# JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 10 (2022), Issue 2 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

## Contents

## Corrigenda and Addenda

Viewpoint

# Standardized Integration of Person-Generated Data Into Routine Clinical Care

Billy Zeng[1], BSc; Riley Bove[2], MMSc, MD; Simona Carini[1], MA; Jonathan Shing-Jih Lee[1], MAS, MD; JP Pollak[3], PhD; Erica Schleimer[2], BSc; Ida Sim[1], MD, PhD

[1]Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]University of California, San Francisco Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, United States

[3]The Commons Project, New York, NY, United States

**Corresponding Author:**
Ida Sim, MD, PhD
Division of General Internal Medicine
University of California, San Francisco
Suite 308
1545 Divisadero St
San Francisco, CA, 94143-0320
United States
Phone: 1 415 514 8686
Fax: 1 415 514 8666
Email: ida.sim@ucsf.edu

## Abstract

Person-generated data (PGD) are a valuable source of information on a person's health state in daily life and in between clinic visits. To fully extract value from PGD, health care organizations must be able to smoothly integrate data from PGD devices into routine clinical workflows. Ideally, to enhance efficiency and flexibility, such integrations should follow reusable processes that can easily be replicated for multiple devices and data types. Instead, current PGD integrations tend to be one-off efforts entailing high costs to build and maintain custom connections with each device and their proprietary data formats. This viewpoint paper formulates the integration of PGD into clinical systems and workflow as a *PGD integration pipeline* and reviews the functional components of such a pipeline. A PGD integration pipeline includes PGD acquisition, aggregation, and consumption. Acquisition is the person-facing component that includes both technical (eg, sensors, smartphone apps) and policy components (eg, informed consent). Aggregation pools, standardizes, and structures data into formats that can be used in health care settings such as within electronic health record–based workflows. PGD consumption is wide-ranging, by different solutions in different care settings (inpatient, outpatient, consumer health) for different types of users (clinicians, patients). The adoption of data and metadata standards, such as those from IEEE and Open mHealth, would facilitate aggregation and enable broader consumption. We illustrate the benefits of a standards-based integration pipeline for the illustrative use case of home blood pressure monitoring. A standards-based PGD integration pipeline can flexibly streamline the clinical use of PGD while accommodating the complexity, scale, and rapid evolution of today's health care systems.

***(JMIR Mhealth Uhealth 2022;10(2):e31048)*** doi:10.2196/31048

**KEYWORDS**

mobile health; data sharing; health care; patient-generated health data; telemedicine

## Introduction

Person-generated data (PGD) are a valuable source of information on a person's health state in daily life and in between clinic visits [1]. PGD can be acquired via apps, sensors, wearables, or simple online forms, which we will collectively call *PGD devices*.

To fully extract value from PGD, health care organizations must be able to smoothly integrate data from PGD devices into routine clinical workflows. For example, in an ideal remote blood pressure (BP) monitoring program, clinicians will "prescribe" a BP monitoring plan (eg, measure BP every morning for the next 2 weeks). The patient will collect and share BP data from their Bluetooth-connected wireless cuff, data that will be

seamlessly integrated into the clinical workflow for clinicians to see during patient management, for example, titration of home medications based on notifications of outlier home BP values. This same workflow should be able to accommodate prescriptions of other PGD such as blood glucose, body weight, or oxygen saturation acquired by any clinically approved PGD device.

Current telemonitoring programs, however, often have a limited scope, address only 1 disease (eg, hypertension, diabetes, or heart failure), acquire only 1 type of remote data [2,3] (eg, BP or blood glucose), and support only a limited number of PGD devices (in terms of brand/model). This restrictiveness is at odds with the current technical capabilities of internet services in which data can be exchanged with a device-agnostic approach [4,5]. Email is a familiar example. Underlying standards permit email to be sent and read regardless of service provider, app, browser, or device used [6]. The current state of PGD-to-clinic integration lacks the seamlessness of email. Instead, health care organizations build and maintain custom connections with each device and their proprietary data formats. Such connections account for a large share of the cost of using PGD devices in clinical care [7], which constitutes a barrier to PGD usage [8].

This viewpoint formulates the integration of PGD into clinical systems and workflow as a *PGD integration pipeline* and reviews the functional components of such pipeline. We contrast the current state of integration to a standards-based pipeline using an example of integrating wireless BP data into primary care. We emphasize throughout the central importance of data standards in facilitating device-agnostic approaches needed to accommodate the complexity, scale, and rapid evolution of today's health care systems.

## Standardized PGD Integration Pipeline

### Overview

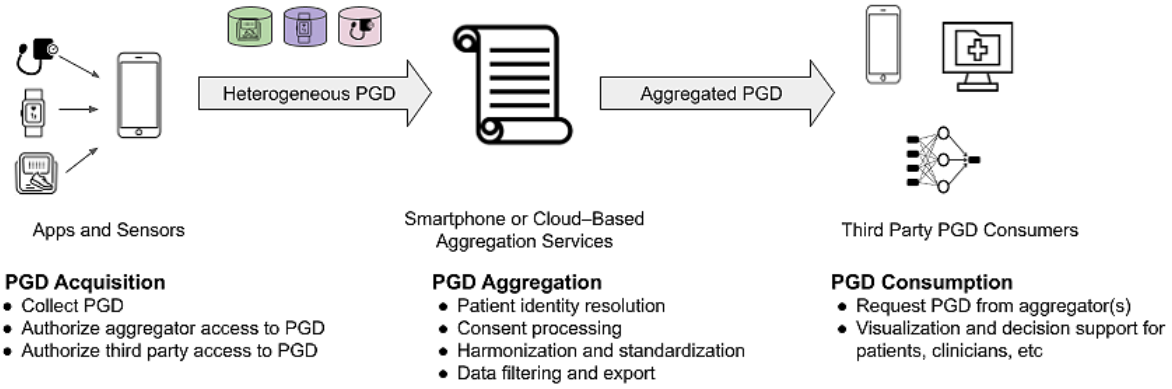
Building custom connections between individual devices and health care organizations is costly and introduces data management inefficiencies. For organizations interested in remotely monitoring multiple types of health data via different PGD devices, one approach is to select 1 or a few device vendors for each data type and develop custom connections for each device to the electronic health record (EHR) [9]. Not only is this approach redundant, costly, and maintenance heavy, the dependence on vendor- or device-specific custom connections reduces flexibility to add or substitute new devices in the future.

We can identify opportunities for streamlining the pipeline if we segment the 3 major functional components of PGD integration:

1. PGD acquisition: this encompasses PGD devices that manage person-facing functions such as consent and data collection;
2. PGD aggregation: this service manages consent, authentication, and authorization; maps data to standardized format(s); provides storage; and a query endpoint for third parties;
3. PGD consumption: third parties including EHRs, decision support systems, and analytic services provide applications that consume PGD to serve users such as clinicians and patients (Figure 1).

Currently, each PGD device manages its own acquisition, storage, and data usage, while each health care organization acts as a third party to multiple query endpoints, with each requiring their own integration into clinical workflow. Data standards would enable PGD from multiple devices to flow through a single pipeline instead of multiple pipelines, with each serving 1 device.

**Figure 1.** The person-generated data (PGD) integration pipeline comprises 3 components: PGD acquisition, aggregation, and consumption.




### PGD Acquisition and Data Sharing Consent

PGD are acquired from patients via a diverse and growing ecosystem of health tracking apps, wearables, and sensors [10]. Typically, a device will require a patient to download a smartphone app to establish an account and pull data from the device to the smartphone through Bluetooth to store on the device company's cloud. Many devices provide an app or online dashboard where a patient or their physician can view tracked data [11-13].

Once data are acquired by the device, patients provide consent for data sharing either directly or via a separate PGD aggregator app that will serve the data to third-party solutions and their users (Figure 1). Existing aggregator apps include Apple Health, Google Fit, CommonHealth, Human API, and Validic. Apple Health and Google Fit allow patients to further share their data with any participating third party in the iOS and Android ecosystem, respectively, but with somewhat opaque rules by which third parties can request data and without any evaluation of clinical validity or security. CommonHealth, a nonprofit entrant to the personal health record/aggregator app space, differs by establishing a Common Trust Framework [14] in which patients can consent to share downloaded EHR or device data with trusted apps and services running on their phone. This framework is a neutral, independent set of rules that is developed through open-community governance.

Data-sharing consent can be granted at different levels of granularity. Patients may authorize their clinician to access only their BP data, while authorizing a clinical trial they are enrolled in access to BP, step count, weight, and calorie tracking data. Consent may also be revoked entirely or temporarily withheld for privacy or other reasons (eg, withholding weight data while on vacation).

## PGD Aggregation

Once patients consent to data sharing, an aggregator app's service processes their consent to mediate data transfer. PGD aggregation service components include authenticating third-party data requests, resolving whose data are being requested, managing authorization and consent, securely storing data (if needed), mapping data to standardized format(s), and exporting data in the desired standardized format to the third party.

## Authentication, Authorization, Identity Resolution, and Consent Management Handling

### *General Practice*

Standard industry procedures such as OAuth2 [15] are used for delegated authorization between PGD aggregator and third parties. Delegated authorization allows patients to authorize different services to access their data without services needing to expose personal credential information to each other. However, identity resolution between multiple services is challenging as it is common for patients to have several health care accounts (eg, their clinic, laboratory, and pharmacy).

Identity resolution within health care accounts, such as EHR services, is mediated via patient Fast Healthcare Interoperability Resources (FHIR) IDs. However, a patient will have different FHIR IDs for every health organization they access, and an organization may have multiple FHIR IDs for each patient depending on the back end implementation of FHIR servers. Without a national unique patient identifier, PGD aggregators will have to approximate patient identity. Linking many FHIR IDs and devices with apps requires tedious combinations of authorization flows subsequently further complicating consent management, as the PGD aggregator must match a patient's data-sharing consent against any third party requests for the patient's data. The complexity of consent management architectures argues strongly for standardized reusable multipurpose PGD integration pipelines.

### *Data Storage*

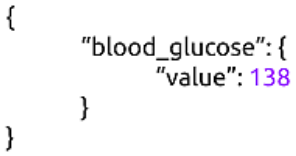
The PGD aggregator can either pass-through or store-and-forward data depending on the business need. With pass-through, the aggregator ingests data from the phone or device cloud and sends them directly to a third party at each request. With store-and-forward, the aggregator persists the data. Benefits of the pass-through approach include lower costs and security risks because the aggregator does not store data. Downsides include increased latency in data access, inability to perform computations (eg, average of requested values), and the need to repeat any mapping to standardized data formats. In a store-and-forward model, data can be persisted in native or in any standardized format.

PGD aggregators often have an on-phone and a cloud component. Some are PGD only (eg, Google Fit) while others (eg, Apple Health, CommonHealth) also aggregate EHR data. Apple Health and CommonHealth keep all synced data on the patient's smartphone; Google Fit uploads the data to Google Cloud.

### *Standardized Data and Metadata Export*

Most existing aggregators export PGD to third parties using their own nonstandardized formats [16-18]. CommonHealth, by contrast, exports data in standardized formats: EHR data are exported in Health Level 7 (HL7) FHIR format and PGD in Open mHealth/IEEE 1752.1 format. This difference is crucial. Clinically relevant contextual information is necessary for making clinical decisions. As shown in Figure 2, a blood glucose data of "138" is clinically meaningless unless the units, any relationship to meals or sleep, and effective time (ie, when the observation applied in the real world, not when the value was reported) are made clear. Standardized selection, definition, and value sets, as in Figure 3, for these contextual variables (eg, Unified Code for Units of Measure [UCUM] for units) would allow third-party systems to reliably and unambiguously understand the meaning of the PGD value, a minimal requirement for using PGD in health care or research.

**Figure 2.** This figure shows a JSON instance of a blood glucose value of 138. No other data or metadata are available.



```
{
    "blood_glucose": {
        "value": 138
    }
}
```


**Figure 3.** This figure shows an Open mHealth–compliant JSON instance of blood glucose with metadata showing that the value of 138 mg/dL is the average fasting value on awakening between February 5 and May 5, 2021.

```
{
    "blood_glucose": {
        "unit": "mg/dL",
        "value": 138
    },
    "effective_time_frame": {
        "time_interval": {
            "start_date_time": "2021-02-05T07:25:00Z",
            "end_date_time": "2021-05-05T07:25:00Z"
        }
    },
    "temporal_relationship_to_meal": "fasting",
    "temporal_relationship_to_sleep": "on waking",
    "descriptive_statistic": "average"
}
```


In addition to clinically relevant contextual information, use of PGD in health care or research also requires metadata [19,20]—data about the data. Examples include the name, model, and unique ID of the source device, and the unique ID of the app [21] installed on the patient's smartphone that acquired the data. Table 1 lists examples of metadata of interest for a sleep digital biomarker.

While there is no end to the types of metadata that would be of interest to someone for some purpose, it is infeasible to collect all possible metadata on all PGD. Nevertheless, a minimal set of critical metadata must be available on all PGD values to enable ecosystem-wide quality assurance, auditing, and regulatory oversight. The PGD ecosystem must therefore coalesce around a core set of data and metadata standards to enable long-term integrity and usability of PGD. Data can be standardized at the point of export by PGD devices or PGD aggregators can harmonize and provide endpoints for standardized PGD. Table 2 lists the standards that are most relevant to PGD. At the device level, standards such as IEEE 11073 [22] and FHIR's device resource [23] address manufacturing, security, privacy, and data export issues. For PGD integration, the Open mHealth/IEEE 1752.1.1 standard is the most directly relevant, covering the most widely used PGD variables for sleep, physical activity, cardiovascular, and other domains with over 80 JSON schemas [24,25]. Value sets are standardized using terms from Systematized Nomenclature of Medicine (SNOMED) or Logical Observation Identifiers Names and Codes (LOINC). A minimal metadata schema is used for the JSON schema header with standardized pointers to externally held metadata information (eg, an UDI registry). Open mHealth schemas are open source, free to all, and are the output of a global community of stakeholders consisting of developers, data scientists, informaticians, researchers, and clinicians. The sleep, physical activity, metadata, and utility schemas (on units, time, etc.) comprise the global standard IEEE 1752.1.1 [25].

**Table 1.** Metadata of a sleep digital biomarker.

| Metadata category | Example questions |
|---|---|
| What is the biomarker about? | Sleep duration? Sleep quality? Sleep refreshment? |
| Definition (eg, for total sleep duration) | Time in bed? Time asleep? With or without micro awakenings? |
| Validity | How does the biomarker compare with a gold standard? |
| Error | How much does it vary from the gold-standard value? |
| Natural variability | What is the natural variability within and among individuals, for comparison to the error range? |
| Uncertainty/Confidence | What is the probability that the person was asleep during this time? |
| Bias | Are there systematic errors in different populations? |
| Identity | Was the measurement collected for the right person? |
| Context | Was there relevant contextual information? For example, at home versus on a trip across time zones. |

**Table 2.** Selected standards relevant to mobile health.

| Standard | Description |
|---|---|
| HL7[a] FHIR[b] | HL7 refers to a set of international standards for transferring clinical and administrative data between health care providers. Within HL7, FHIR describes the data schema and application program interface for exchanging EHR[c] data. |
| IEEE 11073 | A family of standards for medical device communication, including point-of-care clinical devices and personal health devices. |
| IEEE 1752 | A family of standards for representation of person-generated health data, based on work by Open mHealth. |
| CTA[d] | A set of standards specifying how products work and the ways consumers interact with them. A subset of the standards pertain to consumer technologies in the health and fitness space [26]. |

[a]HL7: Health Level 7.

[b]FHIR: Fast Healthcare Interoperability Resources.

[c]EHR: electronic health record.

[d]CTA: Consumer Technology Association.

### PGD Consumption

Third-party users occupying the distal end of the PGD integration pipeline include health care organizations, researchers, and patients themselves (eg, consumers of an app that provides predictive analytics for blood glucose control). Many third parties want to be device agnostic. For example, a company providing decision support for BP management would want to accept BP data from any FDA-cleared brand and model of wireless BP cuff. Many third parties may also need to integrate heterogeneous data sources, such as reconciling sleep data from a smartwatch and a dedicated sleep sensor. Third parties would enjoy great efficiencies if PGD were available in a common data and metadata standard by not needing to divine the contextual meaning or metadata of PGD acquired from different sources. A standardized endpoint from a PGD aggregator would support the ideal of collecting PGD once and reusing them for multiple purposes.

## Illustrative Case

### Home Blood Pressure Integration

Home BP monitoring (HBPM) programs, in which dedicated staff monitor the home BPs of a panel of patients with hypertension for treatment support and adjustment, have shown efficacy in improving BP control [27]. Health care organizations are thus increasingly interested in establishing HBPM programs [8], which are reimbursable under several Centers for Medicare & Medicaid Services (CMS) billing codes but only if home BP measurements are acquired via wireless-connected cuffs and written directly into the health care organization's EHR [28]. We illustrate the PGD Integration Pipeline using the example of integrating wireless BP data into an EHR.

### Current Status: Home Blood Pressure Integration

Currently, home BPs from connected devices can be brought into an HBPM program through several pathways. One pathway is for patients to manually enter home BPs into an EHR patient portal. Despite its simplicity, this approach has many downsides. Using patient portals is challenging for patients with language barriers and low technology skills [29]. Manual reporting may result in fewer datapoints, is difficult to sustain over time [30], and evaluation and management of manually reported BP data are not reimbursable by CMS under the remote physiologic monitoring codes [28].

Another approach involves a partnership between a health care organization and a single wireless BP cuff company which will offer that company's online dashboard for clinicians to view. The need for clinicians to login to the company's website outside of their EHR severely disrupts workflow and is usually vehemently opposed by clinicians. Moreover, to qualify for CMS reimbursement, a custom interface has to be built and maintained to write data from that company into the EHR. Not only is this time-consuming and expensive, but it also severely limits flexibility. Adding another brand of cuff would require

an entirely new integration effort and online dashboard. Inertia to stay only with the initial company would be high. Such "vendor-lock" is inadvisable in a fast-changing digital health world.

An emerging approach takes advantage of Apple Health and Google Fit as PGD aggregators. At University of California, San Francisco (UCSF), which is on the Epic system, a pilot project is allowing clinicians to prescribe HBPM and have that prescription displayed on their patients' MyChart portal. Patients use one of several brands of cuffs, download both the device's app and the MyChart app onto their smartphone, and use Apple Health or Google Fit to consent and direct their BP data from the device's app into Epic. The ingested data can then be viewed within Epic and evaluation and management can be billed under CMS codes. This approach has the benefits of being device-agnostic, billable, and integrated into the EHR-based workflow. However, it is reliant on, and constrained by, Epic, MyChart, Apple Health, and Google Fit functionality and usability. Indeed, as of this writing, Google Fit has "temporarily stopped" accepting connected BP values and other "sensitive" health data types including body temperature and oxygen saturation [31]. Moreover, using PGD aggregators without data standards–driven infrastructure impairs robust PGD validation and use. For example, device manufacturer data are unavailable for query from either Apple Health or Google Fit endpoints.

### The Goal: Standardized Home Blood Pressure Integration

The optimal approach of using data standards throughout the BP integration pipeline offers many benefits. First, if device vendors adhered to data and metadata standards (eg, Open mHealth/IEEE 1752.1.1), the meaning and context of BP and other PGD would be captured for posterity at the source, which is ideal for downstream use, auditing, and regulatory oversight. If BP data are not standardized at the source, a standards-based PGD aggregator such as CommonHealth can ingest and map BP data from multiple vendors into Open mHealth or FHIR for standardized export. Health care organizations using a standards-based PGD aggregator are ensured that BP data will come in a consistent format with the same clinical contextual information and metadata, regardless of the cuff's brand or model. This device-agnostic predictability within a single integration pipeline yields great flexibility: multiple types of PGD from different device vendors can be integrated.

For health care organizations, standardization can facilitate data integration into workflow and writing into the EHR for billing. EHRs in the United States must now by law support HL7 FHIR data and protocol standards [32]. This allows EHRs to receive and display PGD using SMART-on-FHIR [33] protocols to launch dashboards directly in the EHR without requiring separate login. The mPROVE project at UCSF is taking this approach [34], displaying patient-reported outcomes and BP data in the BRIDGE SMART-on-FHIR dashboard [35]. Using SMART-on-FHIR frees data display and decision support presentation from the constraints of the EHR. While still in early adoption, SMART-on-FHIR technology has tremendous promise to augment the distal end of the PGD integration pipeline.

Another valuable benefit of data standardization for health care organizations is increased efficiency of data integration [36]. Instead of having to build and maintain custom connections to multiple device vendors, an organization receiving PGD in a common predictable format such as Open mHealth/IEEE 1752.1.1 can reuse the same interface for bringing PGD into their EHR or clinical workflow. Going forward, this singular interface can accommodate any new PGD data type that is supported by the data standard. The organization can flexibly switch to any other PGD aggregator that supports the same data standard because the PGD remains consistent for interfacing into the EHR.

Finally, the promise of PGD will be realized only if patients trust how their PGD will be handled, and if collecting, consenting, understanding, and sharing PGD are sufficiently easy to do [37]. To the extent that standardization of the PGD integration pipeline reduces data silos, multiple identities and accounts, and a profusion of opaque data-sharing mechanisms, trust will be enhanced for all parties and PGD integrity and value will be increased.

## Discussion

### Highlights

Today's mHealth data ecosystem—where multiple apps, devices, and proprietary aggregators each export data in their own data formats with little context or metadata—is suboptimal for unleashing the full capabilities of mHealth technologies to improve clinical care. Standards are key to successful data interchangeability and should be adopted broadly to enable device-agnostic solutions and modularity and to simplify the PGD ecosystem while simultaneously supporting data validation and data integrity.

### Relationship to Digital Biomarker Validation and App Frameworks

Deployment of PGD solutions in clinical care needs to extend beyond interoperability and integration. Various frameworks and best practices exist for choosing and deploying mHealth apps and sensors. HL7's Consumer Mobile Health Applications Functional Framework (cMHAFF) provides industry guidance and common methods to assess the "foundational characteristics," including but not limited to security, privacy, data access, data export, and transparency/disclosure of conditions, of mHealth apps [38]. HL7's App Data Exchange (ADE) project documents the functional requirements and provides a framework supporting data exchange between mHealth devices, apps, and other parts of the health IT Infrastructure [39]. The ADE project references mHealth data standards such as Open mHealth/IEEE 1752.1.1 and IEEE 11073.

By themselves, neither cMHAFF nor ADE address the clinical validity or value of an mHealth solution. The DiMe Playbook is a "comprehensive 'how-to' guide" on developing, selecting, and deploying digital biomarkers. It addresses digital biomarker verification, analytical validation, and clinical validation (V3) as well as the role of standards such as Open mHealth/IEEE 1752.1.1 in data integration [40].

### Toward Interoperability by Design

Like privacy, data provenance and interoperability should be intentionally designed into a system up-front rather than shoehorned into it later on [41]. For the mHealth ecosystem, a mix of frameworks, official data standards and protocols, and best practices as reviewed above is beginning to paint a path out of today's fragmented silos. For the PGD integration pipeline in particular, the path includes PGD devices and mHealth apps exporting and consuming digital biomarkers in the Open mHealth/IEEE 1752.1 format where appropriate; expanding the data types standardized by Open mHealth/IEEE 1752.1; data aggregators exporting PGD in both their current format (to ensure backward compatibility) and Open mHealth/IEEE 1752.1 format (to transition toward standardized interoperability); and finally, wide adoption of the Open mHealth-to-FHIR implementation guide as the common FHIR observation resource profile for PGD [42]. These steps offer a glidepath for the ecosystem to transition to data and metadata standards that themselves evolve to accommodate new digital biomarkers and new metadata frameworks. Further research is needed on scalable metadata acquisition and management, biomarker validation platforms, and interoperability with the broader internet of things.

### Conclusion

The clinical value of PGD from mHealth apps and sensors is currently limited by difficult and inefficient integration into routine clinical care. Major components of the PGD integration pipeline include PGD acquisition, PGD aggregation, and third-party solutions that consume PGD to deliver end value for clinical care and clinical research, all while retaining people's control on their data and trust in the process. Standardization of data and metadata along the entire PGD integration pipeline is crucial for ensuring device-agnostic, modular, flexible, multipurpose, and thus lower-cost integration into clinical workflow. The value of efficient integration of PGD data will increase revenue streams, reduce overhead, improve data integrity, and facilitate patient trust. PGD aggregation services that offer standards-based PGD integration play a vital role in transitioning from today's siloed friction-heavy data ecosystem to a low-friction interoperating system that our patients deserve. Health leaders responsible for remote monitoring and other PGD programs should seek out and adopt pipeline-based approaches to standardize the integration of PGD into clinical care.

### Acknowledgments

This work was partially supported by grants R24EB025845 and U18HS26883, and by a Mount Zion Health Fund grant (no. 20201224) from the University of California, San Francisco.

### Authors' Contributions

IS, SC, and JPP conceived the central idea of the paper. BZ and IS wrote the initial draft. IS secured funding that partially supported this work. All coauthors (BZ, RB, SC, JSJL, JPP, ES, and IS) contributed to, reviewed, and approved the manuscript.

### Conflicts of Interest

IS is a cofounder of Open mHealth and a trustee of The Commons Foundation, and declares the following competing interests: financial support from Myovant and Vivli, as well as nonfinancial support from Open mHealth, 98point6, and The Commons Project Foundation. RB receives research support from the National Institutes of Health, California Initiative to Advance Precision Medicine, National Multiple Sclerosis Society (Harry Weaver Award), Hilton Foundation, Sherak Foundation, as well as Biogen, Novartis and Roche Genentech. She has received scientific advisory board and consulting fees from Alexion, Biogen, EMD Serono, Genzyme Sanofi, Novartis, and Roche Genentech. JPP is a cofounder of and declares financial support from The Commons Project Foundation. SC is a data scientist at Open mHealth, and declares nonfinancial support from Open mHealth and The Commons Project Foundation. BZ, JSJL, and ES have no financial disclosures to declare.

