization; 2020 Jan 21. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf [accessed 2020-01-21]
4. WHO Coronavirus (COVID-19) Dashboard. Geneva, Switzerland: World Health Organization URL: https://covid19.who.int/ [accessed 2021-03-21]
5. Cabore JW, Karamagi HC, Kipruto H, Asamani JA, Droti B, Seydi ABW, et al. The potential effects of widespread community transmission of SARS-CoV-2 infection in the World Health Organization African Region: A predictive model. BMJ Glob Health 2020 May;5(5):e002647 [FREE Full text] [doi: 10.1136/bmjgh-2020-002647] [Medline: 32451366]
6. Klinkenberg D, Fraser C, Heesterbeek H. The effectiveness of contact tracing in emerging epidemics. PLoS One 2006 Dec 20;1:e12 [FREE Full text] [doi: 10.1371/journal.pone.0000012] [Medline: 17183638]
7. Application software. Wikipedia. URL: https://en.wikipedia.org/wiki/Application_software [accessed 2021-02-21]
8. Computing platform. Wikipedia. URL: https://en.wikipedia.org/wiki/Computing_platform [accessed 2021-02-21]
9. Modules and packages. learnpython.org. URL: https://www.learnpython.org/en/Modules_and_Packages [accessed 2021-02-21]
10. Diseases surveillance. The Task Force for Global Health. URL: https://taskforce.org/disease-surveillance/ [accessed 2021-02-21]
11. Programming tool. Wikipedia. URL: https://en.wikipedia.org/wiki/Programming_tool [accessed 2021-02-21]
12. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
13. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 10;69(14):411-415 [FREE Full text] [doi: 10.15585/mmwr.mm6914e1] [Medline: 32271722]
14. Ng Y, Li Z, Chua Y, Chaw W, Zhao Z, Er B, et al. Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore - January 2-February 29, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 20;69(11):307-311 [FREE Full text] [doi: 10.15585/mmwr.mm6911e1] [Medline: 32191691]
15. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 among 627,386 persons in contact with the Diamond Princess Cruise Ship passengers who disembarked in Taiwan: Big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
16. Moonan PK, Bayona M, Quitugua TN, Oppong J, Dunbar D, Jost KC, et al. Using GIS technology to identify areas of tuberculosis transmission and incidence. Int J Health Geogr 2004 Oct 13;3(1):1 [FREE Full text] [doi: 10.1186/1476-072x-3-23]