### References

1. Goldsack J, Dowling A, 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]
2. Majumder S, Mondal T, Deen M. Wearable Sensors for Remote Health Monitoring. Sensors (Basel) 2017 Jan 12;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
3. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
4. Ray P. A survey on Internet of Things architectures. Journal of King Saud University - Computer and Information Sciences 2018 Jul;30(3):291-319. [doi: 10.1016/j.jksuci.2016.10.003]
5. Meinert E, Van Velthoven M, Brindley D, Alturkistani A, Foley K, Rees S, et al. The Internet of Things in Health Care in Oxford: Protocol for Proof-of-Concept Projects. JMIR Res Protoc 2018 Dec 04;7(12):e12077 [FREE Full text] [doi: 10.2196/12077] [Medline: 30514695]

6. La Lau R. Email Basics. In: La Lau R, editor. Practical Internet Server Configuration. Berkeley, CA: Apress; Aug 2021:281-307.
7. UPMC Vice President Katie Scott explains the importance of providing a quality digital patient experience. 2020. URL: https://connectedmed.com/resources/why-health-systems-must-prioritize-the-digital-patient-experience/ [accessed 2021-05-19]
8. Parati G, Dolan E, McManus R, Omboni S. Home blood pressure telemonitoring in the 21st century. J Clin Hypertens (Greenwich) 2018 Jul;20(7):1128-1132 [FREE Full text] [doi: 10.1111/jch.13305] [Medline: 30003701]
9. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018 Jun 20;1(1):23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
10. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
11. Remote Patient Monitoring | Chronic Conditions Management | RPM | CCM Internet. URL: https://ihealthlabs.com/pages/remote-patient-monitoring-rpm [accessed 2021-05-19]
12. OMRON Apps for Connected Products & Bluetooth Devices Internet. URL: https://omronhealthcare.com/service-and-support/connected-health/ [accessed 2021-05-19]
13. Withings Health Mate: a Fitness, Activity and Health Tracker App. URL: https://www.withings.com/us/en/health-mate [accessed 2021-05-19]
14. CommonTrust Network. URL: https://www.commontrustnetwork.org [accessed 2021-05-25]
15. Sciancalepore S, Piro G, Caldarola D, Boggia G, Bianchi G. OAuth-IoT: An access control framework for the Internet of Things based on open standards. 2017 Sep 04 Presented at: 2017 IEEE Symposium on Computers and Communications (ISCC); July 3-7, 2017; Heraklion, Greece. [doi: 10.1109/iscc.2017.8024606]
16. Data types | Google Fit. URL: https://developers.google.com/fit/datatypes [accessed 2021-05-19]
17. Data Types | Apple Developer Documentation. URL: https://developer.apple.com/documentation/healthkit/data_types [accessed 2021-05-19]
18. Getting Started: Human API. URL: https://reference.humanapi.co/reference [accessed 2021-05-19]
19. Gagnon M. Ontology-based integration of data sources. 2007 Dec 26 Presented at: 2007 10th International Conference on Information Fusion; August 2007; Quebec, QC, Canada p. 1-8 URL: https://www.researchgate.net/publication/4301389_Ontology-based_integration_of_data_sources [doi: 10.1109/icif.2007.4408086]
20. Hume S, Sarnikar S, Noteboom C. Enhancing Traceability in Clinical Research Data through a Metadata Framework. Methods Inf Med 2020 May 07;59(2-03):75-85. [doi: 10.1055/s-0040-1714393] [Medline: 32894879]
21. Unique Mobile Health Application Identifier (UMHAI) | Interoperability Standards Advisory (ISA). URL: https://www.healthit.gov/isa/uscdi-data/unique-mobile-health-application-identifier-umhai [accessed 2021-05-01]
22. IEEE 11073-10207-2017 - IEEE Health informatics--Point-of-care medical device communication Part 10207: Domain Information and Service Model for Service-Oriented Point-of-Care Medical Device Communication. URL: https://standards.ieee.org/standard/11073-10207-2017.html [accessed 2021-05-01]
23. Device - FHIR v4.0.1. URL: https://www.hl7.org/fhir/device.html [accessed 2021-05-25]
24. Documentation Open MHealth. URL: https://www.openmhealth.org/documentation/ [accessed 2021-04-06]
25. IEEE Standard for Open Mobile Health Data—Representation of Metadata, Sleep, and Physical Activity Measures. IEEE Std 17521-2021. URL: https://standards.ieee.org/ieee/1752.1/6982/ [accessed 2022-01-29]
26. Standards. URL: https://shop.cta.tech/collections/standards/health-and-fitness [accessed 2021-05-25]
27. Muntner P, Einhorn PT, Cushman WC, Whelton PK, Bello NA, Drawz PE, 2017 National Heart, Lung,Blood Institute Working Group. Blood Pressure Assessment in Adults in Clinical Practice and Clinic-Based Research: JACC Scientific Expert Panel. J Am Coll Cardiol 2019 Jan 29;73(3):317-335 [FREE Full text] [doi: 10.1016/j.jacc.2018.10.069] [Medline: 30678763]
28. Medicare Program; CY 2021 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Medicaid Promoting Interoperability Program Requirements for Eligible Professionals; Quality Payment Program; Coverage of Opioid Use Disorder Services Furnished by Opioid Treatment Programs; Medicare Enrollment of Opioid Treatment Programs; Electronic Prescribing for Controlled Substances for a Covered Part D Drug; Payment for Office/Outpatient Evaluation and Management Services; Hospital IQR Program; Establish New Code Categories; Medicare Diabetes Prevention Program (MDPP) Expanded Model Emergency Policy; Coding and Payment for Virtual Check-in Services Interim Final Rule Policy; Coding and Payment for Personal Protective Equipment (PPE) Interim Final Rule Policy; Regulatory Revisions in Response to the Public Health Emergency (PHE) for COVID-19; and Finalization of Certain Provisions from the March 31st, May 8th and September 2nd Interim Final Rules in Response to the PHE for COVID-19. URL: https://www.federalregister.gov/documents/2020/12/28/2020-26815/medicare-program-cy-2021-payment-policies-under-the-physician-fee-schedule-and-other-changes-to-part [accessed 2021-05-19]
29. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

30. Celler B, Argha A, Varnfield M, Jayasena R. Patient Adherence to Scheduled Vital Sign Measurements During Home Telemonitoring: Analysis of the Intervention Arm in a Before and After Trial. JMIR Med Inform 2018 May 09;6(2):e15 [FREE Full text] [doi: 10.2196/medinform.9200] [Medline: 29631991]
31. Write Blood Pressure Data | Google Fit. URL: https://developers.google.com/fit/scenarios/write-bp-data [accessed 2021-05-19]
32. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. URL: https://tinyurl.com/2p8bktc4 [accessed 2021-05-25]
33. SMART on FHIR API. 2019. URL: https://smarthealthit.org/smart-on-fhir-api/ [accessed 2021-10-26]
34. Improving the Management of Multiple Chronic Conditions with mPROVE (California) | AHRQ Digital Healthcare Research: Informing Improvement in Care Quality, Safety, and Efficiency. URL: https://digital.ahrq.gov/ahrq-funded-projects/improving-management-multiple-chronic-conditions-mprove [accessed 2021-05-15]
35. UCSF BRIDGE. URL: https://bridge.ucsf.edu/ [accessed 2021-03-10]
36. Ismail L, Materwala H, Karduck AP, Adem A. Requirements of Health Data Management Systems for Biomedical Care and Research: Scoping Review. J Med Internet Res 2020 Jul 07;22(7):e17508 [FREE Full text] [doi: 10.2196/17508] [Medline: 32348265]
37. Kukafka R. Digital Health Consumers on the Road to the Future. J Med Internet Res 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: 10.2196/16359] [Medline: 31750835]
38. HL7 Consumer Mobile Health Application Functional Framework (CMHAFF) Internet. URL: https://cmhaff.healtheservice.com/ [accessed 2021-05-02]
39. HL7 FHIR® Implementation Guide: mHealth ADE Assessment Framework, Release 1 - Mobile Health - Confluence. URL: https://confluence.hl7.org/pages/viewpage.action?pageId=80121539 [accessed 2021-05-26]
40. The Playbook - DiMe. URL: https://playbook.dimesociety.org/playbooks/the-playbook/ [accessed 2021-05-03]
41. Vaidya A. Interoperability has to move from data exchange to data management Internet. MedCity News. URL: https://tinyurl.com/3tpac8sv [accessed 2021-03-10]
42. openmhealth/OMH-on-FHIR Internet. URL: https://github.com/openmhealth/OMH-on-FHIR [accessed 2021-05-10]


## Abbreviations

**ADE:** App Data Exchange
**BP:** blood pressure
**cMHAFF:** Consumer Mobile Health Applications Functional Framework
**CMS:** Center for Medicare and Medicaid Services
**EHR:** electronic health record
**FHIR:** Fast Healthcare Interoperability Resources
**HBPM:** home blood pressure monitoring
**HL7:** Health Level 7
**LOINC:** Logical Observation Identifiers Names and Codes
**PGD:** person-generated data
**SNOMED:** Systematized Nomenclature of Medicine
**UCSF:** University of California, San Francisco
**UCUM:** Unified Code for Units of Measure


*Edited by L Buis; submitted 07.06.21; peer-reviewed by A Hidki, J Seitz, I Shubina; comments to author 27.07.21; revised version received 31.10.21; accepted 20.12.21; published 10.02.22.*

*Please cite as:*
*Zeng B, Bove R, Carini S, Lee JSJ, Pollak JP, Schleimer E, Sim I*
*Standardized Integration of Person-Generated Data Into Routine Clinical Care*
*JMIR Mhealth Uhealth 2022;10(2):e31048*
*URL: https://mhealth.jmir.org/2022/2/e31048*
*doi:10.2196/31048*
*PMID:35142627*

Viewpoint

# Smart Speakers: The Next Frontier in mHealth

Jacob Sunshine[1,2], MD

[1]Department of Anesthesiology & Pain Medicine, University of Washington, Seattle, WA, United States
[2]Paul G Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, United States

**Corresponding Author:**
Jacob Sunshine, MD
Department of Anesthesiology & Pain Medicine
University of Washington
1959 NE Pacific Street
Box 356540
Seattle, WA, 98195
United States
Phone: 1 206 543 6814
Email: jesun@uw.edu

## Abstract

The rapid dissemination and adoption of smart speakers has enabled substantial opportunities to improve human health. Just as the introduction of the mobile phone led to considerable health innovation, smart speaker computing systems carry several unique advantages that have the potential to catalyze new fields of health research, particularly in out-of-hospital environments. The recent rise and ubiquity of these smart computing systems holds significant potential for enhancing chronic disease management, enabling passive identification of unwitnessed medical emergencies, detecting subtle changes in human behavior and cognition, limiting isolation, and potentially allowing widespread, passive, remote monitoring of respiratory diseases that impact public health. There are 3 broad mechanisms for how a smart speaker can interact with a person to improve health. These include (1) as an intelligent conversational agent, (2) as a passive identifier of medically relevant diagnostic sounds, and (3) by active sensing using the device's internal hardware to measure physiologic parameters, such as with active sonar, radar, or computer vision. Each of these different modalities has specific clinical use cases, all of which need to be balanced against potential privacy concerns, equity concerns related to system access, and regulatory frameworks which have not yet been developed for this unique type of passive data collection.

***(JMIR Mhealth Uhealth 2022;10(2):e28686)*** doi:10.2196/28686

**KEYWORDS**

digital health; mobile health; machine learning; smart speaker; smartphone

## Background

The rapid dissemination and adoption of smart speakers has enabled substantial opportunities to improve human health. Just as the introduction of the mobile phone led to considerable health innovation, and ultrasound enabled new opportunities for point-of-care diagnosis and procedural optimization, smart speaker computing systems carry several unique advantages that can catalyze new fields of research, particularly in out-of-hospital environments. The recent rise and ubiquity of these smart computing systems, which are often cheaper than smartphones and substantially less expensive than medical grade equipment, holds significant potential for enhancing chronic disease management, enabling passive identification of unwitnessed medical emergencies, detecting subtle changes in human behavior and cognition, limiting isolation, and potentially allowing widespread, passive, remote monitoring of respiratory-based infectious diseases which impact public health, all while still providing general utility for users. Advances in machine-based classification of disease states, capable of being run on-device and securely in the cloud, can enable rapid diagnostic and predictive functions at a low cost while preserving privacy. This confluence of factors has created a significant opportunity involving these devices, which currently reside in 1 of 4 US households, when applied thoughtfully to carefully chosen health conditions [1].

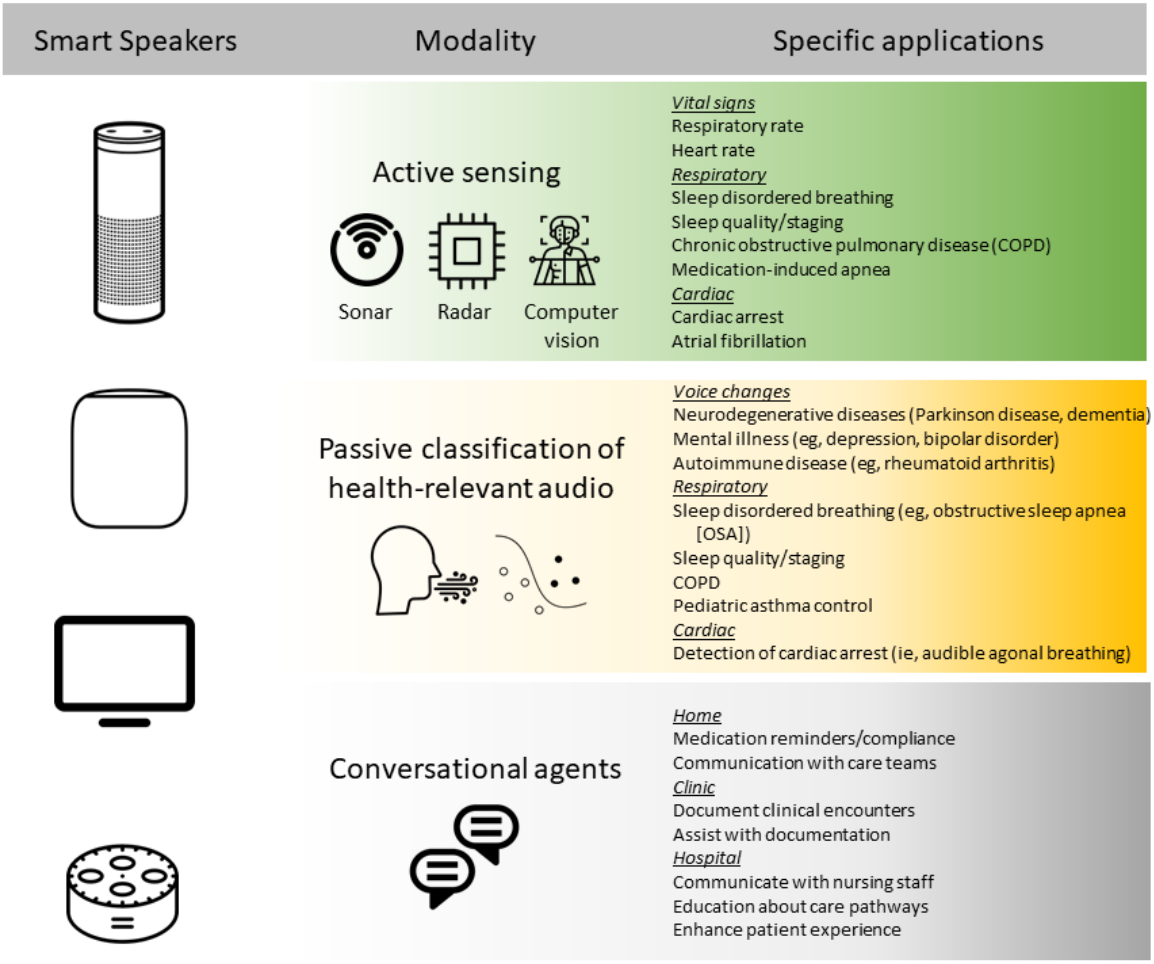
## What Are Smart Speakers and How Are They Different From Smartphones?

At its most basic form, a smart speaker is a system comprising a speaker, a microphone array, an embedded computer, a

software- and machine learning–based intelligent assistant, and wireless connectivity that enables data integration with the cloud, nearby smart devices, and other information technology (IT) infrastructures outside of the home. The increasing computational horsepower of embedded platforms coupled with advances in machine learning have enabled on-device capabilities that remove the need to transmit audio to the cloud. As such, the system has the capability to continuously monitor the home environment and instruct a patient on or converse with them about a medically relevant topic, identify health-related audible biomarkers, sense the environment for contextually relevant health-related motion, and much more. And because these computing systems have wireless capability, they can transmit data to the cloud for secure storage and analysis, if desired. Such connectivity also, in theory, enables integration with medical IT infrastructures, so a trained provider can interpret, triage, and act upon relevant information from a smart speaker, or in an emergent context, connect with an emergency response system (eg, 911) to summon help. Key differentiators of these devices compared to mobile phones include that they are plugged in, thus avoiding power constraints that are associated with charging a device; they are predominantly stationary, enabling long-term, passive, and continuous monitoring; and their range of measurements is greater than a phone, which generally must be interacted with when it is directly in a user's hands. The inherent constraints of their placement, moreover, provide a substantive benefit by reducing the number of "edge cases" that invariably arise when building intelligent sensing systems. Yet, perhaps smart speakers' biggest advantage over mobile phones and other wearable devices is their ability to foster compliance [2,3] by not requiring patients to wear or do anything after initial setup (ie, they can be truly "set and forget").

Against this background, there are 3 broad mechanisms for how a smart speaker can interact with a patient to improve health. These include (1) as an intelligent conversational agent, (2) as a passive identifier of medically relevant diagnostic sounds, and (3) by active sensing using the device's internal hardware to measure physiologic parameters, such as with active sonar, radar, or computer vision (Figure 1).

**Figure 1.** Overview of how smart speakers can enhance health and well-being.




## Smart Speakers as Health Conversation Agents

The first deployed and most straightforward use for smart speakers is as intelligent conversational agents and facilitators. These applications generally rely on voice user interfaces (VUIs), which enable the user to interact with the system using their voice and allow them to receive medically relevant auditory feedback [4]. In the home environment, conversational use cases include the system providing reminders to take medications, retrieving recent lab results (eg, blood sugar), managing medical

appointments, and tracking wellness goals [5]. These systems are also capable of reducing isolation, particularly in older adults, by providing a low-barrier way to facilitate communication (eg, with family members, caretakers, social workers), and detecting signals in the environment where a check-in may be warranted (eg, a change or reduction in activities of daily living). Outside of the home, these devices also have a role in the clinic and the inpatient environment. Within the clinic, these devices may soon be used to help liberate physicians from their computers, as provider-patient conversations are passively captured, parsed, and analyzed to efficiently document medical encounters [6]. Devices have also been deployed in hospitals, particularly in patient rooms, primarily as a way to improve the patient experience [7], and in the era of COVID-19, to provide a crucial means of communication with the care team and family members unable to visit the patient [8].

## Classification of Medically Relevant Diagnostic Sounds

The next level of interaction with these devices is as a classifier of medically relevant, contextually appropriate biosignals that represent signs and symptoms of disease. There have been major advances in sound classification research in the computing community [9-11] that have implications for medically informative audio [12]. Researchers are examining publicly available data sets from the computing community, such as AudioSet [13], to relabel and train new models for medically relevant sounds [14]. In this use case, which would predominate in home environments, the computing system classifies certain audible biomarkers for the purposes of diagnosis or to better inform disease management. Similar to invoking certain trigger words (eg, “Hey Siri,” “Alexa,” “Hey Bixby,” “OK, Google”), these systems are capable of passively identifying specific audio signatures that are contextually relevant and of medical utility. Building on classification guidance from the National Institutes of Health (NIH) and the US Food and Drug Administration (FDA), Coravos et al [15] have proposed a useful framework of digital biomarkers, which classifies signals as they relate to susceptibility or risk, diagnosis, monitoring, prognostication, and prediction. These audio biomarkers can be used to detect and classify coughs [16,17], discern voice changes arising from neurodegenerative diseases such as Parkinson disease [18] or dementia [19], characterize voice changes related to depression [20,21] or other mental illnesses [22], classify breathing patterns associated with obstructive sleep apnea (OSA) [23], identify deteriorating asthma [24], and even identify unwitnessed cardiac arrest by detecting the presence of agonal breathing [11].

## Active Sensing Using Smart Devices

The final way that these computing systems can be used is perhaps the most innovative and involves turning these devices into contactless active sensing systems using computer vision, sonar, or radar for the purposes of physiologic monitoring. If the smart speaker has a camera, this enables important diagnostic capabilities aided by computer vision, which enables a machine to make inferences based on dynamic images and subtle changes in pixelation. Notable potential use cases for computer vision include the detection of falls [25], respiratory and heart rate monitoring [26,27], identifying significant changes in activity in older populations [28], self-monitoring of physical therapy, monitoring of acute and chronic wounds [29], and postoperative- and posthospitalization-based rehabilitation within the home. In addition, because these devices have speakers and microphones, they are capable of active sonar and echolocation utilizing high (>18 kHz), inaudible frequencies to detect medically relevant motion. Some smart speakers are already enabling these features for activity sensing and gesture detection. A benefit of this method is that, because it utilizes inaudible frequencies, it can collect relevant data while filtering out all audible speech and thus preserves privacy. Similarly, in radar-based systems, electromagnetic waves are transmitted into the environment and phase changes in the reflected signals can be used to classify medically relevant motion. The potential use cases of these sonar- and radar-based active sensing modalities include monitoring of chest motion or breathing [30] and its perturbations (pertinent for asthma [31], chronic obstructive pulmonary disease [COPD] [32], OSA [33], and opioid overdose [34]), sleep disturbance (eg, insomnia), identification of incipient respiratory infection, measurement of cardiac activity (eg, heart rate and atrial fibrillation) [35], monitoring of activity levels based on movement, epilepsy monitoring, and more.

## Privacy

As with any ubiquitous computing system, a critical consideration relates to privacy, which can mean different things to different people. For a health monitoring context, this refers to monitoring that, similar to the default functionality of these devices, enables continuous “listening,” but only processes and stores (if the user desires) relevant health data. In practice, using asthma or COPD as an example, the system would not store or analyze conversations, though it would recognize, document, and analyze increases in nocturnal cough or relevant changes in respiration, such as dyspnea or audible wheezing. It is important that any health-related data approved to be stored are stored securely within an environment designed to be compliant with the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GPDR), and that the data belong to and are made easily accessible to individuals. Just as there are potential privacy concerns associated with smartphones and personal computers, there is a point where their utility outweighs their real and perceived privacy concerns. The adoption arc for smart speakers is undoubtedly affected by these concerns, presenting a challenge but also an opportunity to develop innovative privacy-preserving functionality that would make the collection of health data more comfortable and trustworthy. These efforts would be greatly enhanced by manufacturers taking straightforward, transparent actions that foster trust and maximize control of information for the monitored user.

## Barriers to Implementation and Future Directions

Although there is tremendous potential for this new computing platform to potentially improve human health, there remain several barriers. The first major barrier is the lack of an open ecosystem, compared to the development environment and regulatory frameworks for applications that can run on smartphones, tablets, or PCs. Crucially, there is no app store or developer environment that provides the level of access to firmware that would enable flexible development of innovative, high-quality, medically relevant applications which take advantage of a device's internal hardware. For example, unlike on Android or iOS, a developer cannot leverage the smart speaker's camera, individual speaker(s), or microphones for the purposes of app development. Although the major smart speaker manufactures allow for the development of "skills" or plug-ins within a highly constrained design framework, including at least one enabling secure transmission of health information [36], they do not offer the openness and flexibility that exists for the development of health-related applications intended for smartphones. Such an ecosystem would represent a substantial opportunity for health-related software development and would leverage these devices' full computational capabilities.

Control of data flow for regulatory and HIPAA standards is also critical in health care use cases. Regulatory organizations, health system stakeholders, and computing communities need to come together to develop an agreement on the responsible use of data for these emerging technologies. In particular, it is unclear what protections are needed for data generated in the home that *could* be used for health purposes compared to data that is generated in a clinic or hospital encounter, where protections are clearly enumerated for patient data. Current regulatory guidance does not take into account these new sources of data generated in the home, which will have to be addressed as these computing systems become more common for health purposes. Relatedly, thoughtful care must be taken when using voice or medically relevant audio as a passively measured biomarker. Such measurements are primarily relevant to the intended monitored user, who would have consented to these biosignals being collected, processed, and stored. Yet, such a design has implications when others are in close proximity to these systems because their biosignals could be captured without having provided explicit consent. Although there are several examples of people being monitored in everyday life without their explicit consent (eg, security-based audiovisual observation or being in the presence of others' smart devices), passive health sensing must be undertaken with particular care given the nature of the data being collected.

Another critical consideration with passive systems deployed on ubiquitous devices is the need to minimize false positives. Generally, it is not wise to use these systems for asymptomatic screening of healthy populations given the dangers of excessive false positives. Using these systems to monitor specific patient populations at risk for certain physiologic perturbations that are clinically meaningful is more likely to be useful to the patient and care teams generally. Toward this end, following identification of a given biomarker or aberrant trend, effective uses of these systems will likely require a level of interactivity (via screen or voice) to collect further information, such as pertinent positives and negatives, before consequential actions or referrals are executed. Additionally, as these computing systems mature as tools for research, they will require a research platform that can enable vetted, high-quality studies at scale, similar to Apple's ResearchKit, Sage Bionetwork's Bridge Platform, and CareEvolution's MyDataHelps. Such research is essential to demonstrate the health utility of these platforms, which will require actual clinical evidence to gain trust from patients, care teams and health systems. Finally, when used for health purposes, it is essential these devices do not exacerbate health disparities, for example, by being differentially accessible to certain populations. Concrete ways to reduce inequities include programs that make smart speakers, when indicated, accessible to those who desire them but may not be able to afford the cost. Similarly, if used for health purposes and prescribed by a care team, these systems should be readily covered by payers. Lastly, it is imperative that application VUIs and non-VUIs encompass as many languages as possible and, particularly for VUIs, that performance differences across language, age, sex, and gender are actively minimized and eventually eliminated.

## Conclusion

In summary, smart speakers represent a new, ubiquitous computing platform within our home environments, which hold considerable untapped potential to improve human health at low cost, and if done thoughtfully, in ways that foster high compliance and preserve privacy. The primary health benefits are likely to be observed with enhanced chronic disease management, early detection of unwitnessed emergencies and indolent neurodegenerative processes, and enhancements of the patient and provider experience in clinic and inpatient environments. Achieving this unrealized potential will require smart speaker manufacturers to open their platforms to developers as they have with smartphones, develop an ecosystem specifically for medically oriented applications and research, and enable and relentlessly prioritize privacy-preserving functionality.

**Acknowledgments**

JS is supported by the National Science Foundation (Grants 1914873149 and 181255) and the NIH (Grants K23DA046686, R44DA050339). The author thanks Shwetak Patel, PhD for feedback on the manuscript.

## Conflicts of Interest

JS holds an equity stake in Sound Life Sciences Inc.

## References


1. The smart audio report. National Public Media. URL: https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/ [accessed 2020-02-25]
2. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
3. Galarnyk M, Quer G, McLaughlin K, Ariniello L, Steinhubl SR. Usability of a wrist-worn smartwatch in a direct-to-participant randomized pragmatic clinical trial. Digit Biomark 2019;3(3):176-184 [FREE Full text] [doi: 10.1159/000504838] [Medline: 32095776]
4. Stigall B, Waycott J, Baker S, Caine K. Older adults' perception and use of voice user interfaces: a preliminary review of the computing literature. In: Proceedings of the 31st Australian Conference on Human-Computer-Interaction. 2019 Dec 02 Presented at: 31st Australian Conference on Human-Computer-Interaction; December 2-5, 2019; Fremantle, Australia p. 423-427. [doi: 10.1145/3369457.3369506]
5. Jiang R. Introducing new Alexa healthcare skills. Amazon Developer Services and Technologies: Amazon Alexa. 2019. URL: https://developer.amazon.com/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills [accessed 2020-03-01]
6. Langston J. Microsoft and Nuance join forces in quest to help doctors turn their focus back to patients. Official Microsoft Blog. 2019. URL: https://blogs.microsoft.com/ai/nuance-exam-room-of-the-future/ [accessed 2020-02-22]
7. Dietsche E. Pilot project brings Alexa to Cedars-Sinai patients. MedCity News. 2019. URL: https://medcitynews.com/2019/02/alexa-cedars-sinai/ [accessed 2020-01-28]
8. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
9. Siri Team. Hey siri: an on-device DNN-powered voice trigger for Apple's personal assistant. Apple Machine Learning Research. 2017. URL: https://machinelearning.apple.com/research/hey-siri [accessed 2022-02-09]
10. Models for Audioset: a large scale dataset of audio events. GitHub. URL: https://github.com/tensorflow/models/tree/master/research/audioset [accessed 2020-03-01]
11. Muda L, Begam M, Elamvazuthi I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. J Comput 2010;2(3):138-143 [FREE Full text]
12. Chan J, Rea T, Gollakota S, Sunshine JE. Contactless cardiac arrest detection using smart devices. NPJ Digit Med 2019;2:52 [FREE Full text] [doi: 10.1038/s41746-019-0128-7] [Medline: 31304398]
13. Gemmeke J, Ellis D, Freedman D, Jansen A, Lawrence W, Moore R, et al. Audio Set: an ontology and human-labeled dataset for audio events. 2017 Presented at: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP); March 5-9, 2017; New Orleans, LA p. 776-780. [doi: 10.1109/icassp.2017.7952261]
14. Al Hossain F, Lover AA, Corey GA, Reich NG, Rahman T. FluSense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2020 Mar 18;4(1):1-28. [doi: 10.1145/3381014]
15. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
16. Larson E, Lee T, Liu S, Rosenfeld M, Patel S. Accurate and privacy preserving cough sensing using a low-cost microphone. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: 13th international conference on Ubiquitous computing; 2011; Beijing, China p. 375-384. [doi: 10.1145/2030112.2030163]
17. Sun X, Lu Z, Hu W, Cao G. SymDetector: detecting sound-related respiratory symptoms using smartphones. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka, Japan p. 97-108. [doi: 10.1145/2750858.2805826]
18. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
19. Chen R, Jankovic F, Marinsek N, Foschini L, Kourtis L, Signorini A, et al. Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019; Anchorage, AK p. 2145-2155. [doi: 10.1145/3292500.3330690]
20. Tasnim M, Stroulia E. Detecting depression from voice. In: Advances in Artificial Intelligence. 2019 Presented at: 32nd Canadian Conference on Artificial Intelligence; May 28–31, 2019; Kingston, Canada p. 472-478. [doi: 10.1007/978-3-030-18305-9_47]

21. Huang Z, Epps J, Joachim D, Chen M. Depression detection from short utterances via diverse smartphones in natural environmental conditions. 2018 Presented at: Interspeech 2018; September 2-6, 2018; Hyderabad, India p. 3393-3397. [doi: 10.21437/interspeech.2018-1743]
22. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6:e856. [doi: 10.1038/tp.2016.123] [Medline: 27434490]
23. Dafna E, Tarasiuk A, Zigel Y. Sleep staging using nocturnal sound analysis. Sci Rep 2018 Sep 07;8(1):13474 [FREE Full text] [doi: 10.1038/s41598-018-31748-0] [Medline: 30194402]
24. Larson E, Goel M, Boriello G, Heltshe S, Rosenfeld M, Patel S. SpiroSmart: using a microphone to measure lung function on a mobile phone. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: 2012 ACM Conference on Ubiquitous Computing; September 2012; Pittsburgh, PA p. 280-289. [doi: 10.1145/2370216.2370261]
25. Anderson D, Keller J, Skubic M, Chen X, He Z. Recognizing falls from silhouettes. 2006 Presented at: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 30-September 3, 2006; New York, NY p. 6388-6391. [doi: 10.1109/iembs.2006.259594]
26. Chatterjee A, Prathosh A, Praveena P. Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20, 2016; Orlando, FL p. 2708-2711. [doi: 10.1109/embc.2016.7591289]
27. Yan BP, Lai WHS, Chan CKY, Au ACK, Freedman B, Poh YC, et al. High-throughput, contact-free detection of atrial fibrillation from video with deep learning. JAMA Cardiol 2020 Jan 01;5(1):105-107. [doi: 10.1001/jamacardio.2019.4004] [Medline: 31774461]
28. Luo Z, Hsieh JT, Balachandar N, Yeung S, Pusiol G, Luxenberg J, et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. In: Proceedings of Machine Learning Research. 2018 Presented at: Machine Learning for Healthcare Conference, 17-18 August 2018, Palo Alto, California; August 17-18, 2018; Palo Alto, CA p. 1-18.
29. Gunter R, Fernandes-Taylor S, Mahnke A, Awoyinka L, Schroeder C, Wiseman J, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113 [FREE Full text] [doi: 10.2196/mhealth.6023] [Medline: 27683059]
30. Wang G, Munoz-Ferreras J, Gu C, Li C, Gomez-Garcia R. Linear-frequency-modulated continuous-wave radar for vital sign monitoring. 2014 Presented at: 2014 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet); January 19-23, 2014; Newport Beach, CA p. 1387-1399. [doi: 10.1109/wisnet.2014.6825496]
31. Huffaker MF, Carchia M, Harris BU, Kethman WC, Murphy TE, Sakarovitch CCD, et al. Passive nocturnal physiologic monitoring enables early detection of exacerbations in children with asthma. A proof-of-concept study. Am J Respir Crit Care Med 2018 Aug 01;198(3):320-328 [FREE Full text] [doi: 10.1164/rccm.201712-2606OC] [Medline: 29688023]
32. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2000 May;161(5):1608-1613. [doi: 10.1164/ajrccm.161.5.9908022] [Medline: 10806163]
33. Nandakumar R, Gollakota S, Watson N. Contactless sleep apnea detection on smartphones. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services. 2015 Dec 23 Presented at: 13th Annual International Conference on Mobile Systems, Applications, and Services; May 18-22, 2015; Florence, Italy p. 22-24. [doi: 10.1145/2867070.2867078]
34. Nandakumar R, Gollakota S, Sunshine JE. Opioid overdose detection using smartphones. Sci Transl Med 2019 Jan 09;11(474):eaau8914. [doi: 10.1126/scitranslmed.aau8914] [Medline: 30626717]
35. Wang A, Nguyen D, Sridhar AR, Gollakota S. Using smart speakers to contactlessly monitor heart rhythms. Commun Biol 2021 Mar 09;4(1):319 [FREE Full text] [doi: 10.1038/s42003-021-01824-9] [Medline: 33750897]
36. Ross C. Amazon Alexa now HIPAA-compliant, allows secure access to data. Stat. 2019. URL: https://www.statnews.com/2019/04/04/amazon-alexa-hipaa-compliant/ [accessed 2019-04-03]


## Abbreviations

**COPD:** chronic obstructive pulmonary disease
**FDA:** US Food and Drug Administration
**GPDR:** General Data Protection Regulation
**HIPAA:** Health Insurance Portability and Accountability Act
**IT:** informational technology
**NIH:** National Institutes of Health
**OSA:** obstructive sleep apnea
**VUI:** voice user interface

*Edited by L Buis; submitted 10.03.21; peer-reviewed by R Nandakumar, E Sezgin, A Mahnke; comments to author 15.05.21; revised version received 11.06.21; accepted 07.01.22; published 21.02.22.*

*Please cite as:*
*Sunshine J*
*Smart Speakers: The Next Frontier in mHealth*
*JMIR Mhealth Uhealth 2022;10(2):e28686*
*URL: https://mhealth.jmir.org/2022/2/e28686*
*doi:10.2196/28686*
*PMID:35188467*

Review

# The Effects of mHealth-Based Gamification Interventions on Participation in Physical Activity: Systematic Review

Linqi Xu[1,2], MS; Hongyu Shi[1], MS; Meidi Shen[3], MS; Yuanyuan Ni[4], MS; Xin Zhang[1], MS; Yue Pang[1], BS; Tianzhuo Yu[1], BS; Xiaoqian Lian[1], BS; Tianyue Yu[1], BS; Xige Yang[4], MS; Feng Li[1], PhD

[1]School of Nursing, Jilin University, Changchun, China

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

[3]School of Nursing, Peking University, Beijing, China

[4]Department of Anaesthesia, Bethune First Hospital of Jilin University, Changchun, China

**Corresponding Author:**
Feng Li, PhD
School of Nursing
Jilin University
965 Xinjiang Street
Changchun, 130012
China
Phone: 86 1 779 008 9009
Email: fli@jlu.edu.cn

## Abstract

**Background:** It is well known that regular physical exercise has associated benefits; yet, participation remains suboptimal. Mobile health (mHealth) has become an indispensable medium to deliver behavior change interventions, and there is a growing interest in the gamification apps in mHealth to promote physical activity (PA) participation. Gamification could use game design elements (such as points, leaderboards, and progress bars), and it has the potential to increase motivation for PA and engagement. However, mHealth-based gamification interventions are still emerging, and little is known about the application status and efficacy of such interventions.

**Objective:** This systematic review aims to investigate gamification apps in mHealth for improving PA levels and simultaneously summarize the impact of gamification interventions on PA participation.

**Methods:** We searched PubMed, Scopus, Web of Science, Embase, CINAHL (EBSCO host), and IEEE Xplore from inception to December 20, 2020. Original empirical research exploring the effects of gamification interventions on PA participation was included. The papers described at least one outcome regarding exercise or PA participation, which could be subjective self-report or objective indicator measurement. Of note, we excluded studies about serious games or full-fledged games.

**Results:** Of 2944 studies identified from the database search, 50 (1.69%) were included, and the information was synthesized. The review revealed that gamification of PA had been applied to various population groups and broadly distributed among young people but less distributed among older adults and patients with a disease. Most of the studies (30/50, 60%) combined gamification with wearable devices to improve PA behavior change, and 50% (25/50) of the studies used theories or principles for designing gamified PA interventions. The most frequently used game elements were goal-setting, followed by progress bars, rewards, points, and feedback. This review demonstrated that gamification interventions could increase PA participation; however, the results were mixed, and modest changes were attained, which could be attributed to the heterogeneity across studies.

**Conclusions:** Overall, this study provides an overview of the existing empirical research in PA gamification interventions and provides evidence for the efficacy of gamification in enhancing PA participation. High-quality empirical studies are needed in the future to assess the efficacy of a combination of gamification and wearable activity devices to promote PA, and further exploration is needed to investigate the optimal implementation of these features of game elements and theories to enhance PA participation.

***(JMIR Mhealth Uhealth 2022;10(2):e27794)*** doi:10.2196/27794

**KEYWORDS**

mobile health; gamification; physical activity; systematic review; mobile phone

# Introduction

## Background

Regular physical activity (PA) correlates with varied physical and mental health benefits [1-4]. Guidelines reviewed by the Physical Activity Guidelines Advisory Committee recommended that even small increases in light-intensity PA participation can lead to health benefits [5-7]. However, despite proven benefits of PA participation, approximately one-third of the global adult population is insufficiently active and fails to fulfill the minimum PA guideline recommendations [8]. Moreover, an average adult spends approximately 8 hours of the day in sedentary mode [9], resulting in poor health outcomes, including an increased risk of cardiovascular disease and type 2 diabetes [10,11]. Therefore, innovative behavior change interventions are required to improve PA levels.

Mobile health (mHealth), as defined by the American Heart Association's scientific statement, is "the use of mobile computing and communication technologies (eg, mobile phones, wearable devices) for health services and information" [12]. It has become an essential medium to bring about behavior change interventions and has demonstrated a promising role in improving PA levels [13]; for example, wearable activity trackers enable users to objectively monitor their PA levels when used in conjunction with a mobile app. The real-time feedback relating to daily steps from the app may provide ongoing support and motivation for maintaining healthy PA behavior [14].

Gamification is the use of game design elements (such as points, leaderboards, progress bars, and badges) in nongame contexts (such as management, education, marketing, and health care) to increase motivation and engagement [15]. There is a growing interest in the application of gamification in mHealth to promote healthy behavior change [16-19], especially in promoting PA levels [20]. For example, Patel et al [21] used gamification combined with social incentives to reward behaviors and finally increased PA among adults who were overweight and obese. As the concept of gamification is relatively new [15], empirical evidence is still emerging on the efficacy of gamification PA behavior change interventions.

To the best of our knowledge, no systematic review of quantitative studies has assessed the efficacy of gamification on PA behavior change. A systematic review in 2016 examined the amount and quality of empirical evidence for the efficacy of gamification on health and well-being [19]; however, the wide variability in gamification studies was limited in terms of the conclusions that could be drawn. Besides, the use of gamification in behavior change interventions is a young but rapidly growing research field; therefore, it would be timely to conduct a systematic review that combines all the empirical evidence related to the efficacy of gamification on PA participation.

## Aims

This systematic review aims to explore gamification apps in mHealth for improving PA levels and simultaneously summarize the effects of gamification interventions on PA participation. Specifically, this study aims to (1) determine the most commonly used type of mHealth (eg, wearable devices and mobile apps) to deliver PA gamification interventions, (2) describe the most commonly used game elements applied to mHealth for improving PA levels, (3) determine the behavior change theories used in PA gamification interventions, and (4) summarize the impact of gamification interventions on PA outcomes (including daily step counts and time spent in PA) and sedentary behavior.

# Methods

## Operationalizing Gamification

*Gamification* was defined and operationalized as the use of digital game elements in nongame contexts, which needs to be differentiated from creating immersive, full-fledged games as in serious games [15,22]. Serious games, sometimes referred to as *games with a purpose*, provide pure gaming experiences by creating a complete and immersive game (eg, augmented reality exergames such as Pokémon Go), whereas gamification attempts to affect users' behavior and motivation through an experience reminiscent of games using game elements such as badges and points (eg, a wearable device combined with a mobile app used points and leaderboards to promote PA levels). However, the actual difference between the 2 concepts could be vague and highly subjective [22]. In cases where the concepts were indistinguishable, 3 investigators (LX, XY, and FL) discussed the issue and arrived at the final decision.

## Search Strategy

This systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and Cochrane guidelines for systematic reviews [23,24]. Candidate papers were searched in PubMed, Scopus, Web of Science, Embase, CINAHL (EBSCO host), and IEEE Xplore from inception to December 20, 2020. In addition, relevant papers from other systematic reviews were included. The search strategy used controlled vocabulary (Medical Subject Headings), natural language terms, and synonyms. The search keywords were *gamification*, *game element*, and *physical activity*. Multimedia Appendix 1 provides further details on the search strategy.

## Selection Criteria

The search results were imported into EndNote X9 (Clarivate) citation management software after removing the duplicates. All titles and abstracts of the candidate papers were screened by 2 investigators (LX and XZ). After the initial screening, 2 other investigators (MS and YP) independently reviewed the full text of the identified papers. Papers that fulfilled the following criteria were included in the systematic review:

1. Original empirical research, including qualitative and quantitative research (must be experimental research). Reviews (eg, systematic reviews, meta-analyses, narrative reviews, and scoping reviews), design documents, nonexperimental research, and protocols were excluded.
2. Peer-reviewed papers such as published papers, doctoral theses, and conference papers.
3. Full text is available in English.

4. Clearly specify gamification or the use of at least one game element. Research where gamification was only mentioned but not analyzed was excluded.
5. Gamification is delivered through digital devices (eg, PCs, tablets, smartphones, and wearable devices).
6. The purpose of gamification is to promote PA.
7. Serious games and full-fledged games (eg, video games as well as immersive virtual reality games and augmented reality exergames) were excluded.
8. The papers describe at least one outcome regarding exercise or PA participation, which could be subjective self-report or objective indicator measurement.
9. If there was a dispute over a reference, help from a third investigator was sought to resolve the issue and arrive at a final agreement.

### Study Quality

The quality of both the randomized controlled trials (RCTs) and quasi-experimental studies was evaluated by 2 authors (LX and MS). For all studies included in the systematic review, we performed a quality assessment using the Cochrane Effective Practice and Organization of Care Group controlled before-and-after studies risk-of-bias assessment recommendation [25]; this risk-of-bias assessment tool was equally applicable to the quality assessment of RCTs and quasi-experimental studies. A total of 9 risk-of-bias criteria, including selection, performance, and reporting, were used to assess the included studies for potential bias; besides, each criterion was rated as *low risk*, *high risk*, or *unclear risk*. We summarized the quality evaluation results using a diagram. Any disputes were resolved through discussion with a third investigator (Tianzhuo Y) to reach a final agreement.

### Data Extraction and Analysis

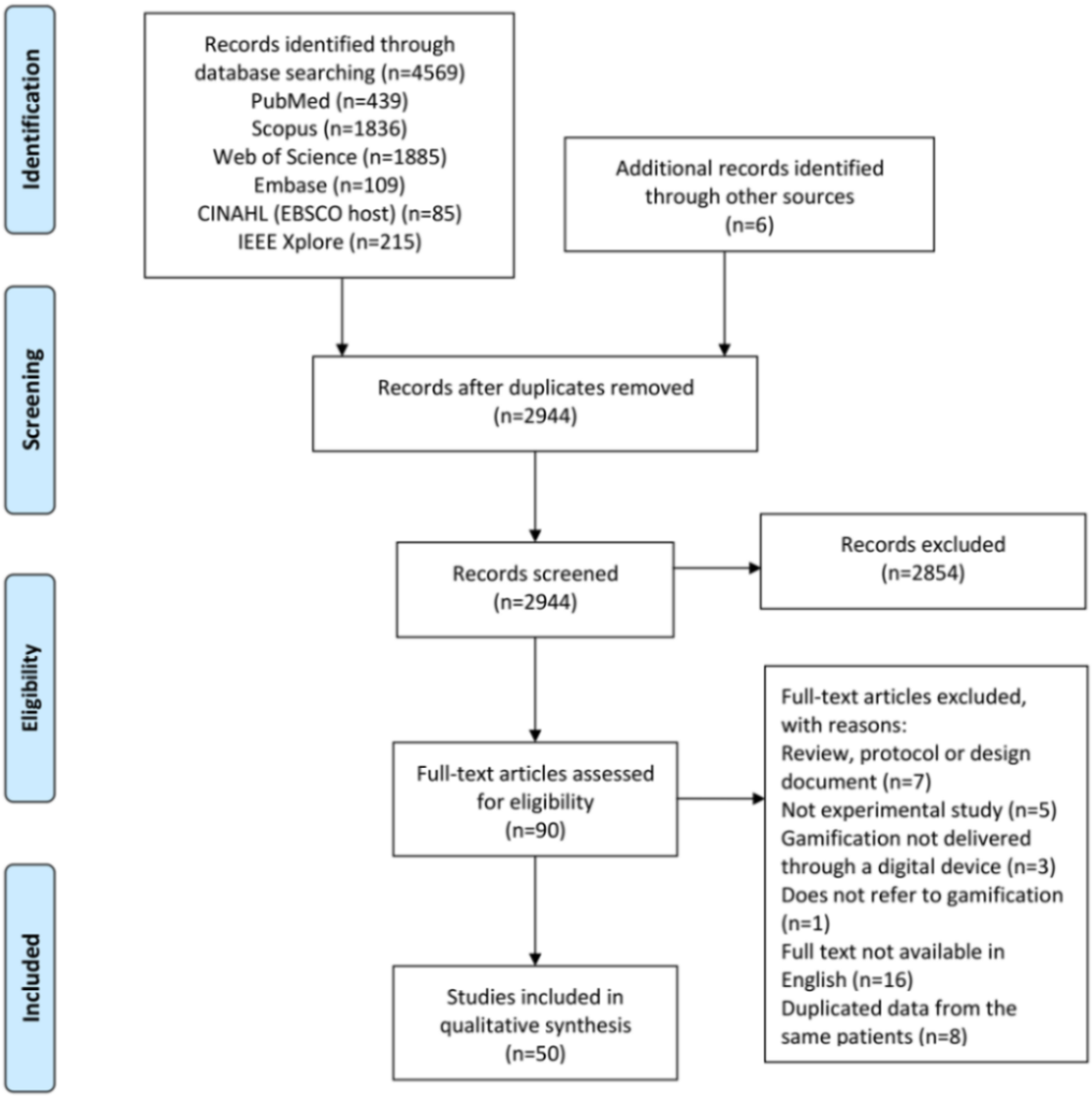
Working independently, 2 investigators (Tianyue Y and XL) extracted information from the selected studies into a prepared Microsoft Access form that was developed specifically for this systematic review. In cases of disagreement, the final decisions were taken after a discussion between the 2 investigators (Tianyue Y and XL). The recorded data in the systematic review included the name of the first author, publication year, country, study design, participant characteristics (population type, mean age, and percentage of the participants who were women), intervention characteristics (sample size, study setting, modality, and duration), gamification characteristics (game name, game elements, and theory used), and PA outcomes (PA measure, domains, and results). For the systematic review, the PA results comprised daily step counts, time spent in light PA (LPA), moderate PA (MPA), vigorous PA (VPA), moderate to vigorous PA (MVPA), percentage of goal reached, and PA motivation. Because of multiple definitions proposed for the term gamification, the subsequent classification methods of game elements were also divided. In this study, we used a combination of the taxonomy of game elements provided by Cugelman [26], Johnson et al [19], Lister et al [17], Sardi et al [16], and Vermeir et al [27]. The studies included in the systematic review had variations in study designs and insufficient data, which did not allow us to perform a meta-analysis. Therefore, we present the analysis of the PA outcomes and sedentary behavior in the form of a narrative review, with the results summarized in a table. Furthermore, we compared the inconsistencies of the intervention and gamification features between positive and negative studies to identify potential explanations.

## Results

### Search Results

A total of 4569 papers were identified through database searching, and an additional 6 papers were identified through other sources. Of these 4575 papers, after removal of duplicates, 2944 (64.35%) were screened by title or abstract. Of these 2944 papers, 2854 (96.94%) were excluded because they did not meet the inclusion and exclusion criteria, leaving 90 (3.06%) for full-text review. After careful evaluation, 44% (40/90) of the papers were excluded for the following reasons: 18% (7/40) were reviews, protocols, or design documents; 13% (5/40) were not experimental studies; 3% (1/40) did not refer to gamification; the gamification of 8% (3/40) was not delivered by means of a digital device; the full texts of 40% (16/40) were not available in English; and 20% (8/40) had duplicate data from the same patients. Finally, of the 90 studies, 50 (56%) were included and evaluated in our systematic review. Figure 1 shows the profile of the study selection.

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




## Study Characteristics

Multimedia Appendix 2 [21,28-76] presents the characteristics of all 50 papers included in our systematic review. The studies were published between 2013 and 2020, and 84% (42/50) were published after 2015, which indicated that research on using gamification to enhance PA was an emerging field and had a rapidly rising trend. The studies were distributed globally: 36% (18/50) in European countries, 24% (12/50) in the United States, 16% (8/50) in Asian countries, 10% (5/50) in Canada, 8% (4/50) in Australia, 4% (2/50) in Brazil, and 2% (1/50) in Singapore. The studies that were selected were primarily from two different types: RCTs (24/50, 48%) and quasi-experimental studies (26/50, 52%). Of the 26 quasi-experimental studies, 7 (27%) used a non-RCT design and 19 (73%) used a single-group pretest–posttest design. Both the RCTs and non-RCTs used a between-group design with 2, 3, 4, and 5 groups.

## Participant Characteristics

The systematic review included a total of 9977 participants, and evaluation was performed. Sample sizes varied from 7 to 3637 participants, with 84% (42/50) of the sample sizes consisting of <200 participants. When reported (45/50, 90%), participant types in 58% (26/45) of the studies were classified as low risk, including healthy adults (10/45, 22%), healthy adolescents (5/45, 11%), children (5/45, 11%), undergraduate students (3/45, 7%), and family (3/45, 7%), whereas participant types in 42% (19/45) of the studies were classified as high risk, including older adults (5/45, 11%); adults who were overweight or obese (4/45, 9%); insufficiently active people (3/45, 7%); and patients with rheumatoid arthritis (1/45, 2%), chronic obstructive pulmonary disease (1/45, 2%), childhood cancer (1/45, 2%), chronic back pain (1/45, 2%), coronary heart disease (1/45, 2%), ovarian cancer (1/45, 2%), and type 2 diabetes (1/45, 2%), indicating that the gamification of PA had been applied to a variety of population groups. The age of the participants ranged from 8 to 71 years, with the gamification interventions broadly distributed among young people but less distributed among older adults and patients with a disease. The proportion of women varied from 0% to 88%; of the 50 studies, 1 (2%) included only male participants and 7 (14%) did not report the gender ratio.