17. Nöremark M, Widgren S. EpiContactTrace: an R-package for contact tracing during livestock disease outbreaks and for risk-based surveillance. BMC Vet Res 2014 Mar 17;10:71 [FREE Full text] [doi: 10.1186/1746-6148-10-71] [Medline: 24636731]
18. Zhang Y, Dang Y, Chen YD, Chen H, Thurmond M, King CC, et al. BioPortal Infectious Disease Informatics research: Disease surveillance and situational awareness. In: Proceedings of the 9th International Conference on Digital Government Research. 2008 Presented at: 9th International Conference on Digital Government Research; May 18-21, 2008; Montreal, QC p. 393-394.
19. Coronavirus: Africa could be next epicentre, WHO warns. BBC News. 2020 Apr 17. URL: https://www.bbc.com/news/world-africa-52323375 [accessed 2021-02-21]
20. Touray K, Mkanda P, Tegegn SG, Nsubuga P, Erbeto TB, Banda R, et al. Tracking vaccination teams during polio campaigns in northern Nigeria by use of geographic information system technology: 2013-2015. J Infect Dis 2016 May 01;213 Suppl 3:S67-S72 [FREE Full text] [doi: 10.1093/infdis/jiv493] [Medline: 26609004]
21. Hartung C, Lerer A, Anokwa Y, Tseng C, Brunette W, Borriello G. Open Data Kit: Tools to build information services for developing regions. In: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development. 2010 Presented at: 4th ACM/IEEE International Conference on Information and Communication Technologies and Development; December 13-16, 2010; London, UK p. 1-12. [doi: 10.1145/2369220.2369236]
22. Chala K. WHO commits to improve health situation in Gambella, Ethiopia. World Health Organization. 2015 Nov 05. URL: https://www.afro.who.int/news/who-commits-improve-health-situation-gambella-ethiopia [accessed 2022-02-24]
23. Rimpiläinen S, Thomas J, Morrison C. Global Examples of COVID-19 Surveillance Technologies. Flash Report. Glasgow, Scotland: Digital Health & Care Institute, University of Strathclyde; 2020 Apr 16. URL: https://strathprints.strath.ac.uk/72028/1/Rimpilainen_etal_DHI_2020_Global_examples_of_Covid_19_surveillance_technologies.pdf [accessed 2022-02-24]
24. The Transformation Agenda Series 4: Sustained Progress Towards Polio Eradication in the WHO African Region. Brazzaville, Republic of the Congo: WHO Regional Office for Africa; 2019 May. URL: https://apps.who.int/iris/bitstream/handle/10665/326840/9789290234289-eng.pdf?sequence=1&isAllowed=y [accessed 2022-02-24]
25. Brunette W, Sundt M, Dell N, Chaudhri R, Breit N, Borriello G. Open Data Kit 2.0: Expanding and refining information services for developing regions. In: Proceedings of the 14th Workshop on Mobile Computing Systems and Applications. 2013 Presented at: 14th Workshop on Mobile Computing Systems and Applications; February 26-27, 2013; Jekyll Island, GA p. 1-6. [doi: 10.1145/2444776.2444790]
26. ArcGIS Online. URL: https://www.arcgis.com/index.html [accessed 2021-02-21]
27. Microsoft Power BI. URL: https://powerbi.microsoft.com/en-us/ [accessed 2021-02-21]
28. KoBoToolbox. URL: https://www.kobotoolbox.org/ [accessed 2021-02-21]
29. DHIS2. URL: https://dhis2.org/ [accessed 2021-02-21]
30. Obermair HM, Dodd RH, Bonner C, Jansen J, McCaffery K. 'It has saved thousands of lives, so why change it?' Content analysis of objections to cervical screening programme changes in Australia. BMJ Open 2018 Feb 13;8(2):e019171 [FREE Full text] [doi: 10.1136/bmjopen-2017-019171] [Medline: 29440214]
31. Clara A, Do TT, Dao AT, Tran PD, Dang TQ, Tran QD, et al. Event-based surveillance at community and healthcare facilities, Vietnam, 2016-2017. Emerg Infect Dis 2018 Sep;24(9):1649-1658 [FREE Full text] [doi: 10.3201/eid2409.171851] [Medline: 30124198]
32. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ 2020 Oct 15;739:140033 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.140033] [Medline: 32534320]
33. Sarfo AK, Karuppannan S. Application of geospatial technologies in the COVID-19 fight of Ghana. Trans Indian Natl Acad Eng 2020 Jul 04;5(2):193-204. [doi: 10.1007/s41403-020-00145-3]
34. WHO in Africa offers small research grants on the innovative use of GIS towards polio eradication. World Health Organization. 2019 Oct 11. URL: https://www.afro.who.int/fr/node/11856 [accessed 2021-02-21]
35. World Health Organization GIS Centre for Health. AFRO GIS: Reaching all populations. ArcGIS StoryMaps. 2021 Jan 06. URL: https://storymaps.arcgis.com/stories/3a26d81380234693a8ed2c0c38342ac5 [accessed 2021-02-21]
36. The top five tech solutions that helped eradicate wild poliovirus in the African Region. Africa Kicks Out Wild Polio. URL: https://www.africakicksoutwildpolio.com/the-top-five-tech-solutions-that-helped-eradicate-wild-poliovirus-in-the-african-region/ [accessed 2021-02-21]
37. Salehi F, Ahmadian L. The application of geographic information systems (GIS) in identifying the priority areas for maternal care and services. BMC Health Serv Res 2017 Jul 12;17(1):482 [FREE Full text] [doi: 10.1186/s12913-017-2423-9] [Medline: 28701226]

Abbreviations

AFRO: African Regional Office
API: application programming interface
DHIS2: District Health Information Software 2

EPR: emergency preparedness and response
GIS: geographic information system
IPC: infection prevention and control
ISS: Integrated Supportive Supervision
ODK: Open Data Kit
SORMAS: Surveillance Outbreak Response Management and Analysis System
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by L Jordan, S Lalmuanawma; comments to author 06.10.20; revised version received 01.02.21; accepted 08.05.21; published 17.03.22.

Please cite as:
Akpan GU, Bello IM, Touray K, Ngofa R, Oyaole DR, Maleghemi S, Babona M, Chikwanda C, Poy A, Mboussou F, Ogundiran O, Impouma B, Mihigo R, Yao NKM, Ticha JM, Tuma J, A Mohamed HF, Kanmodi K, Ejiofor NE, Kipterer JK, Manengu C, Kasolo F, Seaman V, Mkanda P
Leveraging Polio Geographic Information System Platforms in the African Region for Mitigating COVID-19 Contact Tracing and Surveillance Challenges: Viewpoint
JMIR Mhealth Uhealth 2022;10(3):e22544
URL: https://mhealth.jmir.org/2022/3/e22544
doi:10.2196/22544
PMID:34854813

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org