## Intervention Characteristics

Most of the study interventions (35/50, 70%) were conducted on the web, 12% (6/50) at homes, 8% (4/50) at schools, 4% (2/50) at workplaces, 4% (2/50) in communities, and 2% (1/50) in laboratories. The gamification of PA was delivered by means of several digital methods: mobile apps only (14/50, 28%), website only (6/50, 12%), activity monitors (eg, wristband and bracelet) only (7/50, 14%), website combined with activity monitors (9/50, 18%), and mobile apps combined with activity monitors (14/50, 28%), showing that most of the studies (30/50, 60%) combined gamification with wearable devices to improve PA behavior change. To be more specific, most of the wearable devices used in gamification were wrist worn (eg, Fitbit). The duration of the intervention ranged from 72 hours to 2 years; most (38/50, 76%) had no follow-up duration, indicating that further evaluations of PA gamified interventions are required to determine longer-term sustainability in the future.

## Gamification Characteristics

Table 1 and Figure 2 show the gamification characteristics of the studies included in our systematic review. The number of game elements used in PA gamified interventions ranged from 1 to 10, with most including 5 game elements. The most frequently used game elements were goal-setting, followed by progress bars, rewards, points, and feedback.

Of the 50 studies, 25 (50%) used theories or principles for designing gamified PA interventions. As depicted in Table 2, self-determination theory (SDT) was used in 32% (8/25) of the studies, behavioral economics (BE) in 20% (5/25), social cognitive theory in 12% (3/25), theory of planned behavior in 12% (3/25), behavior change technology in 12% (3/25), the transtheoretical model in 12% (3/25), the Whole Person Wellness Model in 4% (1/25), theories of perceived value in 4% (1/25), fun theory in 4% (1/25), sociocognitive learning theory in 4% (1/25), and the health action process approach in 4% (1/25). Furthermore, most of the studies (22/25, 88%) used a single theory and 12% (3/25) used a combination of 2 theories.

**Table 1.** Type of game elements used in the selected studies (N=50).

| Game elements | Values, n (%) |
|---|---|
| **Achievement and progression oriented** | |
| Challenges | 6 (12) |
| Goal-setting | 30 (60) |
| Feedback | 21 (42) |
| Progress bars | 26 (52) |
| Points | 22 (44) |
| Levels | 7 (14) |
| Leaderboards | 12 (24) |
| Badges | 6 (12) |
| Rewards | 25 (50) |
| **Social interaction oriented** | |
| Competition | 16 (32) |
| Collaboration | 16 (32) |
| Social support | 2 (4) |
| **Immersion oriented** | |
| Story or theme | 9 (18) |
| Avatars | 2 (4) |

**Figure 2.** Number of game elements used in the selected studies.

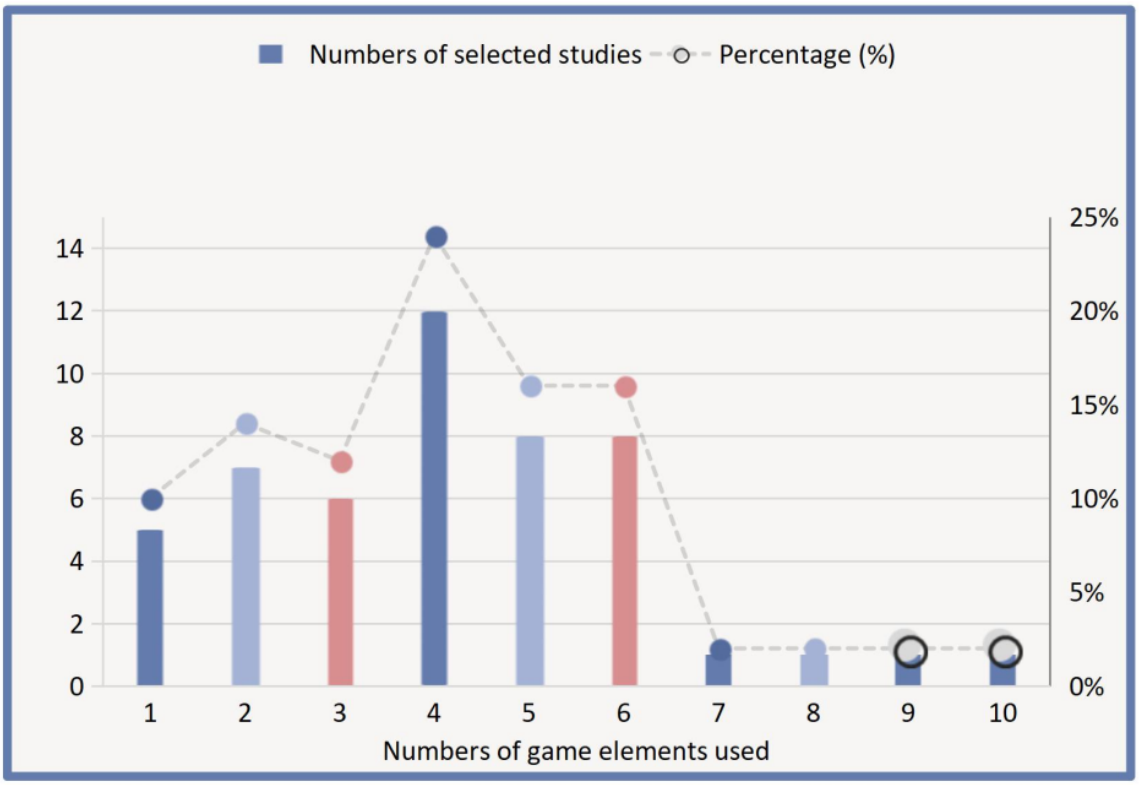



**Table 2.** Number of theories and principles used in the selected studies (N=25).

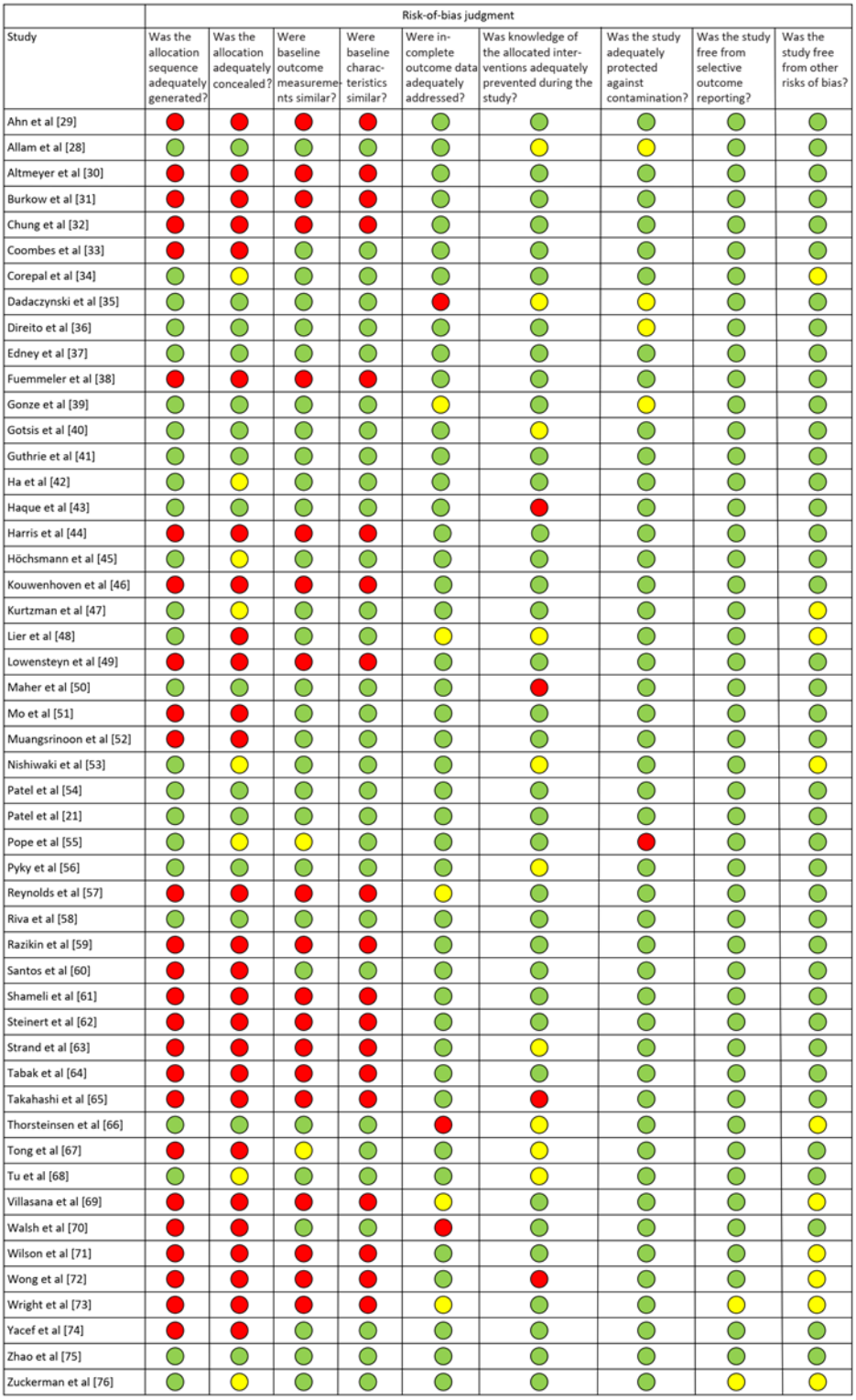
| Theory | Values, n (%) |
|---|---|
| SDT[a] | 8 (32) |
| BE[b] | 5 (20) |
| SCT[c] | 3 (12) |
| TPB[d] | 3 (12) |
| BCT[e] | 3 (12) |
| TTM[f] | 3 (12) |
| WPWM[g] | 1 (4) |
| Theories of perceived value | 1 (4) |
| Fun theory | 1 (4) |
| Sociocognitive learning theory | 1 (4) |
| HAPA[h] | 1 (4) |

[a]SDT: self-determination theory.
[b]BE: behavioral economics.
[c]SCT: social cognitive theory.
[d]TPB: theory of planned behavior.
[e]BCT: behavior change technology.
[f]TTM: transtheoretical model.
[g]WPWM: Whole Person Wellness Model.
[h]HAPA: health action process approach.

## Assessment of Study Quality

As mentioned in Figure 3, the quality of the 50 studies included in the systematic review was summarized using the Cochrane Effective Practice and Organization of Care Group risk-of-bias criteria. Generally, 58% (29/50) of the studies performed well, with at least 6 of the 9 evaluation criteria reported as low risk. As the RCTs and single-group pretest–posttest studies involved random sequence generation and allocation concealment, they were high risk. Furthermore, because 38% (19/50) of the studies had no control group, the applicable criteria relating to between-group comparisons were not fulfilled.

**Figure 3.** Risk-of-bias summary [21,28-76].



### Effects of Gamification on Outcome of PA

The PA behavior domains comprised daily step counts, time spent in LPA, MPA, VPA, and MVPA measured by objective activity monitors (34/50, 68%) or self-reported questionnaires (16/50, 32%). Multimedia Appendix 3 [21,28-76] provides a detailed summary of outcome measures, domains, and results for all included studies. Table 3 includes a summary of selected outcomes by study design. The controlled studies compared the differences between the intervention group and the control group, and the single-group studies simply compared the pre–post data in 1 group. Moreover, we compared the differences in intervention and gamification characteristics between positive and negative studies to identify potential reasons in Multimedia Appendix 4 [21,28-56,58-62,64,66-72,74,76].

**Table 3.** Summary of selected outcomes by study design in the included studies (N=50).

| Outcome and studies that assessed them (randomized controlled trials [study and effect]) | Quasi-experimental studies (study and effect) | |
|---|---|---|
| | Nonrandomized controlled studies | Single-group (pre–post) studies |
| **Step counts (n=23)** | | |
| Corepal et al [34][a] | Coombes et al [33][a] | Ahn et al [29][b] |
| Direito et al [36][a] | Muangsrinoon et al [52][c] | Altmeyer et al [30][b] |
| Gonze et al [39][a] | Santos et al [60][c] | Chung et al [32][d] |
| Höchsmann et al [45][d] | Tong et al [67][c] | Shameli et al [61][b] |
| Kurtzman et al [47][a] | Walsh et al [70][a] | Tabak et al [64][d] |
| Lier et al [48][c] | —[e] | Takahashi et al [65][d] |
| Nishiwaki et al [53][c] | — | Wright et al [73][d] |
| Patel et al [54][c] | — | — |
| Patel et al [21][c] | — | — |
| Pope et al [55][a] | — | — |
| Tu et al [68][c] | — | — |
| **Time spent in overall PA[f] (n=15)** | | |
| Allam et al [28][b] | Mo et al [51][c] | Altmeyer et al [30][b] |
| Gotsis et al [40][a] | — | Burkow et al [31][d] |
| Haque et al [43][c] | — | Harris [44][b] |
| Maher et al [50][a] | — | Lowensteyn et al [49][b] |
| Nishiwaki et al [53][c] | — | Razikin et al [59][b] |
| Riva et al [58][a] | — | Steinert et al [62][b] |
| Thorsteinsen et al [66][a] | — | Villasana et al [69][a] |
| — | — | Wong et al [72][b] |
| **Time spent in LPA[g] (n=7)** | | |
| Corepal et al [34][a] | Mo et al [51][c] | — |
| Dadaczynski et al [35][c] | Yacef et al [74][a] | — |
| Direito et al [36][a] | — | — |
| Maher et al [50][c] | — | — |
| Zuckerman et al [76][c] | — | — |
| **Time spent in MPA[h] (n=6)** | | |
| Corepal et al [34][a] | Mo et al [51][c] | — |
| Dadaczynski et al [35][a] | Yacef et al [74][c] | — |
| Direito et al [36][a] | — | — |
| Maher et al [50][a] | — | — |
| **Time spent in VPA[i] (n=6)** | | |
| Corepal et al [34][a] | Mo et al [51][c] | — |

| Outcome and studies that assessed them (randomized controlled trials [study and effect]) | Quasi-experimental studies (study and effect) | |
|---|---|---|
| | Nonrandomized controlled studies | Single-group (pre–post) studies |
| Dadaczynski et al [35][a] | Yacef et al [74][c] | — |
| Direito et al [36][a] | — | — |
| Maher et al [50][a] | — | — |
| **Time spent in MVPA[j] (n=9)** | | |
| Corepal et al [34][a] | Coombes et al [33][c] | Fuemmeler et al [38][b] |
| Direito et al [36][a] | — | Kouwenhoven-Pasmooij et al [46][b] |
| Edney et al [37][a] | — | Wilson et al [71][a] |
| Guthrie et al [41][c] | — | — |
| Ha et al [42][c] | — | — |
| **Sedentary behavior (n=4)** | | |
| Direito et al [36][c] | Yacef et al [74][a] | Fuemmeler et al [38][b] |
| Pyky et al [56][a] | — | — |
| **Percentage of goal reached (n=3)** | | |
| Patel et al [54][c] | — | — |
| Patel et al [21][c] | — | — |
| Zuckerman et al [76][c] | — | — |
| **PA motivation (n=3)** | | |
| Zhao et al [75][c] | — | Reynolds et al [57][d] |
| — | — | Strand et al [63][b] |

[a]The between-group difference or the pre–post difference is not significant.
[b]The pre–post difference between groups is statistically significant.
[c]The difference between the intervention and control groups is statistically significant.
[d]There is a trend toward improvement, but the improvement is not significant.
[e]Not available.
[f]PA: physical activity.
[g]LPA: light physical activity.
[h]MPA: moderate physical activity.
[i]VPA: vigorous physical activity.
[j]MVPA: moderate to vigorous physical activity.

## Step Counts

Of the 50 included studies, 23 (46%) assessed the impact of PA gamification interventions on step counts. Of these 23 studies, 11 (48%) were RCTs, 5 (22%) were non-RCTs, and 7 (30%) were single-group studies. As depicted in Table 3, the results were quite consistent between the controlled studies and the single-group studies. The controlled studies (16/23, 65%) reported mixed results; 50% (8/16) [21,48,52-54,60,67,68] reported that the gamification interventions exerted a positive impact on step counts, 44% (7/16) [33,34,36,39,47,55,70] reported that no difference existed between the intervention and control groups for step counts, and 6% (1/16) [45] suggested a trend toward an increase in step counts after the gamification interventions, although the difference was not significant. The single-group studies (7/23, 30%) also reported mixed results; 43% (3/7) [29,30,61] reported that the pre–post difference within groups was statistically significant for step counts, whereas 57% (4/7) [32,64,65,73] reported that the pre–post difference was not significant.

## Time Spent in PA

### *Overview*

Of the 50 included studies, 8 (16%) controlled studies and 8 (16%) single-group studies assessed the time spent in PA, as shown in Table 3, and the results were quite different between

the controlled studies and the single-group studies. In the controlled studies, only 3 (3/8, 38%) [43,51,53] reported that the difference between the intervention and control groups was statistically significant. However, for the single-group studies, most of the studies (6/8, 75%) [30,44,49,59,62,72] demonstrated that the time spent in PA significantly increased after the gamification intervention. Only the study by Villasana et al [69] reported no trend toward improvement; the pre–post difference was not significant after the gamification intervention, and the study used just 1 game element (challenge) and did not use any theory (Multimedia Appendix 4).

We further examined the impact of gamification interventions on LPA, MPA, VPA, and MVPA.

#### *Impact on LPA*

Among the 50 included studies, time spent in LPA was assessed in 5 (10%) RCTs [34-36,50,76] and 2 (4%) non-RCTs [51,74] with mixed results. Of the 5 RCTs, 3 (60%) [35,50,76] showed that compared with the control groups, the intervention groups spent more time in LPA; however, the other 2 (40%) RCTs [34,36] reported that the differences between the intervention and control groups were not significant. In the non-RCTs, the study by Mo et al [51] reported that the gamification intervention exerted a positive impact on LPA, whereas the study by Yacef et al [74] reported no significant difference between the intervention and control groups. After comparing these 2 studies, we found that applying multiple and integrated gamification elements (>2 game elements) could be associated with positive effects on LPA.

#### *Impact on MPA*

Of the 50 included studies, 4 (8%) RCTs [34-36,50] and 2 (4%) non-RCTs [51,74] measured the time spent in MPA; the 4 (100%) RCTs [34-36,50] reported that the differences between the intervention and control groups were not significant, whereas the 2 (100%) non-RCTs [51,74] showed significant effects. The difference in the results between the RCTs and the non-RCTs could be attributed to the selection bias in the non-RCTs.

#### *Impact on VPA*

Among the 50 included studies, the outcomes of VPA were reported in 4 (8%) RCTs [34-36,50] and 2 (4%) non-RCTs [51,74]; of note, the results were different between these 2 types of studies. The RCTs [34-36,50] reported that no difference existed between the intervention and control groups for VPA; however, the non-RCTs [51,74] reported that the VPA was significantly increased in the intervention group compared with the control group.

#### *Impact on MVPA*

Of the 50 included studies, 9 (18%) studies reported the time spent in MVPA. Of these 9 studies, 6 (67%) were controlled studies [33,34,36,37,41,42] and 3 (33%) were single-group studies [38,46,71]; the results in both were mixed. In the 6 controlled studies, 3 (50%) [33,41,42] reported that the gamification intervention had positive effects on MVPA, whereas 3 (50%) [34,36,37] reported no significant difference between the intervention and control groups. In the 3 single-group studies, the pre–post difference between the groups for time spent in MVPA was significant in 2 (67%) studies [38,46] but not in the study by Wilson et al [71].

### Effects of Gamification on Sedentary Behavior

Sedentary behavior was reported as daily sitting time. Of the 50 included studies, 2 (4%) RCTs [36,56], 1 (2%) non-RCT [74], and 1 (2%) single-group study [38] reported this outcome; the results of the controlled studies were mixed, but the single-group study reported that the gamification intervention exerted a positive impact on sedentary behavior. In the 3 controlled studies, 1 (33%) RCT [36] reported that the intervention group spent less time in sitting compared with the control group, whereas the 2 (67%) other studies [56,74] reported no statistically significant differences between the intervention and control groups for daily sitting time. However, the single-group (pre–post) study [38] reported a significant decrease after the gamification intervention.

## Discussion

### Principal Findings

This study aims to offer a review of the gamification of PA. A total of 50 studies were included in the systematic review, suggesting that gamification in PA was still developing and lacked high-quality empirical research that could validate the efficacy of such interventions. The review revealed that gamification of PA had been applied to a variety of population groups and broadly distributed among young people but less distributed among older adults and patients with a disease. Most of the studies (30/50, 60%) combined gamification with wearable devices to improve PA behavior change. The most frequently used game elements were goal-setting, followed by progress bars, rewards, points, and feedback; besides, the most used theory in PA gamification was SDT. This systematic review revealed mixed findings for the efficacy of gamification interventions for improving PA participation and sedentary behavior. Both controlled studies and single-group studies reported mixed results on step counts, MVPA, and sedentary behavior. In addition, the controlled studies reported mixed results on time spent in LPA, MPA, and VPA. However, most of the single-group studies (6/8, 75%) revealed that gamified interventions might positively affect time spent in overall PA. Of note, these findings were limited because of the small number of studies.

### Gamification and mHealth

In the systematic review, the types of mHealth technologies used for delivering PA gamification interventions varied, with most of the studies using activity monitors (30/50, 60%), followed by mobile apps (28/50, 56%). To be more specific, most of the wearable devices used were wrist worn (eg, Fitbit). There is a growing interest in the use of wearable activity trackers to facilitate behavior management, when combined with the use of mobile apps; they might enhance users' motivation for PA and help to better manage their health [77,78]. Wearable activity trackers could provide real-time feedback related to daily steps and energy expenditure by means of specifically designed algorithms or through health professionals [79,80], and when combined with gamification, they may

markedly help in improving PA motivation and participation. However, there are few high-quality empirical studies. Thus, more empirical research is required in the future to explore the efficacy of a combination of gamification and wearable activity devices in promoting PA.

### Game Elements Used in PA Gamification

In the systematic review, the most frequently used game elements were achievement and progress oriented, such as goal-setting, progress bars, rewards, points, and feedback, which is consistent with previous reviews [26,27], suggesting that these were also the most frequently used elements in PA gamification interventions. Goal-setting (30/50, 60%) is a key technique for behavior change [26], and when it is combined with progress and feedback, it could markedly facilitate intrinsic motivation [81]. However, few scholars believe that rewards promote extrinsic motivation compared with intrinsic motivation; therefore, there may be a poor maintenance effect of the interventions [82].

The second most frequently used game elements in PA gamification interventions were social interaction oriented, such as competition and collaboration; these 2 elements increase users' experience of fun and promote motivation for PA participation through social incentives. However, studies have demonstrated that different types and applications of social incentives might affect the efficacy of gamification interventions [21]. For example, gamification with collaboration among families led to significant increases in PA; however, the intervention was ineffective when conducted with participants who were previously unknown to each other [21,54]. Among such participants, competition became a more effective incentive method to promote PA. Therefore, future research needs to investigate the efficacy of gamification combined with different types of social incentives to promote PA participation.

### Gamification and Behavior Change Theories

In the systematic review, half of the studies used theories or principles for designing gamified PA interventions, and SDT (8/25, 32%) was the most commonly used theory, followed by BE (5/25, 20%). These findings were consistent with a previous systematic review [19]. SDT is a well-established motivation theory that has become a key framework for health behavior interventions because the motivation of individuals was recognized as the main factor driving behavior change [83]. However, intrinsic motivation or extrinsic motivation has different effects on behavior change, and existing research reveals that intrinsic motivation can promote not only behavior change in a more stable manner but also psychological and social well-being [19]. Hence, future research could consider applying gamification to promote intrinsic motivation to aid in improving PA participation.

The second most commonly used theory in PA gamification interventions was BE. In recent years, there has been a trend to use BE principles to guide interventions for improving PA [84]. From the perspective of BE principles, the decision to participate in PA is considered an investment in future health. An individual who is willing to *pay* the immediate costs of PA (eg, time and energy expenditure) to obtain health benefits in the future is regarded as having patient time preferences. We identified some predictable decision biases and chose interventions that persuade patients to choose a healthier decision (eg, participating in PA). Common BE principles embedded within PA gamification interventions included loss aversion, regret aversion, precommitment, and social norms [21,54].

### Effects of Gamification on PA and Sedentary Behavior

Overall, the evidence regarding the use of gamification to facilitate PA participation was inconclusive. Therefore, it is essential to consider potential explanations for the inconsistencies between the positive and negative studies. Regarding the time spent in overall PA, the positive impact of gamified interventions on PA was observed in 75% (6/8) of the single-group studies; these findings were consistent with a previous published systematic review [19], which reported that the positive impact of gamified interventions on PA was observed in 80% (8/10) of the studies. We further compared the differences in intervention and gamification characteristics between positive and negative studies. Of the 8 single-group studies, only 1 (13%) showed no trend toward improvement, and the pre–post difference was not significant in terms of the time spent in overall PA after the gamification intervention; the study used just 1 game element (challenge) and did not use any theory. These findings revealed that a combination of multiple game elements could be more effective for PA participation than a single game element, and gamification intervention using theory guidance could be more effective than a gamification intervention without any theory guidance. Furthermore, we tried to identify the appealing game features that could be associated with a positive effect; however, it is difficult to draw a definite conclusion because many studies have applied ≥2 gamification elements, and we cannot separate them to make a judgment. In addition, some of the studies [9,39] reported that participants liked the self-monitoring of progress and leader board aspects, which might be associated with the positive effect on PA outcomes. However, this should be interpreted with caution because of the heterogeneity of the selected studies.

Regarding the time spent in MPA and VPA, of the 50 included studies, 6 (12%) controlled studies measured the time spent in MPA and VPA and reported mixed results; the results differed between RCTs and non-RCTs. The bias in the non-RCTs could have potentially led to positive results. Our study reported mixed effects of gamification on daily sitting time. As far as we know, this is the first systematic review to report the impact of gamification on sedentary behavior; however, the results were limited because there were only a few high-quality empirical studies.

### Limitations

Our study includes several limitations. First, because of the variability and heterogeneity of the research interventions and results, the evidence might not be sufficiently strong to determine whether gamification effectively improves PA participation. Second, the studies included in the systematic review had variations in study design and insufficient data, which did not allow a meta-analysis. Third, although the population was diverse, the original articles had insufficient data, which prevented us from conducting a subgroup analysis

based on the population. Fourth, related outcomes were measured immediately after the end of the intervention period, and the long-term effects of gamification in most studies were not observed; therefore, we did not summarize and synthesize the maintenance effect of the gamification interventions. Fifth, the differences in game elements, mHealth technology types, populations, and sample sizes among the included studies might be a major cause of the heterogeneity. Finally, most selected studies in our review were conducted using medical registry databases, which might suffer from an intrinsic risk of coding imprecision and incompleteness.

### Conclusions and Practical Implications

This study demonstrates that gamification interventions can increase PA participation; however, the results were mixed, and modest changes were obtained. This could be attributed to the heterogeneity across studies. Gamification combined with wearable activity trackers could help individuals to self-monitor progress and provide fun and motivation to promote health-related behavior change, especially in improving PA. Therefore, high-quality empirical studies are required in the future to examine the efficacy of a combination of gamification and wearable activity devices to promote PA. Gamification interventions generally have short-term effects, and ongoing contact by means of specifically designed algorithms and through health professionals could increase long-term adherence to PA participation. Hence, gamification combined with wearable activity devices has the potential to assist health professionals to provide ongoing support and motivation to patients who are physically inactive in terms of adherence to PA participation. Moreover, this study reveals that a combination of multiple game elements could be more effective for PA participation than a single game element, and a gamification intervention using theory guidance could be more effective than a gamification intervention without any theory guidance. The combination of different theories and different multiple game elements might produce different effects; hence, further exploration is required to explore the optimal implementation of these features of game elements and theories to improve PA participation. Furthermore, future empirical research on gamification should focus not only on the outcome of PA but also on the impact on sedentary behavior.

## Authors' Contributions

LX and FL contributed to the systematic review’s conception and research question. LX, HS, and MS contributed to the database search. LX, XZ, MS, and YP selected the included studies. LX, Tianyue Y, and XL extracted the data from the selected studies. LX and XY contributed to statistical analysis and writing. All authors gave final approval and agree to be accountable for all aspects of the work to ensure integrity and accuracy.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 15 KB - mhealth_v10i2e27794_app1.docx ]

Multimedia Appendix 2
Summary descriptions of the studies included in the systematic review.
[DOCX File , 36 KB - mhealth_v10i2e27794_app2.docx ]

Multimedia Appendix 3
Summary of outcomes in the selected studies.
[DOCX File , 30 KB - mhealth_v10i2e27794_app3.docx ]

Multimedia Appendix 4
Comparing the differences of intervention and gamification characteristics between positive and negative studies.
[DOCX File , 23 KB - mhealth_v10i2e27794_app4.docx ]

## References

1. Writing GM, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, American Heart Association Statistics Committee, Stroke Statistics Subcommittee. Executive summary: heart disease and stroke statistics--2016 update: a report from the American Heart Association. Circulation 2016 Jan 26;133(4):447-454. [doi: 10.1161/CIR.0000000000000366] [Medline: 26811276]
2. Yang Q, Cogswell ME, Flanders WD, Hong Y, Zhang Z, Loustalot F, et al. Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. J Am Med Assoc 2012 Mar 28;307(12):1273-1283. [doi: 10.1001/jama.2012.339] [Medline: 22427615]

3. 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]
4. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
5. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
6. LaCroix AZ, Bellettiere J, Rillamas-Sun E, Di C, Evenson KR, Lewis CE, Women's Health Initiative (WHI). Association of light physical activity measured by accelerometry and incidence of coronary heart disease and cardiovascular disease in older women. JAMA Netw Open 2019 Mar 01;2(3):e190419 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0419] [Medline: 30874775]
7. Spartano NL, Davis-Plourde KL, Himali JJ, Andersson C, Pase MP, Maillard P, et al. Association of accelerometer-measured light-intensity physical activity with brain volume: the framingham heart study. JAMA Netw Open 2019 Apr;2(4):e192745. [doi: 10.1001/jamanetworkopen.2019.2745] [Medline: 31002329]
8. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
9. Schuna JM, Johnson WD, Tudor-Locke C. Adult self-reported and objectively monitored physical activity and sedentary behavior: NHANES 2005-2006. Int J Behav Nutr Phys Act 2013;10:126 [FREE Full text] [doi: 10.1186/1479-5868-10-126] [Medline: 24215625]
10. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
11. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, Lancet Physical Activity Series 2 Executive Committe, Lancet Sedentary Behaviour Working Group. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet 2016 Sep 24;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 27475271]
12. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
13. McConnell MV, Turakhia MP, Harrington RA, King AC, Ashley EA. Mobile health advances in physical activity, fitness, and atrial fibrillation: moving hearts. J Am Coll Cardiol 2018 Jun 12;71(23):2691-2701. [doi: 10.1016/j.jacc.2018.04.030] [Medline: 29880130]
14. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
15. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; September 28 - 30, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
16. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Dec;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
17. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
18. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
19. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
20. No authors listed. Gamification for physical activity behaviour change. Perspect Public Health 2018 Nov;138(6):309-310. [doi: 10.1177/1757913918801447] [Medline: 30412020]
21. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CA, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the united states: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Sep 09:1-9 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]

22. Johnson D, Horton E, Mulcahy R, Foth M. Gamification and serious games within the domain of domestic energy consumption: a systematic review. Renew Sust Energ Rev 2017 Jun;73:249-264. [doi: 10.1016/j.rser.2017.01.134]
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
25. EPOC Resources for review authors. Cochrane Effective Practice and Organisation of Care (EPOC). 2017. URL: https://epoc.cochrane.org/resources/epoc-resources-review-authors [accessed 2021-01-25]
26. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
27. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644 [FREE Full text] [doi: 10.2196/18644] [Medline: 32773374]
28. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]
29. Ahn SJ, Johnsen K, Ball C. Points-based reward systems in gamification impact children's physical activity strategies and psychological needs. Health Educ Behav 2019 Jun 24;46(3):417-425 [FREE Full text] [doi: 10.1177/1090198118818241] [Medline: 30678507]
30. Altmeyer M, Lessel P, Sander T, Krüger A. Extending a gamified mobile app with a public display to encourage walking. In: Proceedings of the 22nd International Academic Mindtrek Conference. 2018 Presented at: Mindtrek 2018: Academic Mindtrek 2018; October 10 - 11, 2018; Tampere Finland p. 20-29. [doi: 10.1145/3275116.3275135]
31. Burkow TM, Vognild LK, Johnsen E, Bratvold A, Risberg MJ. Promoting exercise training and physical activity in daily life: a feasibility study of a virtual group intervention for behaviour change in COPD. BMC Med Inform Decis Mak 2018 Dec 18;18(1):136 [FREE Full text] [doi: 10.1186/s12911-018-0721-8] [Medline: 30563507]
32. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to health: a novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2016 Jun 16:26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
33. Coombes E, Jones A. Gamification of active travel to school: a pilot evaluation of the Beat the Street physical activity intervention. Health Place 2016 May;39:62-69 [FREE Full text] [doi: 10.1016/j.healthplace.2016.03.001] [Medline: 26974232]
34. Corepal R, Best P, O'Neill R, Kee F, Badham J, Dunne L, et al. A feasibility study of 'The StepSmart Challenge' to promote physical activity in adolescents. Pilot Feasibility Stud 2019 Nov;5:132 [FREE Full text] [doi: 10.1186/s40814-019-0523-5] [Medline: 31832227]
35. Dadaczynski K, Schiemann S, Backhaus O. Promoting physical activity in worksite settings: results of a German pilot study of the online intervention Healingo fit. BMC Public Health 2017 Sep 08;17(1):696 [FREE Full text] [doi: 10.1186/s12889-017-4697-6] [Medline: 28886734]
36. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
37. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):51-62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
38. Fuemmeler BF, Holzwarth E, Sheng Y, Do EK, Miller CA, Blatt J, et al. Mila Blooms: a mobile phone application and behavioral intervention for promoting physical activity and a healthy diet among adolescent survivors of childhood cancer. Games Health J 2020 Aug;9(4):279-289. [doi: 10.1089/g4h.2019.0060] [Medline: 32391734]
39. Gonze BD, Padovani RD, Simoes MD, Lauria V, Proença NL, Sperandio EF, et al. Use of a smartphone app to increase physical activity levels in insufficiently active adults: feasibility sequential multiple assignment randomized trial (SMART). JMIR Res Protoc 2020 Oct 23;9(10):e14322 [FREE Full text] [doi: 10.2196/14322] [Medline: 33094733]
40. Gotsis M, Wang H, Spruijt-Metz D, Jordan-Marsh M, Valente TW. Wellness partners: design and evaluation of a web-based physical activity diary with social gaming features for adults. JMIR Res Protoc 2013 Feb 01;2(1):e10 [FREE Full text] [doi: 10.2196/resprot.2132] [Medline: 23611986]
41. Guthrie N, Bradlyn A, Thompson SK, Yen S, Haritatos J, Dillon F, et al. Development of an accelerometer-linked online intervention system to promote physical activity in adolescents. PLoS One 2015;10(5):e0128639 [FREE Full text] [doi: 10.1371/journal.pone.0128639] [Medline: 26010359]
42. Ha AS, Lonsdale C, Lubans DR, Ng JY. Increasing students' activity in physical education: results of the self-determined exercise and learning for FITness trial. Med Sci Sports Exerc 2020 Mar;52(3):696-704. [doi: 10.1249/MSS.0000000000002172] [Medline: 31652232]

43. Haque MS, Kangas M, Jämsä T. A persuasive mhealth behavioral change intervention for promoting physical activity in the workplace: feasibility randomized controlled trial. JMIR Form Res 2020 May 04;4(5):e15083 [FREE Full text] [doi: 10.2196/15083] [Medline: 32364506]
44. Harris MA. The relationship between physical inactivity and mental wellbeing: findings from a gamification-based community-wide physical activity intervention. Health Psychol Open 2018 Jan;5(1):2055102917753853 [FREE Full text] [doi: 10.1177/2055102917753853] [Medline: 29372067]
45. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
46. Kouwenhoven-Pasmooij TA, Robroek SJ, Ling SW, van Rosmalen J, van Rossum EF, Burdorf A, et al. A blended web-based gaming intervention on changes in physical activity for overweight and obese employees: influence and usage in an experimental pilot study. JMIR Serious Games 2017 Apr 03;5(2):e6 [FREE Full text] [doi: 10.2196/games.6421] [Medline: 28373157]
47. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675. [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
48. Lier LM, Breuer C. The motivating power of gamification: Does the inclusion of game elements increase the effectiveness of worksite health promotion programs? Int J Workplace Health Manag 2019 Dec 23;13(1):1-15. [doi: 10.1108/ijwhm-04-2019-0055]
49. Lowensteyn I, Berberian V, Berger C, Da Costa D, Joseph L, Grover SA. The sustainability of a workplace wellness program that incorporates gamification principles: participant engagement and health benefits after 2 years. Am J Health Promot 2019 Jul;33(6):850-858. [doi: 10.1177/0890117118823165] [Medline: 30665309]
50. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
51. Mo D, Xiang M, Luo M, Dong Y, Fang Y, Zhang S, et al. Using gamification and social incentives to increase physical activity and related social cognition among undergraduate students in Shanghai, China. Int J Environ Res Public Health 2019 Mar 08;16(5):858 [FREE Full text] [doi: 10.3390/ijerph16050858] [Medline: 30857261]
52. Muangsrinoon S, Boonbrahm P. Game elements to promote walking in Thais working adults. In: Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). 2019 Presented at: 16th International Joint Conference on Computer Science and Software Engineering (JCSSE); July 10-12, 2019; Chonburi, Thailand. [doi: 10.1109/jcsse.2019.8864223]
53. Nishiwaki M, Kuriyama A, Ikegami Y, Nakashima N, Matsumoto N. A pilot crossover study: effects of an intervention using an activity monitor with computerized game functions on physical activity and body composition. J Physiol Anthropol 2014 Dec 02;33:35 [FREE Full text] [doi: 10.1186/1880-6805-33-35] [Medline: 25466450]
54. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
55. Pope L, Garnett B, Dibble M. Lessons learned through the implementation of an ehealth physical activity gaming intervention with high school youth. Games Health J 2018 Apr;7(2):136-142. [doi: 10.1089/g4h.2017.0164] [Medline: 29393679]
56. Pyky R, Koivumaa-Honkanen H, Leinonen A, Ahola R, Hirvonen N, Enwald H, et al. Effect of tailored, gamified, mobile physical activity intervention on life satisfaction and self-rated health in young adolescent men: a population-based, randomized controlled trial (MOPO study). Comput Hum Behav 2017 Jul;72:13-22. [doi: 10.1016/j.chb.2017.02.032]
57. Reynolds L, Sosik V, Cosley D. When Wii doesn't Fit: how non-beginners react to Wii Fit's gamification. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications. 2013 Presented at: Gamification '13: Gameful Design, Research, and Applications; October 2 - 4, 2013; Toronto Ontario Canada p. 111-114. [doi: 10.1145/2583008.2583027]
58. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
59. Razikin K, Goh D, Lee C. Examining enjoyment in gamifying physical exercise and a healthy diet. In: Human-Computer Interaction. Interaction Contexts. Cham: Springer; 2017:84-98.
60. Santos LH, Okamoto K, Funghetto SS, Cavalli AS, Hiragi S, Yamamoto G, et al. Effects of social interaction mechanics in pervasive games on the physical activity levels of older adults: quasi-experimental study. JMIR Serious Games 2019 Jul 22;7(3):e13962 [FREE Full text] [doi: 10.2196/13962] [Medline: 31333202]
61. Shameli A, Althoff T, Saberi A, Leskovec J. How gamification affects physical activity: large-scale analysis of walking challenges in a mobile application. Proc Int World Wide Web Conf 2017 Apr;2017:455-463 [FREE Full text] [doi: 10.1145/3041021.3054172] [Medline: 28990011]

62. Steinert A, Buchem I, Merceron A, Kreutel J, Haesner M. A wearable-enhanced fitness program for older adults, combining fitness trackers and gamification elements: the pilot study fMOOC@Home. Sport Sci Health 2018 Jan 10;14(2):275-282. [doi: 10.1007/s11332-017-0424-z]
63. Strand KA, Francis SL, Margrett JA, Franke WD, Peterson MJ. Community-based exergaming program increases physical activity and perceived wellness in older adults. J Aging nd Phys Act 2014 Jul;22(3):364-371. [doi: 10.1123/japa.22.3.364]
64. Tabak M, de Vette F, van Dijk H, Vollenbroek-Hutten M. A game-based, physical activity coaching application for older adults: design approach and user experience in daily life. Games Health J 2020 Jun;9(3):215-226. [doi: 10.1089/g4h.2018.0163] [Medline: 32053023]
65. Takahashi M, Kawasaki H, Maeda A, Nakamura M. Mobile walking game and group-walking program to enhance going out for older adults. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12 - 16, 2016; Heidelberg Germany p. 1372-1380. [doi: 10.1145/2968219.2968415]
66. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014:746232 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]
67. Tong X, Gromala D, Shaw CD. A field study: evaluating gamification approaches for promoting physical activity with motivational models of behavior changes. In: Human-Computer Interaction. Novel User Experiences. Cham: Springer; 2016:417-424.
68. Tu R, Hsieh P, Feng W. Walking for fun or for “likes”? The impacts of different gamification orientations of fitness apps on consumers’ physical activities. Sport Manag Rev 2021 Feb 03;22(5):682-693. [doi: 10.1016/j.smr.2018.10.005]
69. Villasana MV, Pires IM, Sá J, Garcia NM, Teixeira MC, Zdravevski E, et al. Promotion of healthy lifestyles to teenagers with mobile devices: a case study in Portugal. Healthcare (Basel) 2020 Sep 02;8(3):315 [FREE Full text] [doi: 10.3390/healthcare8030315] [Medline: 32887251]
70. Walsh G, Golbeck J. StepCity: a preliminary investigation of a personal informatics-based social game on behavior change. In: Proceedings of the Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April 26 - May 1, 2014; Toronto Ontario Canada p. 2371-2376. [doi: 10.1145/2559206.2581326]
71. Wilson D, Gillett C, Piele G, Williams C, Stuart A. Does a wrist-worn activity monitor increase exercise activity in teenagers with congenital heart disease? University Hospitals Bristol NHS Healthcare Trust. In: Proceedings of the EuroPRevent 2016. 2016 Presented at: EuroPRevent 2016; May 5-7, 2016; Istanbul p. S44. [doi: 10.1177/2047487316668095]
72. Wong RS, Yu EY, Wong TW, Fung CS, Choi CS, Or CK, et al. Development and pilot evaluation of a mobile app on parent-child exercises to improve physical activity and psychosocial outcomes of Hong Kong Chinese children. BMC Public Health 2020 Oct 14;20(1):1544 [FREE Full text] [doi: 10.1186/s12889-020-09655-9] [Medline: 33054753]
73. Wright A, Schrier E, Thompson E, Schumer S, Liu J, Krasner C, et al. Stepping into survivorship: harnessing mobile health and principles of behavioral economics to increase physical activity in ovarian cancer survivors. Gynecol Oncol 2020 Oct;159:322. [doi: 10.1016/j.ygyno.2020.05.585]
74. Yacef K, Caillaud C, Galy O. Supporting learning activities with wearable devices to develop life-long skills in a health education app. In: Artificial Intelligence in Education. Cham: Springer; 2018:394-398.
75. Zhao Z, Arya A, Orji R, Chan G. Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study. JMIR Serious Games 2020 Nov 17;8(4):e19968 [FREE Full text] [doi: 10.2196/19968] [Medline: 33200994]
76. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014 Jul 5;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
77. Mizuno A, Changolkar S, Patel MS. Wearable devices to monitor and reduce the risk of cardiovascular disease: evidence and opportunities. Annu Rev Med 2021 Jan 27;72:459-471. [doi: 10.1146/annurev-med-050919-031534] [Medline: 32886543]
78. Xie Z, Jo A, Hong YR. Electronic wearable device and physical activity among US adults: an analysis of 2019 HINTS data. Int J Med Inform 2020 Dec;144(8):104297. [doi: 10.1016/j.ijmedinf.2020.104297] [Medline: 33059242]
79. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older adults' acceptance of activity trackers. J Appl Gerontol 2017 Feb;36(2):127-155 [FREE Full text] [doi: 10.1177/0733464815624151] [Medline: 26753803]
80. Brickwood K, 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 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
81. Hicks K, Gerling K, Richardson G, Pike T, Burman O, Dickinson P. Understanding the effects of gamification and juiciness on players. In: Proceedings of the IEEE Conference on Games (CoG). 2019 Presented at: IEEE Conference on Games (CoG); Aug. 20-23, 2019; London, UK. [doi: 10.1109/cig.2019.8848105]
82. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]

83. Deci EL, Ryan RM. Motivation, personality, and development within embedded social contexts: an overview of self-determination theory. In: The Oxford Handbook of Human Motivation (1 ed.). Oxfordshire, United Kingdom: Oxford University Press; 2012:85-107.
84. Shuval K, Leonard T, Drope J, Katz DL, Patel AV, Maitin-Shepard M, et al. Physical activity counseling in primary care: insights from public health and behavioral economics. CA Cancer J Clin 2017 May 06;67(3):233-244 [FREE Full text] [doi: 10.3322/caac.21394] [Medline: 28198998]

## Abbreviations

**BE:** behavioral economics
**LPA:** light physical activity
**mHealth:** mobile health
**MPA:** moderate physical activity
**MVPA:** moderate to vigorous physical activity
**PA:** physical activity
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial
**SDT:** self-determination theory
**VPA:** vigorous physical activity

*Edited by L Buis; submitted 07.02.21; peer-reviewed by T Jämsä, S Grover; comments to author 08.04.21; revised version received 29.05.21; accepted 20.12.21; published 03.02.22.*

*Please cite as:*
*Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, Yu T, Lian X, Yu T, Yang X, Li F*
*The Effects of mHealth-Based Gamification Interventions on Participation in Physical Activity: Systematic Review*
*JMIR Mhealth Uhealth 2022;10(2):e27794*
*URL: https://mhealth.jmir.org/2022/2/e27794*
*doi:10.2196/27794*
*PMID:35113034*

Original Paper

# The Quality of Health Apps and Their Potential to Promote Behavior Change in Patients With a Chronic Condition or Multimorbidity: Systematic Search in App Store and Google Play

Alessio Bricca[1,2], PhD; Alessandro Pellegrini[1,2], MSc; Graziella Zangger[1,2], MSc; Jonas Ahler[2], MSc; Madalina Jäger[1,2], PhD; Søren T Skou[1,2], PhD

[1]Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[2]The Research Unit PROgrez, Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Slagelse, Denmark

**Corresponding Author:**
Alessio Bricca, PhD
Research Unit for Musculoskeletal Function and Physiotherapy
Department of Sports Science and Clinical Biomechanics
University of Southern Denmark
Campusvej 55
Odense, 5260
Denmark
Phone: 45 65509510
Email: abricca@health.sdu.dk

## Abstract

**Background:** Mobile apps offer an opportunity to improve the lifestyle of patients with chronic conditions or multimorbidity. However, for apps to be recommended in clinical practice, their quality and potential for promoting behavior change must be considered.

**Objective:** We aimed to investigate the quality of health apps for patients with a chronic condition or multimorbidity (defined as 2 or more chronic conditions) and their potential for promoting behavior change.

**Methods:** We followed the Cochrane Handbook guidelines to conduct and report this study. A systematic search of apps available in English or Danish on App Store (Apple Inc) and Google Play (Google LLC) for patients with 1 or more of the following common and disabling conditions was conducted: osteoarthritis, heart conditions (heart failure and ischemic heart disease), hypertension, type 2 diabetes mellitus, depression, and chronic obstructive pulmonary disease. For the search strategy, keywords related to these conditions were combined. One author screened the titles and content of the identified apps. Subsequently, 3 authors independently downloaded the apps onto a smartphone and assessed the quality of the apps and their potential for promoting behavior change by using the Mobile App Rating Scale (MARS; number of items: 23; score: range 0-5 [higher is better]) and the App Behavior Change Scale (ABACUS; number of items: 21; score: range 0-21 [higher is better]), respectively. We included the five highest-rated apps and the five most downloaded apps but only assessed free content for their quality and potential for promoting behavior change.

**Results:** We screened 453 apps and ultimately included 60. Of the 60 apps, 35 (58%) were available in both App Store and Google Play. The overall average quality score of the apps was 3.48 (SD 0.28) on the MARS, and their overall average score for their potential to promote behavior change was 8.07 (SD 2.30) on the ABACUS. Apps for depression and apps for patients with multimorbidity tended to have higher overall MARS and ABACUS scores, respectively. The most common app features for supporting behavior change were the self-monitoring of physiological parameters (eg, blood pressure monitoring; apps: 38/60, 63%), weight and diet (apps: 25/60, 42%), or physical activity (apps: 22/60, 37%) and stress management (apps: 22/60, 37%). Only 8 out of the 60 apps (13%) were completely free.

**Conclusions:** Apps for patients with a chronic condition or multimorbidity appear to be of acceptable quality but have low to moderate potential for promoting behavior change. Our results provide a useful overview for patients and clinicians who would like to use apps for managing chronic conditions and indicate the need to improve health apps in terms of their quality and potential for promoting behavior change.

*(JMIR Mhealth Uhealth 2022;10(2):e33168)* doi:10.2196/33168

**KEYWORDS**

app; self-management; behavior change; multimorbidity; chronic conditions; health apps; mHealth; mobile health; mobile phone

## Introduction

Osteoarthritis, hypertension, type 2 diabetes, depression, heart conditions, and chronic obstructive pulmonary disease are among the leading causes of global disability [1]. These conditions affect millions of people worldwide and are commonly co-occurring (ie, multimorbidity) [2]. Patients with chronic conditions have poorer physical and psychosocial health than those of people without such conditions, and the higher the number of co-occurring conditions, the greater the impact on the individual and society [3,4]. Importantly, these conditions can be prevented and managed by a healthy lifestyle, highlighting the importance of investigating this population [5].

A healthy lifestyle, which includes physical activity and a healthy diet, is associated with up to a 6.3-year longer lifespan for men and a 7.6-year longer lifespan for women with 1 or more chronic conditions [6]. Different care models and interventions have been designed and implemented for people with multiple chronic conditions [7]. Although there is a paucity of information about the effects of these interventions, physical activity appears safe and beneficial for people with 1 or more chronic conditions [8,9]. The use of mobile apps to improve lifestyle has increasingly gained attention, particularly among patients with chronic conditions at any stage of their lives [10-14]. Apps may offer an opportunity to improve the lifestyles of patients with chronic conditions through, for example, self-monitoring and behavior change by providing access to personalized support and motivation anytime [15,16]. However, although apps are widely used (in 2019, more than 204 billion apps were downloaded) [17], their quality (eg, engagement and functionality), content, and potential for promoting behavior change are unclear [18]. Therefore, this study aimed to provide an overview of available health apps and their quality, content, and potential to promote behavior change in patients living with a chronic condition or multimorbidity.

## Methods

This systematic search of health apps was guided by the recommendations for performing systematic reviews in the Cochrane Handbook [19], and the protocol was made available prior to the app screening phase on Open Science Framework [20].

### Eligibility Criteria

We included apps that targeted lifestyle behaviors, such as physical activity and diet, and were directed at patients with 1 or more of the following conditions (multimorbidity): osteoarthritis of the knee or hip, heart conditions (heart failure and ischemic heart disease), hypertension, type 2 diabetes mellitus, chronic obstructive pulmonary disease, and depression. The rationale for focusing on these conditions was that they share a common risk factor (physical inactivity) and pathogenesis (systemic low-grade inflammation) and the fact that they are highly prevalent and can co-occur with each other. Therefore, the anti-inflammatory effects of lifestyle behaviors may improve the health of this population [21].

### Search Strategy and App Selection

We searched the Apple App Store (Apple Inc; iOS) and Google Play Store (Google LLC; Android) for Danish and English apps. Two authors (AB and AP) designed the search strategy (Table S1 in Multimedia Appendix 1), which was adapted from a prior systematic review [8,22]. One author (AP) performed the search in October 2020 and screened the titles and descriptions of the apps. Three authors (AP, GZ, and JA) independently downloaded the apps onto a smartphone. In pairs, they assessed the quality of the apps and their potential for promoting behavior change by using the Mobile App Rating Scale (MARS; number of items: 23) and the App Behavior Change Scale (ABACUS; number of items: 21), respectively. The five highest-rated apps and the five most downloaded apps were included. Quality and potential for promoting behavior change were only assessed for the free apps, given that the cost to purchase apps is a barrier to using mobile health apps [23]. Furthermore, apps were excluded if they did not target patients (eg, apps that targeted clinicians or organizations).

### Data Extraction and Outcomes

A complete overview of the data extraction process is available in the study protocol [20]. The following outcomes were assessed: app quality and the potential to promote behavior change.

App quality was assessed by using the MARS [24]. This validated and objective tool allows for the classification and assessment of the quality of apps. It is a 23-item scale that includes the following five categories: engagement, functionality, aesthetics, information quality, and subjective quality. Each item is assessed on a 5-point scale (1=inadequate; 2=poor; 3=acceptable; 4=good; 5=excellent).

The potential for behavior change was assessed by using the ABACUS [25]. This validated and objective tool includes 21 items that are grouped into the following four categories: knowledge and information, goals and planning, feedback and monitoring, and actions. The score for each item is dichotomous (yes or no), and an overall score (range 0-21) can be calculated. The higher the score, the higher the potential for promoting behavior change.

In addition, we extracted the characteristics of the apps, such as the number and type of self-monitoring tools (eg, a step count and BMI calculator), by using the 42matters website [26] for data that were not available in the iOS and Android stores. Mean scores for the MARS and ABACUS were calculated by averaging the ratings across all of the domains of the scales. The SDs were estimated accordingly.

### Synthesis of Results

We performed a narrative synthesis of the results and presented the results in a tabular and graphical format.

## Results

### App Selection and Characteristics

A total of 453 apps were identified, of which 150 were downloaded and screened for potential eligibility, and 60 were ultimately included (Figure S1 in Multimedia Appendix 1). Most of the included apps (35/60, 58%) were available for both iOS and Android. The apps were all available in English, and 25 out of the 60 apps (42%) were also available in other languages, including Danish, Arabic, and Chinese. The app size varied from 2.4 MB to 278 MB. Table S2 in Multimedia Appendix 1 presents a complete description of the apps.

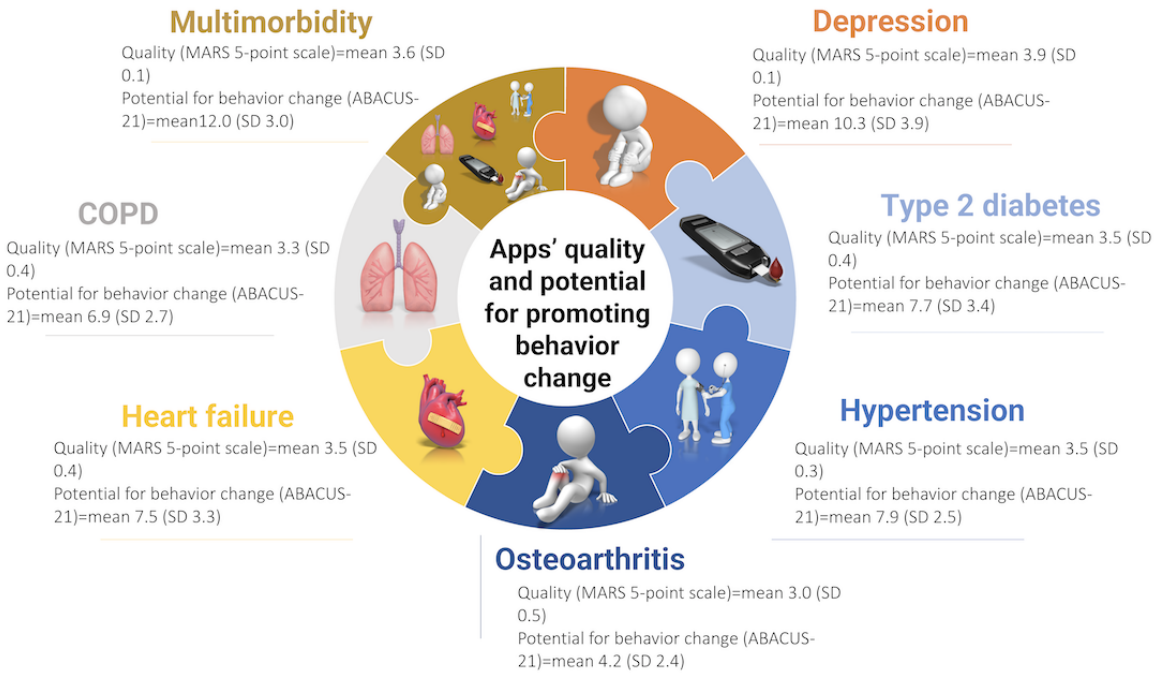
### Quality of the Apps

The overall quality of the apps was acceptable (MARS score: mean 3.48, SD 0.28; range 0-5). However, apps for depression tended to have a higher overall MARS score (mean 3.89, SD 0.13), and apps for osteoarthritis tended to have a lower overall score (mean 3.0, SD 0.48) and lower scores for individual items of the MARS (Tables S3 and S4 in Multimedia Appendix 1).

### Apps' Potential to Promote Behavior Change

The overall potential for behavior change was low to moderate (ABACUS score: mean 8.07, SD 2.30; range 0-21). Apps for patients with multimorbidity tended to have a higher overall ABACUS score (mean 12.0, SD 3.03), while apps for osteoarthritis tended to have the lowest overall scores (mean 4.22, SD 2.39) and the lowest scores for the individual categories of the ABACUS (Figure 1, Table S5 in Multimedia Appendix 1).

**Figure 1.** Summary of the findings for the quality of apps for osteoarthritis, hypertension, type 2 diabetes, depression, heart conditions, COPD, and multimorbidity and their potential for promoting behavior change. ABACUS: App Behavior Change Scale; COPD: chronic obstructive pulmonary disease; MARS: Mobile App Rating Scale.




### Features of the Apps That Supported Behavior Change

The most common features presented in the apps that supported behavior change were the self-monitoring of physiological parameters (eg, blood pressure monitoring; apps: 38/60, 63%), weight and diet (apps: 25/60, 42%), or physical activity (apps: 22/60, 37%) and stress management (apps: 22/60, 37%). Only 8 out of the 60 apps (13%) were completely free.

## Discussion

### Principal Findings

To the best of our knowledge, this is the first study to assess the quality of health apps for patients with 1 or more chronic conditions and their potential for promoting behavior change. The assessed apps' quality is acceptable, but their potential for promoting behavior change in patients with osteoarthritis, hypertension, type 2 diabetes, depression, heart conditions, chronic obstructive pulmonary disease, or multimorbidity is low to moderate. This highlights the need for future studies to develop and evaluate apps with both high quality and high potential for promoting behavior change in patients with chronic conditions and multimorbidity.

The results of this study are comparable to the results of systematic reviews that investigated apps' quality and potential for promoting behavior change in patients with a single chronic condition [27-29]. In these reviews, both low to moderate quality and low to moderate potential for promoting behavior change were found. Apps for multimorbidity tended to have higher quality and higher potential for promoting behavior change than

those of apps for a single chronic condition despite the fact that research on multimorbidity is still in its infancy [30]. Future app studies should focus on improving quality and the potential for behavior change, especially among apps for conditions such as osteoarthritis, which had the lowest-quality apps [28,31]. Future studies should also test the effectiveness of apps via high-quality randomized controlled trials.

The features of the apps were similar across the chronic conditions, including multimorbidity, and focused mainly on the self-monitoring and tracking of physiological and behavioral parameters, such as medication intake, step count, and diet. In contrast, only a minority of apps (22/60, 37%) focused on psychosocial support, although mental and social health plays a major role in managing chronic conditions and multimorbidity [32-34]. This should be kept in mind when designing new apps.

Most of the top-rated and most downloaded apps for patients with a chronic condition or multimorbidity were not completely free (52/60, 87%). Notably, there were no free apps for depression. Nevertheless, the quality of apps for depression and their potential for behavior change were higher than those of apps for osteoarthritis, and 78% (7/9) of osteoarthritis apps were completely free. Although the development of apps has a cost, 1 in 2 smartphone users have never paid for an app [35], and the cost of apps is a barrier to using them [23]. This should be considered when designing new app-based interventions.

### Limitations

A possible limitation of this study is that we only assessed the free content of the apps. However, the potential for promoting behavior change appeared to be similar among free apps and apps with in-app purchases [27]. Furthermore, we only focused on English or Danish apps, meaning that our findings may not be generalizable to apps in other languages. We were also unable to identify apps targeting patients with ischemic heart disease and extract data on the characteristics of people who downloaded the apps (eg, age). This limited the generalizability of the findings related to apps for heart conditions to apps for patients with heart failure. This also limited our ability to conduct stratified subgroup assessments. Finally, the limited number of apps available for each condition prevented the meaningful comparison of the MARS and ABACUS subscales within and between conditions.

### Conclusions

Our results provide patients and clinicians with an overview of apps for managing 1 or more chronic conditions and indicate the need to improve the quality of apps and their potential for promoting behavior change, particularly among apps for patients with osteoarthritis.

## Acknowledgments

We would like to acknowledge Benedicte Lauge Hansen for conducting a pilot search of apps targeting patients with 1 or more chronic conditions as part of her internship. This study was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (MOBILIZE; grant 801790); Næstved, Slagelse and Ringsted Hospitals' Research Fund; and The Association of Danish Physiotherapists Research Fund and Region Zealand (Exercise First). The funding sources were not involved in any aspect of this work.

## Authors' Contributions

STS procured the funding for this work. AB, MJ, and STS drafted the protocol, and AP finalized it. AB and AP designed the search strategy. AP screened the titles and descriptions of the apps, and AP, JA, and GZ independently extracted data from the included apps and assessed their quality and potential for behavior change. AB wrote the manuscript. All authors read, provided feedback for, and approved the study design, methods, protocol, and manuscript drafts as well as approved the final manuscript.

## Conflicts of Interest

STS has received grants from The Lundbeck Foundation and personal fees from Munksgaard and TrustMe-Ed, of which all are outside the submitted work. STS is a cofounder of GLA:D (Good Life With osteoArthritis in Denmark), a not-for-profit initiative hosted at University of Southern Denmark that aims to implement clinical guidelines for osteoarthritis in clinical practice.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 126 KB - mhealth_v10i2e33168_app1.docx ]

## References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Bayliss EA, Bayliss MS, Ware Jr JE, Steiner JF. Predicting declines in physical function in persons with multiple chronic medical conditions: what we can learn from the medical problem list. Health Qual Life Outcomes 2004 Sep 07;2:47 [FREE Full text] [doi: 10.1186/1477-7525-2-47] [Medline: 15353000]

3. Navickas R, Petric VK, Feigl AB, Seychell M. Multimorbidity: What do we know? What should we do? J Comorb 2016 Feb 17;6(1):4-11 [FREE Full text] [doi: 10.15256/joc.2016.6.72] [Medline: 29090166]
4. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22 Suppl 3(Suppl 3):391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
5. Ioakeim-Skoufa I, Poblador-Plou B, Carmona-Pírez J, Díez-Manglano J, Navickas R, Gimeno-Feliu LA, et al. Multimorbidity patterns in the general population: Results from the EpiChron Cohort Study. Int J Environ Res Public Health 2020 Jun 14;17(12):4242 [FREE Full text] [doi: 10.3390/ijerph17124242] [Medline: 32545876]
6. Chudasama YV, Khunti K, Gillies CL, Dhalwani NN, Davies MJ, Yates T, et al. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: A longitudinal cohort study. PLoS Med 2020 Sep 22;17(9):e1003332. [doi: 10.1371/journal.pmed.1003332] [Medline: 32960883]
7. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2021 Jan 15;1(1):CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub4] [Medline: 33448337]
8. Bricca A, Harris LK, Jäger M, Smith SM, Juhl CB, Skou ST. Benefits and harms of exercise therapy in people with multimorbidity: A systematic review and meta-analysis of randomised controlled trials. Ageing Res Rev 2020 Nov;63:101166 [FREE Full text] [doi: 10.1016/j.arr.2020.101166] [Medline: 32896665]
9. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
10. 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 Aug 04;8(8):e15779 [FREE Full text] [doi: 10.2196/15779] [Medline: 32459654]
11. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019 Oct 10;5:2055207619882181 [FREE Full text] [doi: 10.1177/2055207619882181] [Medline: 31656632]
12. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: A scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
13. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A short message service intervention to support adherence to home-based strengthening exercise for people with knee osteoarthritis: Intervention design applying the behavior change wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
14. Gray CS, Wodchis WP, Upshur R, Cott C, McKinstry B, Mercer S, Project Collaborators And Technology Partner, QoC Health Inc. Supporting goal-oriented primary health care for seniors with complex care needs using mobile technology: Evaluation and implementation of the Health System Performance Research Network, Bridgepoint Electronic Patient Reported Outcome Tool. JMIR Res Protoc 2016 Jun 24;5(2):e126 [FREE Full text] [doi: 10.2196/resprot.5756] [Medline: 27341765]
15. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
16. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
17. Sydow L. The state of mobile in 2020: The key stats you need to know. App Annie. URL: https://www.appannie.com/en/insights/market-data/state-of-mobile-2020-infographic/ [accessed 2020-08-26]
18. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018 May 09;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
19. Cochrane Handbook for Systematic Reviews of Interventions, Version 6, 2019. Cochrane Training. 2019. URL: https://training.cochrane.org/handbook/archive/v6 [accessed 2022-02-01]
20. Bricca A, Jäger M, Skou ST, Pihl K, Harris LK, Dideriksen M. Mobile health applications (apps) to support self-management in people living with a chronic condition or multimorbidity: A systematic search in App Store and Google Play. Open Science Framework. URL: https://osf.io/4xkpt/ [accessed 2022-01-27]
21. Gleeson M, Bishop NC, Stensel DJ, Lindley MR, Mastana SS, Nimmo MA. The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease. Nat Rev Immunol 2011 Aug 05;11(9):607-615. [doi: 10.1038/nri3041] [Medline: 21818123]
22. Bricca A, Harris LK, Saracutu M, Smith SM, Juhl CB, Skou ST. The benefits and harms of therapeutic exercise on physical and psychosocial outcomes in people with multimorbidity: Protocol for a systematic review. J Comorb 2020 May 12;10:2235042X20920458 [FREE Full text] [doi: 10.1177/2235042X20920458] [Medline: 32426294]

23. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. 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]
26. Mobile app intelligence. 42matters. URL: https://42matters.com/ [accessed 2020-10-25]
27. 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 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
28. Choi W, Zheng H, Franklin P, Tulu B. mHealth technologies for osteoarthritis self-management and treatment: A systematic review. Health Informatics J 2019 Sep;25(3):984-1003 [FREE Full text] [doi: 10.1177/1460458217735676] [Medline: 29090628]
29. Wali S, Demers C, Shah H, Wali H, Lim D, Naik N, et al. Evaluation of heart failure apps to promote self-care: Systematic app search. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e13173 [FREE Full text] [doi: 10.2196/13173] [Medline: 31710298]
30. The Lancet. Making more of multimorbidity: an emerging priority. Lancet 2018 Apr 28;391(10131):1637. [doi: 10.1016/S0140-6736(18)30941-3] [Medline: 29726322]
31. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
32. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
33. de Alba IGF, Gimeno-Miguel A, Poblador-Plou B, Gimeno-Feliu LA, Ioakeim-Skoufa I, Rojo-Martínez G, et al. Association between mental health comorbidity and health outcomes in type 2 diabetes mellitus patients. Sci Rep 2020 Nov 11;10(1):19583 [FREE Full text] [doi: 10.1038/s41598-020-76546-9] [Medline: 33177607]
34. Triolo F, Harber-Aschan L, Murri MB, Calderón-Larrañaga A, Vetrano DL, Sjöberg L, et al. The complex interplay between depression and multimorbidity in late life: risks and pathways. Mech Ageing Dev 2020 Dec;192:111383 [FREE Full text] [doi: 10.1016/j.mad.2020.111383] [Medline: 33045250]
35. How people discover, use, and stay engaged with apps. Think with Google. 2016 Oct. URL: https://www.thinkwithgoogle.com/_qs/documents/331/how-users-discover-use-apps-google-research.pdf [accessed 2021-08-22]


## Abbreviations

**ABACUS:** App Behavior Change Scale
**GLA:D:** Good Life With osteoArthritis in Denmark
**MARS:** Mobile App Rating Scale


*Edited by L Buis; submitted 26.08.21; peer-reviewed by S Davis, I Ioakeim Skoufa, B Chaudhry; comments to author 18.10.21; revised version received 29.10.21; accepted 07.12.21; published 04.02.22.*

*Please cite as:*
*Bricca A, Pellegrini A, Zangger G, Ahler J, Jäger M, Skou ST*
*The Quality of Health Apps and Their Potential to Promote Behavior Change in Patients With a Chronic Condition or Multimorbidity: Systematic Search in App Store and Google Play*
*JMIR Mhealth Uhealth 2022;10(2):e33168*
*URL: https://mhealth.jmir.org/2022/2/e33168*
*doi:10.2196/33168*
*PMID:35119367*

Original Paper

# Persuasive Design Solutions for a Sustainable Workforce: Review of Persuasive Apps for Real-Time Capability Support for Rural Health Care Professionals

Sabrina Winona Pit[1], MSc, PhD; Aaron J H Tan[1], BMed, MD; Robyn Ramsden[1], BEd, PhD; Kristy Payne[1], BPsych, MClinEpi; Winona Freihaut[1]; Oliver Hayes[1]; Benjamin Eames[1], BBus; Mike Edwards[1], BCompSc; Richard Colbran[1], BA

New South Wales Rural Doctors Network, Hamilton, Australia

**Corresponding Author:**
Sabrina Winona Pit, MSc, PhD
New South Wales Rural Doctors Network
1/53 Cleary St
Hamilton, 2303
Australia
Phone: 61 429455720
Email: spit@nswrdn.com.au

## Abstract

**Background:** There is a need to further investigate how persuasive design principles can change rural health professionals' behaviors to look after their own health workforce capability. Several theories are used when developing apps to persuade people to change behavior, including the Persuasive System Design Model, consisting of primary task, dialogue, system credibility, and social support categories, and Cialdini's principles of persuasion. These have not been analyzed yet in the field of health workforce capability.

**Objective:** This study aims to determine the persuasive design techniques used in capability building–related apps and to provide recommendations for designing a health workforce app to increase their persuasiveness.

**Methods:** A Python script was used to extract a total of 3060 apps from Google Play. Keywords centered around health workforce capability elements. App inclusion criteria were as follows: been updated since 2019, rated by users on average 4 and above, and more than 100,000 downloads. Next, 2 experts reviewed whether 32 persuasive strategies were used in the selected apps, and these were further analyzed by capability categories: competencies and skills, health and personal qualities, values and attitudes, and work organization.

**Results:** In all, 53 mobile apps were systematically reviewed to identify the persuasive design techniques. The most common were surface credibility (n=48, 90.6%) and liking (n=48), followed by trustworthiness (n=43, 81.1%), reminders (n=38, 71.7%), and suggestion (n=30, 56.6%). The techniques in the social support domain were the least used across the different apps analyzed for health workforce capability, whereas those in the primary task support domain were used most frequently. The recommendations reflect learnings from our analysis. These findings provided insight into mobile app design principles relevant to apps used in improving health workforce capability.

**Conclusions:** Our review showed that there are many persuasive design techniques that can assist in building health workforce capability. Additionally, several apps are available in the market that can assist in improving health workforce capability. There is, however, a specific lack of digital, real-time support to improve health workforce capability. Social support strategies through using social support persuasive design techniques will need to be integrated more prominently into a health workforce capability app. An app to measure and monitor health workforce capability scores can be used in conjunction with direct real-world person and real-time support to discuss and identify solutions to improve health workforce capability for rural and remote health professionals who are at high risk of burnout or leaving the rural health workforce.

***(JMIR Mhealth Uhealth 2022;10(2):e33413)*** doi:10.2196/33413

**KEYWORDS**

health; wellness; mobile apps; persuasive strategies; behavior change; review; health workforce; capability; career; employment; rural; workforce planning

## Introduction

Sustainability of the rural health workforce is paramount to meeting the health needs of rural Australia. Significant resources have been dedicated to identifying solutions to problems around recruitment or retention of the rural health workforce [1]. However, although extensive research has been conducted into the drivers behind retention and attrition of the rural health workforce, the literature does not adequately address how technology can be used to support rural health professionals in real time to improve their workforce capability. Health workforce capability can be defined as "the intersection between individual capacity and ability to respond to work, considering the whole of rural life, including work, family, schools, partner, education, and social options." In other words, health workforce capability "describes a health professional's overall level of capability in fulfilling their health care role."

A 2021 scoping review found that digital solutions can enhance the capability and retention of rural health professionals [2]. The authors concluded that online platforms or digital solutions can address many of the challenges experienced by rural health professionals, by improving knowledge and skills, access, translation of knowledge into practice, empowerment, confidence, engagement, and provision of support. To this effect, for any digital app to be effective in improving rural health workforce capability, the desired behavior change that is needed must be considered during the development of the technical solutions. Many theories exist around user acceptance of IT solutions, such as the Mobile Application Rating Scale [3], the Technology Acceptance Model [4], and the Health Information Technology Acceptance Model [5]. However, they do not provide a clear systematic analysis and design criteria for developing persuasive software solutions to increase the likelihood of achieving the desired behavioral change [6].

One theory that can be used when developing apps to persuade people to change behavior include the Persuasive System Design Model (PSD-Model) [6]. This model can be used to identify the software functionality that may be useful in a product. Specifically, 28 design principles are provided and categorized into 4 main domains of persuasive techniques. These are the primary task, dialogue, system credibility, and social support domains. These 4 domains form part of the "design of system qualities" and have been applied to several health-related topics, such as arthritis [7] and mental health [8]. In their work on deconstructing persuasive principles and their implementation in health apps, Oyebode et al [9] expanded the scope of the PSD-Model by augmenting its 28 design principles with select techniques described in Cialdini's principles of persuasion [10,11]. The authors selected only 4 persuasive techniques from Cialdini's principles of persuasion because 2 were already present reflected in the PSD-Model, being authority and liking. This resulted in Oyebode et al [9] analyzing 32 persuasive design techniques in total. These 32 design techniques are classified into 5 domains, including the 4 from the PSD-Model listed above and Cialdini's principles of persuasion. Specifically, Thach and Phan [8] reviewed the users' perception of persuasiveness of mental health apps by qualitatively evaluating user reviews and concluded that when the principles of the PSD-Model are integrated into the design, users are happy with the design.

The question remains whether these design principles can be used when developing a technical solution to assist rural health professionals in maintaining or improving their health workforce capability. There are no clear data on this yet. A 2017 study that has provided some insight was a meta-analysis conducted by Carolan et al [12] of 21 randomized controlled trials (RCTs) involving web-based psychological interventions delivered in the workplace to improve employee psychological well-being and increase work effectiveness. In addition to exploring intervention effectiveness, the study identified features, including persuasive strategies, associated with greater engagement and adherence. The authors found that online interventions improve both psychological wellbeing and work effectiveness, but no differences were found between cognitive behavioral therapy versus other approaches, guidance versus no guidance, and targeted workplace populations versus the general workplace population. Work effectiveness was measured as work engagement, productivity, job-specific effectiveness, work-related self-efficacy, and work-related remuneration. Further analyses identified the following effective features that may likely increase engagement:

- Guidance delivered during a shorter period (6-7 weeks)
- Using secondary modalities to deliver the intervention (eg, emails, text messages, short messaging service)
- Elements of persuasive technology (eg, self-monitoring and tailoring)

Tailoring (57%), self-monitoring (43%), and tunneling (14%) were found to be used as persuasion strategies in studies included in the review and associated with the highest rates of engagement and adherence. Although the meta-analysis provided some useful insights, it did not focus on improving health workforce capability as such. There is a gap in the literature around how persuasive design techniques in digital solutions can change rural health professionals' behaviors to look after their own health workforce capability.

Therefore, this study aimed to determine:

- The persuasive design techniques used to improve use of capability building–related apps; and
- Recommendations for incorporating persuasive design techniques in the design of health workforce capability apps.

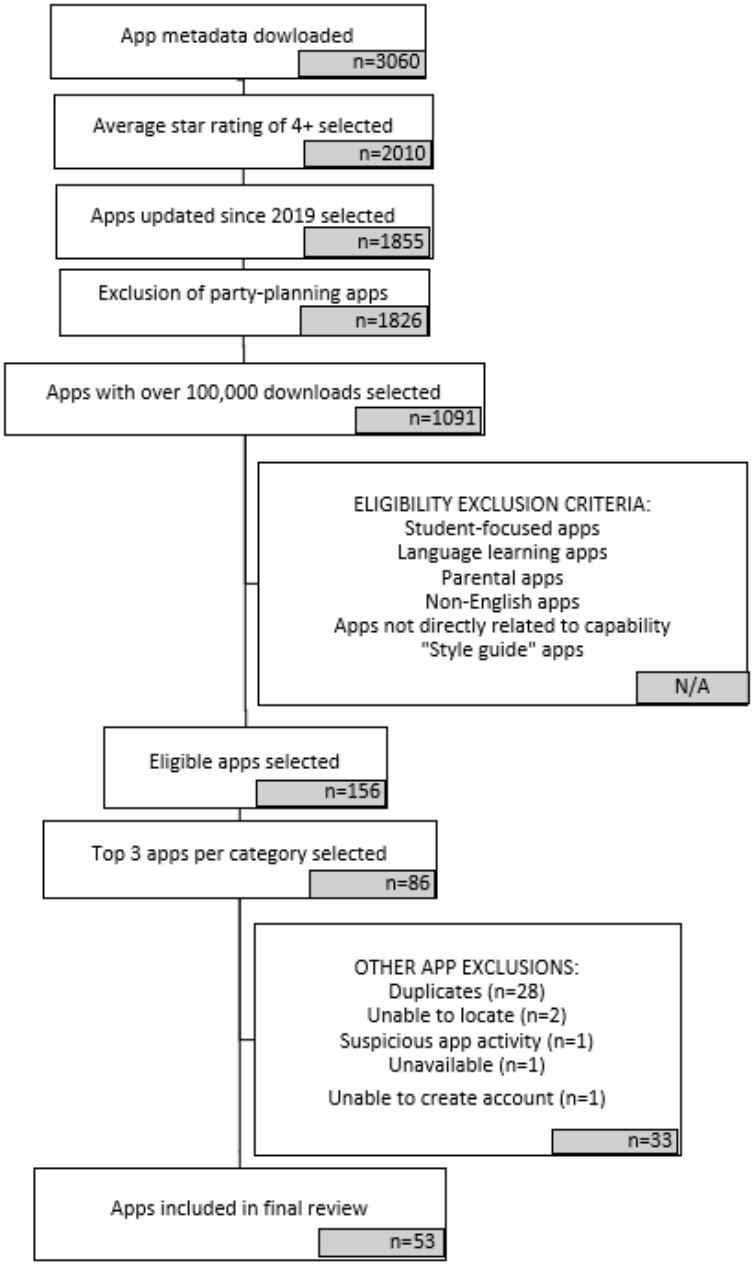
## Methods

### Inclusion and Exclusion Criteria and Search Criteria

We based our methodology primarily on the study conducted by Oyebode et al [9]. A Python script was developed to extract apps from Google Play. Several rounds were conducted to refine the keywords (see Multimedia Appendix 1). Keywords centered around health workforce capability elements. The script limited searches to only English language, Australia only, and both free and paid apps. Apps with an average star rating of 4 and above were included to ensure only higher-quality apps from the user perspective were examined. Authors SP and RR reviewed 1091 apps to determine inclusion.

Figure 1 describes the main steps in the app selection process, and Multimedia Appendix 1 provides additional information.

**Figure 1.** App selection and exclusion process.




### Data Extraction, Coding, and Data Analyses of Selected Apps

Two theories were used in the apps to persuade people to change behavior: the PSD-Model [6] and Cialdini's principles of persuasion [11]. Both theories have been extensively used when analyzing persuasive technologies [9]. Oyebode et al [9] determined that 32 techniques can be measured based on both theories, as the techniques authority and liking were common to both. Next, 2 authors (WF and OH) with a data sciences and psychology background reviewed whether the 32 persuasive design techniques were applied in the apps.

The reviewers downloaded the apps on an Android smartphone or emulator and used the apps to determine whether the persuasive design techniques were present in the apps. The reviewers coded independently and then came together to compare scores to ensure interrater reliability. If there was a

discrepancy, a discussion was held to reach consensus. The coders double-coded all apps and achieved a κ of .81, indicating very good agreement between the 2 coders.

### Data Analyses

After scoring the 53 apps, the apps were divided in 4 broad workforce capability app categories to determine whether there were differences between the different persuasive design techniques used within each category, including competencies and skills, health and personal qualities, values and attitudes, and work organization. This classification was chosen as it aligned with findings from a qualitative data analysis that had been simultaneously conducted to explore user perspectives on building a health workforce capability app (Ramsden et al, unpublished data, 2021). The classification also mirrors elements of the work ability to explain sustainable employability among general practitioners [13]. Competencies and skills include elements such as clinical competence, drug and clinical references, and medical journals. Health and personal quality apps are related to the health professionals' own lives and not patient-related or medical knowledge–type apps. Work organization includes elements such as patient-centered care and entrepreneurial skills. Value and attitude apps include elements such as resilience and self-motivation.

Following this classification, data were analyzed as follows:

- Total number of persuasive design techniques used per domain: primary task support, dialogue support, system credibility support, social support, and Cialdini's principles of persuasion
- Total number and percentage of each persuasive design technique used in each health workforce capability category: competence and skill, health and personal qualities, values and attitudes, and work organization
- Total number of persuasive design techniques used by the persuasive design technique domain across the 4 workforce capability categories developed for this systematic review: competencies and skills, health and personal qualities, values and attitudes, and work organization

Following the quantitative analyses of apps, implementation examples of the PSD-Model, consisting of primary task, dialogue, system credibility, and social support categories, and Cialdini's principles of persuasion are provided.

### Ethics

Ethical approval to conduct the feasibility study was received from the Northern NSW Local Health District Human Research Ethics Committee (2020/ETH03020).

## Results

### Overall Results

Figure 2 shows that the most common persuasive design techniques were surface credibility (48/53, 90.6%) and liking (48/53, 90.6%), followed by trustworthiness (43/53, 81.1%), reminders (38/53, 71.7%), and suggestion (30/53, 56.6%). Social support persuasive design techniques were the least used across the different apps analyzed for health workforce capability, whereas primary task support design techniques were used most frequently overall, with the exception of simulation.

Table 1 shows the most used persuasive techniques per workforce capability category. Those that were used in 50% or more of the total 53 apps are formatted in italic. It is evident that health apps included the most persuasive techniques, as did the apps related to values and attitudes. Apps related to work organization included the least number of design techniques.

**Figure 2.** Total number of persuasive design techniques used by category.

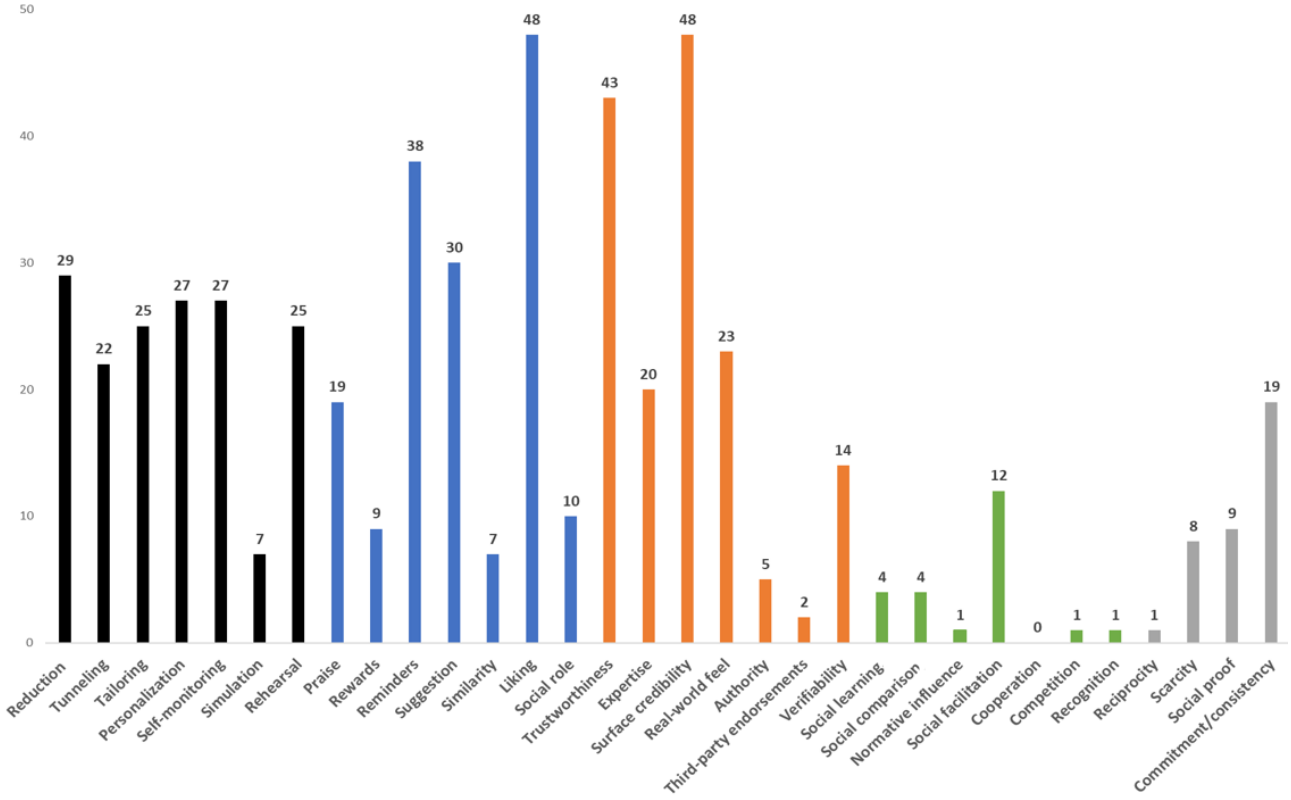


**Table 1.** Persuasive design techniques used by the health workforce capability group.

| Persuasive design technique | Count, n (%) |
|---|---|
| **Competence and skill (n=16)** | |
| *Surface credibility* [a] | 16 (100) |
| *Liking* | 15 (94) |
| *Trustworthiness* | 14 (88) |
| *Reminders* | 10 (63) |
| *Tailoring* | 9 (56) |
| *Personalization* | 9 (56) |
| *Tunneling* | 8 (50) |
| *Expertise* | 8 (50) |
| Reduction | 7 (44) |
| Rehearsal | 7 (44) |
| Suggestion | 7 (44) |
| Self-monitoring | 6 (38) |
| Real-world feel | 6 (38) |
| Verifiability | 6 (38) |
| Simulation | 4 (25) |
| Praise | 4 (25) |
| Scarcity | 4(25) |
| Rewards | 3 (19) |
| Authority | 3 (19) |
| Commitment/consistency | 3 (19) |
| Similarity | 2 (13) |
| Social role | 2 (13) |
| Social proof | 2 (13) |
| Social facilitation | 1 (6) |
| Third-party endorsements | 0 (0) |
| Social learning | 0 (0) |
| Social comparison | 0 (0) |
| Normative influence | 0 (0) |
| Cooperation | 0 (0) |
| Competition | 0 (0) |
| Recognition | 0 (0) |
| Reciprocity | 0 (0) |
| **Health and personal qualities (n=17)** | |
| *Liking* | 17 (100) |
| *Surface credibility* | 17 (100) |
| *Reminders* | 15 (88) |
| *Trustworthiness* | 15 (88) |
| *Suggestion* | 13 (76) |
| *Reduction* | 12 (71) |
| *Self-monitoring* | 11 (65) |
| *Personalization* | 10 (59) |

| Persuasive design technique | Count, n (%) |
|---|---|
| *Real-world feel* | 10 (59) |
| *Rehearsal* | 9 (53) |
| *Commitment/consistency* | 9 (53) |
| Tunneling | 8 (47) |
| Tailoring | 8 (47) |
| Praise | 7 (41) |
| Expertise | 5 (29) |
| Social facilitation | 5 (29) |
| Social proof | 4 (24) |
| Social role | 3 (18) |
| Verifiability | 3 (18) |
| Social learning | 3 (18) |
| Scarcity | 3 (18) |
| Rewards | 2 (12) |
| Social comparison | 2 (12) |
| Similarity | 1 (6) |
| Authority | 1 (6) |
| Reciprocity | 1 (6) |
| Simulation | 0 (0) |
| Third-party endorsements | 0 (0) |
| Normative influence | 0 (0) |
| Cooperation | 0 (0) |
| Competition | 0 (0) |
| Recognition | 0 (0) |
| **Values and attitudes (n=10)** | |
| *Liking* | 10 (100) |
| *Surface credibility* | 8 (80) |
| *Self-monitoring* | 7 (70) |
| *Reminders* | 7 (70) |
| *Trustworthiness* | 7 (70) |
| *Reduction* | 6 (60) |
| *Tailoring* | 6 (60) |
| *Rehearsal* | 6 (60) |
| *Praise* | 5 (50) |
| *Suggestion* | 5 (50) |
| *Commitment/consistency* | 5 (50) |
| Tunneling | 4 (40) |
| Personalization | 4 (40) |
| Real-world feel | 4 (40) |
| Rewards | 3 (30) |
| Social role | 3 (30) |
| Expertise | 3 (30) |
| Verifiability | 3 (30) |

| Persuasive design technique | Count, n (%) |
|---|---|
| Similarity | 2 (20) |
| Social facilitation | 2 (20) |
| Simulation | 1 (10) |
| Third-party endorsement | 1 (10) |
| Social comparison | 1 (10) |
| Social proof | 1 (10) |
| Authority | 0 (0) |
| Social learning | 0 (0) |
| Normative influence | 0 (0) |
| Cooperation | 0 (0) |
| Competition | 0 (0) |
| Recognition | 0 (0) |
| Reciprocity | 0 (0) |
| Scarcity | 0 (0) |
| **Work organization (n=10)** | |
| *Trustworthiness* | 7 (70) |
| *Surface credibility* | 7 (70) |
| *Reminders* | 6 (70) |
| *Liking* | 6 (60) |
| *Suggestion* | 5 (50) |
| Reduction | 4 (40) |
| Personalization | 4 (40) |
| Expertise | 4 (40) |
| Social facilitation | 4 (40) |
| Rehearsal | 3 (30) |
| Self-monitoring | 3 (30) |
| Praise | 3 (30) |
| Real-world feel | 3 (30) |
| Tunneling | 2 (20) |
| Tailoring | 2 (20) |
| Simulation | 2 (20) |
| Similarity | 2 (20) |
| Social role | 2 (20) |
| Verifiability | 2 (20) |
| Social proof | 2 (20) |
| Commitment/consistency | 2 (20) |
| Rewards | 1 (10) |
| Authority | 1 (10) |
| Third-party endorsement | 1 (10) |
| Social learning | 1 (10) |
| Social comparison | 1 (10) |
| Normative influence | 1 (10) |
| Competition | 1 (10) |

| Persuasive design technique | Count, n (%) |
|---|---|
| Recognition | 1 (10) |
| Scarcity | 1 (10) |
| Cooperation | 0 (0) |
| Reciprocity | 0 (0) |

[a]Design techniques used in 50% or more of the total 53 apps are formatted in italic.

## Persuasive Design Techniques' Domains by Workforce Capability Category

Figures 3-7 show the various persuasive design techniques' domains across the 4 workforce capability categories developed for this systematic review: competencies and skills, health and personal qualities, values and attitudes, and work organization.

The following observations were made:

- Primary task support techniques demonstrated use among all workforce capability categories, with *reduction* and *self-monitoring* being less evident in competence-related apps when compared to the health apps.
- For dialogue support, the most frequently used technique was liking, which was high among all categories followed by *reminders* (Figure 5).
- Among system credibility support techniques, *surface credibility* featured most strongly among health (n=16) and competence (n=16) apps, which was similar for *trustworthiness* (n=15 for health and n=14 for competences), while *real-world feel* was most frequently observed in health apps (Figure 6).
- Social support techniques did not feature strongly across all domains, with the exception of *social facilitation* (n=5 for health and n=4 for work organization) (Figure 7).
- Among Cialdini's principles of persuasion, *commitment and consistency* played a large role in health and personal quality–related apps (n=9), followed by value- and attitude-related apps (n=5), while *reciprocity* barely played a role in any workforce capability category (Figure 7).

**Figure 3.** Total number of persuasive design techniques used by the primary task support category.

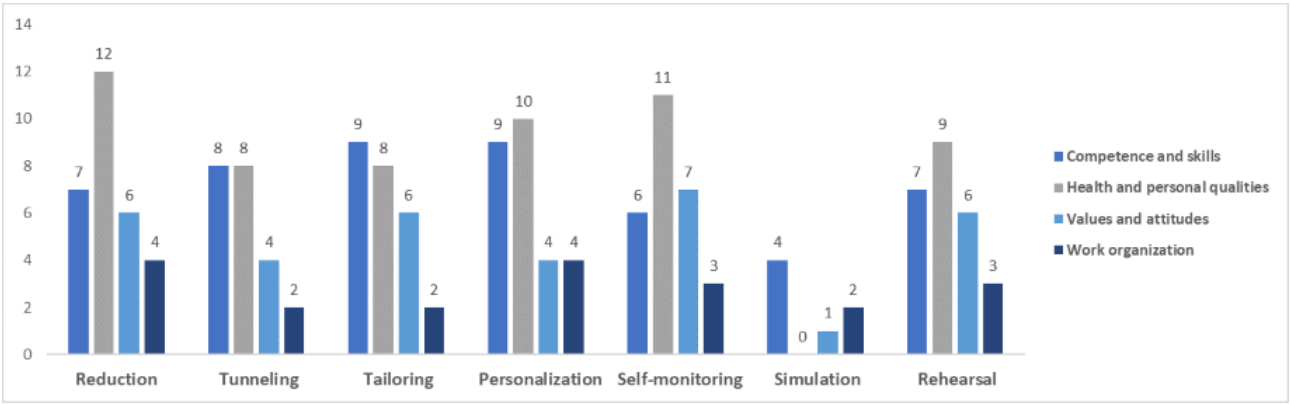



**Figure 4.** Total number of persuasive design techniques used by the dialogue support category.

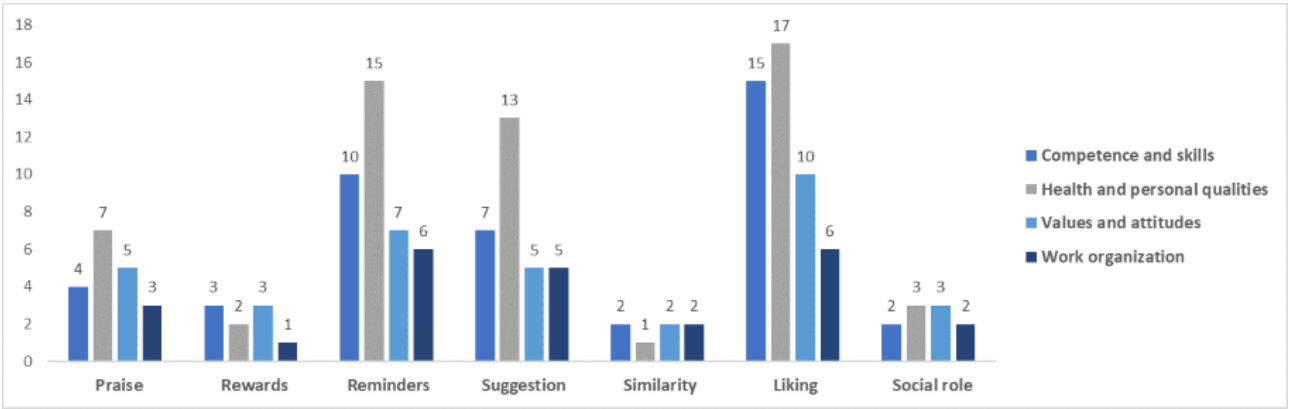


**Figure 5.** Total number of persuasive design techniques used by the system credibility support category.

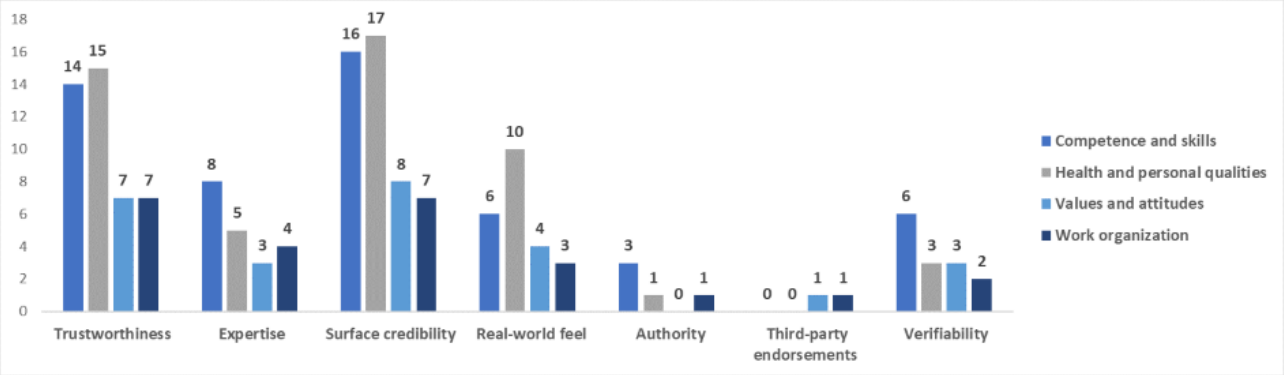



**Figure 6.** Total number of persuasive design techniques used by the of social support category.

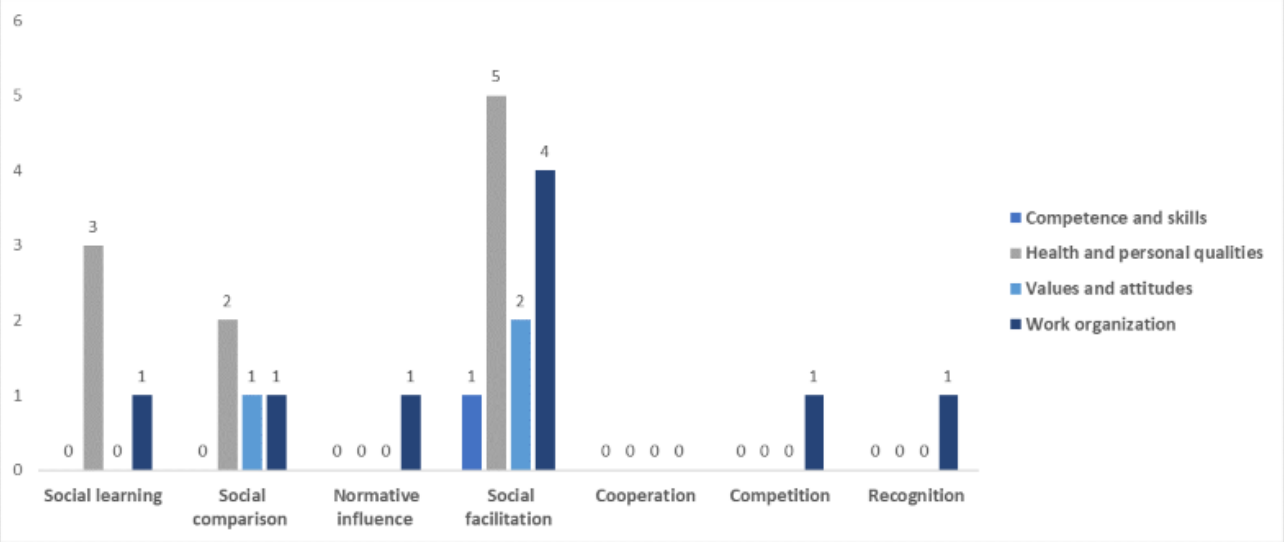
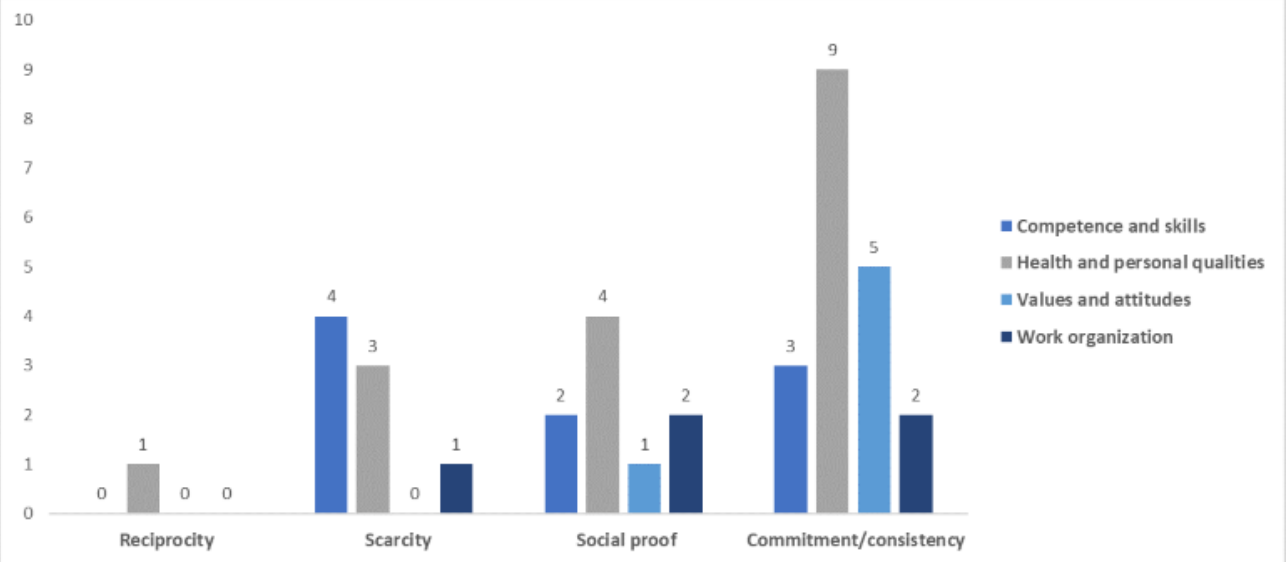



**Figure 7.** Total number of persuasive design techniques used by Cialdini's principles of persuasion.




## Persuasive Design Techniques and Implementation Suggestions Based on Existing Apps

Tables 2-6 show the various persuasive design techniques used in the apps analyzed in this study and how these can be extrapolated to create suggestions for developing a health workforce capability app.

**Table 2.** Persuasive design techniques: definitions and implementation examples for developing a health workforce capability app in the primary task support domain.

| Persuasive design technique | Persuasive design technique definition[a] | Implementation examples |
|---|---|---|
| Reduction | A system that reduces complex behavior into simple tasks helps users perform the target behavior, and it may increase the benefit/cost ratio of a behavior. | • Health workforce capability goals are broken down into smaller steps. |
| Tunneling | Using the system to guide users through a process or experience provides opportunities to persuade along the way. | • A user should be able to choose a pathway that would respond to their specific health workforce capability need. |
| Tailoring | Information provided by the system will be more persuasive if it is tailored to the potential feeds, interests, personality, usage context, or other factors relevant to a user group. | • A user should be able to choose what capability area they are particularly interested in. |
| Personalization | A system that offers personalized content or services has a greater capability for persuasion. | • The system adjusts to the health care profession and the user's age and offers localized services to improve health workforce capability. |
| Self-monitoring | A system that keeps track of one's own performance or status supports the user in achieving goals. | • The system asks health professional to rate their health workforce capability. |
| Simulation | Systems that provide simulations can persuade by enabling users to observe immediately the link between cause and effect. | • The system allows for health professionals to observe other health professionals working interprofessionally and see improved patient outcomes. |
| Rehearsal | A system providing means with which to rehearse a behavior that can enable people to change their attitudes or behavior in the real world. | • A telehealth simulation course is offered to rehearse real-world practice and improve use of telehealth. |

[a]Source: Oinas-Kukkonen and Harjumaa [6].

**Table 3.** Persuasive design techniques: definitions and implementation examples for developing a health workforce capability app in the dialogue support domain.

| Persuasive design technique | Persuasive design technique definition[a] | Implementation examples |
|---|---|---|
| Praise | By offering praise, a system can make users more open to persuasion. | • Texts and symbols are used to offer praise after measuring their own capability score. |
| Rewards | Systems that reward target behaviors may have great persuasive powers. | • The game rewards users by altering media items, such as sounds and background colors. |
| Reminders | If a system reminds users of their target behavior, the users will more likely achieve their goals. | • The user is given a reminder of a selected task to improve capability. |
| Suggestion | Systems offering fitting suggestions will have greater persuasive powers. | • Suggestions are given to be mindful at work or build capability. |
| Similarity | People are more readily persuaded through systems that remind them of themselves in some meaningful way. | • Videos/pictures of health professionals are shown. |
| Liking | A system that is visually attractive for its users is likely to be more persuasive. | • The application has an integrated system that links well with easy-to-read graphs and trends in health workforce capability. |
| Social role | If a system adopts a social role, users will more likely use it for persuasive purposes. | • A dementia expert supports online education for trainees with dementia. |

[a]Source: Oinas-Kukkonen and Harjumaa [6].

**Table 4.** Persuasive design techniques: definitions and implementation examples for developing a health workforce capability app in the system credibility support domain.

| Persuasive design technique | Persuasive design technique definition[a] | Implementation examples |
|---|---|---|
| Trustworthiness | A system that is viewed as trustworthy will have increased powers of persuasion. | • The system includes a privacy statement.<br>• The app demonstrates that the organization has access to funding to support health professionals in their health workforce capability and demonstrates successful examples. |
| Expertise | A system that is viewed as incorporating expertise will have increased powers of persuasion. | • The app demonstrates that the organization has a long-standing reputation in providing health workforce capability support. |
| Surface credibility | People make initial assessments of the system credibility based on a firsthand inspection. | • There are no commercial ads. |
| Real-world feel | A system that highlights people or the organization behind its content or services will have more credibility. | • Users are able to contact the organization to request health workforce capability support. |
| Authority | A system that leverages roles of authority will have enhanced powers of persuasion. | • A health professional national college provides a statement on the importance of health workforce capability. |
| Third-party endorsements | Third-party endorsements, especially from well-known and respected sources, boost perceptions on system credibility. | • A well-respected, known, experienced rural health professional endorses the app. |
| Verifiability | Credibility perceptions will be enhanced if a system makes it easy to verify the accuracy of site content via outside sources. | • The app offers links and support by well-established services. |

[a]Source: Oinas-Kukkonen and Harjumaa [6].

**Table 5.** Persuasive design techniques: definitions and implementation examples for developing a health workforce capability app in the social support domain.

| Persuasive design technique | Persuasive design technique definition[a] | Implementation examples |
|---|---|---|
| Social learning | A person will be more motivated to perform a target behavior if they can use a system to observe others performing the behavior. | • The app includes stories of other rural health professionals who have improved their workforce capability. |
| Social comparison | System users will have a greater motivation to perform the target behavior if they can compare their performance with the performance of others. | • Users can share information real time on how to do something more efficient. |
| Normative influence | A system can leverage normative influence or peer pressure to increase the likelihood that a person will adopt a target behavior. | • The app shows that self-care is key to long-term employability by using examples of other professionals. |
| Social facilitation | System users are more likely to perform a target behavior if they discern via the system that others are performing the behavior along with them. | • Health professionals know that many other people are also participating in the app and can choose to discuss with other users. |
| Cooperation | A system can motivate users to adopt a target attitude or behavior by leveraging human beings' natural drive to cooperate. | • The app demonstrates that working in a team leads to better patient health outcomes. |
| Competition | A system can motivate users to adopt a target attitude or behavior by leveraging human beings' natural drive to compete. | • Health workforce capability scores can be used to determine the personal best in a specific area they wish to work on. |
| Recognition | By offering public recognition for an individual or group, a system can increase the likelihood that a person/group will adopt a target behavior. | • The app demonstrates the "Rural health professional of the month" and how they benefited from improving their health workforce capability. |

[a]Source: Oinas-Kukkonen and Harjumaa [6].

**Table 6.** Persuasive design techniques: definitions and implementation examples for developing a health workforce capability app in the domain of Cialdini's principles of persuasion.

| Persuasive design technique | Persuasive design technique definition[a] | Implementation examples |
|---|---|---|
| Commitment/consistency | These are a pair of interrelated attributes in the sense that people often adhere (consistently) to their significant choices (commitments). | • The app has the ability to record health workforce capability goals (commitment) and timings (consistency) so that health professionals can commit to goals. |
| Scarcity | This causes people to almost panic out of the fear that something will disappear or become unavailable, so they make an intent effort to acquire or preserve it. | —[b] |
| Social proof | This explains the human tendency to look around at others in society for reinforcement and direction in taking action. | • The app shows the number of health professionals that have joined the health workforce capability app. |
| Reciprocity | This describes the human desire to make others feel appreciated by responding in ways that return good gestures. | • The app allows users to post their own health workforce capability issues and also to respond to other users' posts. |

[a]Source: Oyebode et al [9].

[b]Not available.

A wide variety of implementation examples were drawn from analyzing the apps. In summary, the *primary task support* domain focuses on techniques that help the health professional focus on an element of health workforce capability that is important to them. The *dialogue support* techniques are about the dialogue that occurs between the health professional and the digital system to improve the health professional's ability to work on their health workforce capability. *System credibility support* techniques include using only high-quality products and services and stakeholders that are associated with the app, for example, respected health professionals, health organizations, and Australian clinical guidelines endorsed by colleges representing clinical groups. *Social support* techniques play a role in showcasing how and how many other health professionals manage their health workforce capability and facilitating collaboration and information sharing between health professionals.

# Discussion

## Principal Findings

A systematic review of 53 apps that are related to health workforce capability was performed through deconstruction of persuasive design techniques and their implementation. The findings demonstrated that health professional needs and digital solutions broadly align with the 4 health workforce capability categories used in this study: competencies and skills, health and personal qualities, values and attitudes, and work organization. These categories were matched with 32 persuasive design techniques and provided suggestions on how to further improve the persuasiveness of apps used to improve health workforce capability.

### *Persuasive Design Techniques*

Of the 53 apps, the most common persuasive design techniques were surface credibility (n=48, 90.6%) and liking (n=48), followed by trustworthiness (n=43, 81.1%), reminders (n=38, 71.7%), and suggestion (n=30, 56.6%). Social support persuasive design techniques were the least used across the different apps analyzed for health workforce capability, whereas primary task support techniques were the most common, with the exception of simulation. The apparent lower inclusion of persuasive design techniques around social support is a contrast, given that previous qualitative analyses have shown that rural health professionals perceive social and professional isolation to play a major role in building and maintaining their health workforce capability (Ramsden et al, unpublished data, 2021), and also provide multiple suggestions on how this could be achieved in this study. Examples given were related to online communities of practices, interprofessional learning (social comparison and learning), and telehealth improving trust between different disciplines, but also around having an online career coach (normative influence) and demonstrating online effectiveness of team performance and peer recognition, for example, by showcasing good news stories in health workforce capability. This study was based on the analyses by Oyebode et al [9], who reviewed 80 popular mHealth apps using the 32 techniques as described above. Briefly, the most common persuasive categories identified in Oyebode et al's [9] study were personalization (n=77, 96.3%), surface credibility (n=69, 86.3%), trustworthiness (n=66, 82.5%), self-monitoring (n=64, 80%), real-world feel (n=59, 73.8%), reminders (n=57, 71.3%), suggestion (n=56, 70%), liking (n=52, 65%), expertise (n= 52), commitment/consistency (n=47, 58.8%), reduction (n=45, 56.3%), and tunneling (n=40, 50%). The authors provided persuasive strategy suggestions that can be used by app developers to increase the likelihood of behavior change. Similar to our study, Oyebode et al [9] pointed out that social interaction can motivate people to reach their behavioral targets, yet social support is rarely used in mHealth apps (17 of 80 apps, 21%). They thus recommend for future apps to include social support strategies.

Across health workforce capability domains, we found that health apps include the most persuasive techniques, as do the apps related to values and attitudes. Interestingly, the apps related to work organization have the least persuasive techniques, even though work organization plays a large role in maintaining or building health workforce capability [13]. For

example, the ability to link to local allied health professionals through an online booking system would enhance a general practitioner's (GP) workforce capability (Ramsden et al, unpublished data, 2021) by improving interdisciplinary, holistic care. The online system would also facilitate sending and sharing reports between the GP and an allied health professional. The online booking system would also reduce the likelihood of the patient not following up with an allied health professional as they get an appointment on the spot (social facilitation persuasive design technique between the GP, allied health professional, and patient). When designing digital tools that focus on work organization to increase capability, designers could focus on including persuasive design techniques, such as real-world feel, tunneling, and tailoring, which are currently underrepresented in the work organization category.

It appears that health apps include a wide array of persuasive techniques, and thus, building a health workforce capability app could learn from the design of capability-related health aps. It should be kept in mind, however, that Oyebode et al [9] did not find a direct relationship between the number of persuasive strategies used and app effectiveness (measured by user app ratings). The authors, therefore, argued that adopting fewer persuasive design techniques could potentially be as effective, as it potentially reduces the complexity of the app and thus prevents user cognitive overload. The number of techniques used is still under debate, and more research is needed in this area.

### *Health Workforce Capability Digital Support Tool*

Khakurel et al [14] conducted a systematic review of 34 wearable device studies and concluded that wearable technology can potentially increase work efficiency for staff, improve workers' physical well-being, and reduce work-related injuries. However, they also reported that technological, social, policy, data, and economic issues related to wearable technology remain. The authors developed a categorization of wearable devices, including monitoring, assisting, augmenting, tracking, and delivering content. Alhasani et al [15] applied a similar systematic review to 60 stress apps. The authors found that most apps use manual tracking, and pointed out that this can become boring and users might forget to log data. A combination of manual entry and automated tracking sensors (hybrid data capture) is recommended to improve the likelihood of reaching a set behavioral target change. For example, accelerometer data in a mobile phone can be used to predict the stress level at work [16] through the use of a data collection app stored on the phone that measures the continuous process of recording active apps, starting and ending time stamps, and duration of app use. By linking these data with the participant's self-reported stress levels, researchers could accurately predict future stress levels. Ferdous et al [16] demonstrated an average accuracy of 75% and a precision of 85.7% as indicators of overall stress levels in work environments. This information can be used to inform stress reduction, organizational policies, and the interrelation between stress and productivity of workers. Reactions from health professionals varied according to their willingness to trial wearables, but at the same time, qualitative analyses identified that health professionals are receptive to certain elements of persuasive design techniques that relate to wearables, such as monitoring, feedback, reminders, reduction, and tunneling.

Content filtering and personalized interventions that match users' needs have been demonstrated to improve persuasion power [15], such as the persuasive design techniques analyzed in our study: tailoring, tunneling, and reduction. The authors also recommend that users can customize the user interface and app features, such as color background, or their user profile as this gives them increased self-agency, a sense of control, and identity. Persuasive design techniques such as these could be implemented within a capability-building digital solution.

We propose to use a strengths-based approach to improving rural health professionals' capability and include persuasive design techniques in digital solutions. For example, future applications can measure the health professional's vitality or health workforce capability scores at different times throughout the day by asking them to rate their capability at random times or a set time throughout the week. Alternatively, some health professionals may be willing to use a hybrid model by augmenting self-ratings with wearable technology, such as vital signs or step counts. In either case, the health professional can view graphs or trends and analyses showing self-rated health workforce capability levels over time and time of day and display simple statistics, such as minimum, maximum, and average scores. This would reflect the persuasive design technique monitoring. Moving forward, the data can potentially also be used to predict future health workforce capability and recommend specific behavioral change interventions for the health professional to allow for tunneling and reduction. Furthermore, if the health workforce capability score falls, a reminder could pop up for the health professional to take appropriate action. This could include contacting a trusted organization or coach (system credibility persuasive techniques, such as trustworthiness and real-world feel, as linkage occurs with real people) to discuss the health workforce capabilities that they feel are hindering their current capability and identify solutions to increase their capability. It could also involve the health professional doing some exercise or other activity that helps them with their well-being and feelings of capability. Other simple persuasive design techniques, such as limited ads, will increase system credibility support through increased surface credibility. Given that health workforce capability is determined by the interplay of a complex array of factors [2] (Ramsden et al, unpublished data, 2021), this may well be the first step toward building a futuristic, completely digitalized tool to improve the health workforce capability support tool.

### *Ethical Considerations*

There are many ethical aspects that needs to be considered in developing a digital solution supported by real-time support by real people [17].

First, further research is needed on the ethical aspects of this type of data capture, as this will have major security implications. Indeed, Alhasani et al [15] note that every interaction with the app creates behavioral data, via audit trails or sensors, that can be analyzed in real time to predict users' needs. This means that sensitive information about a user is being generated and stored.

Second, there is likely to be a perceived conflict of interest if an organization would own a health workforce capability digital tool and also provide the support to help health professionals with their capability.

Third, the active participation of a health profession can be strongly affected by factors such as the health professionals' perception of how the data are owned, stored, and used. There is thus a need to balance the individual privacy of the data generated with the benefits that an amalgamated data set could provide to the workforce as a whole. As discussed, the ethics of using such granular data need to be further evaluated, and an acceptability study regarding the scope of use for the collected data would be timely. Nonetheless, as alluded to before, it is key to ensure that the data are kept secure and only used for their intended purpose, whatever that is determined to be. Users must also be reassured of this and their confidence maintained.

Lastly, relying on self-rated measures may reinforce unconscious incompetence [18] or result in social desirable answers where a health professional knowingly scores themselves optimistically due to the perception of negative effects on their professional registration or professional standing. This can be mitigated by building a culture of trust and support between the organization and the health professional and a shared understanding that these analyses are being done for the benefit of the health professional.

### Limitations

Although some may see that using only apps from Google Play Store may lead to potential bias in the analyses by excluding Apple App Store products, Meacham et al [19] report that many popular apps can be found on both platforms. They report that developers perceive that it is easier to register their product with Google Play Store than with the Apple App Store. This in turn may lead to some viewing the Apple App Store as of better quality and more unlikely to be free [19]. To ensure the analyses only included higher-quality apps as perceived by the end user, our review excluded apps with user star ratings below 4.

The apps were selected based on search strategy, number of downloads, and reviewer ratings and could be viewed as a proxy for quality. We did not perform a quality assessment of the apps, as this was outside the scope of this review, nor do we endorse any of the apps analyzed. The intent of this review was to analyze the persuasive strategies used in apps that contribute to health workforce capability. Although this can be seen as a study limitation, health workforce capability is a broad and complex concept and this app review should be viewed as exploratory. Furthermore, although the script to select apps from Google Play for the review focused on selecting apps that are available in Australia, the results are transferable to other countries from a persuasive design technique viewpoint. Countries will have their own (clinical) guidelines that may be important for health workforce capability.

It should be considered that the numbers of apps analyzed was relatively small and is therefore only an exploratory investigation of the persuasive design principles by the health workforce capability domain. Nonetheless, the strength lies in the combination of looking at the persuasive design techniques used and in relation to their health workforce capability needs.

A further limitation of the study is that it was beyond the scope to draw conclusions about the impact the number of persuasive techniques included has on app effectiveness. Further research is needed to explore this in relation to health workforce capability apps.

### Conclusion

There are many persuasive design techniques that can assist in building capability. Commonly used techniques are surface credibility, liking, trustworthiness, reminders, and suggestion, while less common are social support persuasive design techniques. Additionally, several apps are available in the market that can assist in improving health workforce capability. There is, however, a specific lack of digital, real-time support to improve health workforce capability. Social support strategies through using social support persuasive design techniques will need to be integrated more prominently into a health workforce capability app. An application to measure and monitor health workforce capability scores can be used in conjunction with direct real-world person and real-time support to discuss and identify solutions to improve health workforce capability for rural and remote health professionals who are at high risk of burnout or leaving the rural health workforce.

## Acknowledgments

We thank NIB Foundation for contributing part funding for the analyses in this study. The funder had no impact on the study design, interpretation, or outcomes of this study. We thank Katie Carlisle for contributing to study design and project management.

## Authors' Contributions

SP, RR, and RC designed the study. SP and RR reviewed the apps for inclusion, with input from AT. WF and OH completed the data collection in collaboration with AT. SP, RR, AT, and KP interpreted study results, with input from RC, OH, and WF. SP and AT drafted the initial manuscript, with major input from RR and KP. ME and BE provided content expertise. All authors contributed to developing the study materials and writing the manuscript, and read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Selection process.
[DOCX File , 18 KB - mhealth_v10i2e33413_app1.docx ]

## References


1. Walsh S, Lyle DM, Thompson SC, Versace VL, Browne LJ, Knight S, et al. The role of national policies to address rural allied health, nursing and dentistry workforce maldistribution. Med J Aust 2020 Dec;213(Suppl 11):S18. [doi: 10.5694/mja2.50881] [Medline: 33314144]
2. Ramsden R, Colbran R, Christopher E, Edwards M. The role of digital technology in providing education, training, continuing professional development and support to the rural health workforce. HE 2021 Feb 18;ahead-of-print(ahead-of-print):1-24 [FREE Full text] [doi: 10.1108/he-11-2020-0109]
3. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
4. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
5. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
6. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. CAIS 2009;24(Article 28):485-500. [doi: 10.17705/1CAIS.02428]
7. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L, Vanden Abeele V. A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. JMIR Mhealth Uhealth 2016 Oct 13;4(4):e118 [FREE Full text] [doi: 10.2196/mhealth.6286] [Medline: 27742604]
8. Thach KS, Phan TP. Persuasive design principles in mental health apps: a qualitative analysis of user reviews. 2019 Mar 20 Presented at: 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF); Mar 20-22, 2019; Danang, Vietnam p. 1-6. [doi: 10.1109/RIVF45261.2019]
9. Oyebode O, Ndulue C, Alhasani M, Orji R. Persuasive mobile apps for health and wellness: a comparative systematic review. 2020 Presented at: International Conference on Persuasive Technology; April 20, 2020; Aalborg, Denmark p. 163-181. [doi: 10.1007/978-3-030-45712-9_13]
10. Cialdini RB. Harnessing the science of persuasion. Harv Bus Rev 2001;79(9):72-81.
11. Hines J. Doctoral dissertation. An Analysis of the Use of Robert B. Cialdini's Six Persuasive Techniques in YouTube Videos about Endurance Training/Life Balance. Edwardsville: Southern Illinois University; 2021. URL: https://media.proquest.com/media/hms/PFT/2/CPzXJ?_s=CVSocBanEfzOZEJU7zzeOzYPX9c%3D [accessed 2021-11-20]
12. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
13. Smyth J, Pit SW, Hansen V. Can the work ability model provide a useful explanatory framework to understand sustainable employability amongst general practitioners: a qualitative study. Hum Resour Health 2018 Jul 24;16(1):32 [FREE Full text] [doi: 10.1186/s12960-018-0292-x] [Medline: 30041658]
14. Khakurel J, Melkas H, Porras J. Tapping into the wearable device revolution in the work environment: a systematic review. ITP 2018 May 14;31(3):791-818 [FREE Full text] [doi: 10.1108/itp-03-2017-0076]
15. Alhasani M, Mulchandani D, Oyebode O, Orji R. A systematic review of persuasive strategies in stress management apps. 2020 Presented at: Persuasive 2020 - Behavior Change Support Systems (BCSS 2020); April 2020; Denmark URL: http://ceur-ws.org/Vol-2662/BCSS2020_paper4.pdf
16. Ferdous R, Osmani V, Mayora O. Smartphone apps usage patterns as a predictor of perceived stress levels at workplace. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); May 20, 2015; Istanbul, Turkey p. 225-228. [doi: 10.4108/icst.pervasivehealth.2015.260192]
17. Jacobs N. Two ethical concerns about the use of persuasive technology for vulnerable people. Bioethics 2020 Jun;34(5):519-526 [FREE Full text] [doi: 10.1111/bioe.12683] [Medline: 31617216]
18. de Weerd H, Degens N. Putting the long-term into behavior change. 2018 Presented at: International Conference on Games and Learning Alliance; December 2018; Palermo, Italy p. 71-81 URL: http://www.harmendeweerd.nl/papers/20181001_Long_term_behavior_change.pdf [doi: 10.1007/978-3-030-11548-7_7]
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]


## Abbreviations

**GP:** general practitioner
**PSD-Model:** Persuasive System Design Model
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 07.09.21; peer-reviewed by G Margellis, U Bork; comments to author 28.09.21; revised version received 29.09.21; accepted 03.10.21; published 07.02.22.*

*Please cite as:*
*Pit SW, Tan AJH, Ramsden R, Payne K, Freihaut W, Hayes O, Eames B, Edwards M, Colbran R*
*Persuasive Design Solutions for a Sustainable Workforce: Review of Persuasive Apps for Real-Time Capability Support for Rural Health Care Professionals*
*JMIR Mhealth Uhealth 2022;10(2):e33413*
*URL: https://mhealth.jmir.org/2022/2/e33413*
*doi:10.2196/33413*
*PMID:35129447*

Review

# Use of Mobile Apps for Visual Acuity Assessment: Systematic Review and Meta-analysis

Lingge Suo[1], MD; Xianghan Ke[1], MD; Di Zhang[1], MD; Xuejiao Qin[2], MD, PhD; Xuhao Chen[1], MD; Ying Hong[1], MD, PhD; Wanwei Dai[1], PhD; Defu Wu[1], MD; Chun Zhang[1*], MD, PhD; Dongsong Zhang[3*], PhD

[1]Department of Ophthalmology, Peking University Third Hospital, Beijing, China

[2]Department of Ophthalmology, Qilu Hospital of Shandong University, Jinan, China

[3]Department of Business Information Systems and Operations Management, University of North Carolina at Charlotte, Charlotte, NC, United States

[*]these authors contributed equally

**Corresponding Author:**
Dongsong Zhang, PhD
Department of Business Information Systems and Operations Management
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC, 28223-0001
United States
Phone: 1 704 687 1893
Email: dzhang15@uncc.edu

## Abstract

**Background:** Vision impairments (VIs) and blindness are major global public health issues. A visual acuity (VA) test is one of the most crucial standard psychophysical tests of visual function and has been widely used in a broad range of health care domains, especially in many clinical settings. In recent years, there has been increasing research on mobile app–based VA assessment designed to allow people to test their VA at any time and any location.

**Objective:** The goal of the review was to assess the accuracy and reliability of using mobile VA measurement apps.

**Methods:** We searched PubMed, Embase, Cochrane Library, and Google Scholar for relevant articles on mobile apps for VA assessment published between January 1, 2008, and July 1, 2020. Two researchers independently inspected and selected relevant studies. Eventually, we included 22 studies that assessed tablet or smartphone apps for VA measurement. We then analyzed sensitivity, specificity, and accuracy in the 6 papers we found through a meta-analysis.

**Results:** Most of the 22 selected studies can be considered of high quality based on the Quality Assessment of Diagnostic Accuracy Studies–2. In a meta-analysis of 6 studies involving 24,284 participants, we categorized the studies based on the age groups of the study participants (ie, aged 3-5 years, aged 6-22 years, and aged 55 years and older), examiner (ie, professional and nonprofessional examiners), and the type of mobile devices (ie, smartphone, iPad). In the group aged 3 to 5 years, the pooled sensitivity for VA app tests versus clinical VA tests was 0.87 (95% CI 0.79-0.93; *P*=.39), and the pooled specificity was 0.78 (95% CI 0.70-0.85; *P*=.37). In the group aged 6 to 22 years, the pooled sensitivity for VA app tests versus clinical VA tests was 0.86 (95% CI 0.84-0.87; *P*<.001), and the pooled specificity for VA app tests versus clinical VA tests was 0.91 (95% CI 0.90-0.91; *P*=.27). In the group aged 55 years and older, the pooled sensitivity for VA app tests versus clinical VA tests was 0.85 (95% CI 0.55-0.98), and the pooled specificity for VA app tests versus clinical VA tests was 0.98 (95% CI 0.95-0.99). We found that the nonprofessional examiner group (AUC 0.93) had higher accuracy than the professional examiner group (AUC 0.87). In the iPad-based group, the pooled sensitivity for VA app tests versus clinical VA tests was 0.86, and the pooled specificity was 0.79. In the smartphone-based group, the pooled sensitivity for VA app tests versus clinical VA tests was 0.86 (*P*<.001), and the pooled specificity for VA app tests versus clinical VA tests was 0.91 (*P*<.001).

**Conclusions:** In this study, we conducted a comprehensive review of the research on existing mobile apps for VA tests to investigate their diagnostic value and limitations. Evidence gained from this study suggests that mobile app–based VA tests can be useful for on-demand VI detection.

***(JMIR Mhealth Uhealth 2022;10(2):e26275)*** doi:10.2196/26275

**KEYWORDS**

smartphone; iPad; eye screening; visual acuity; app; meta-analysis

## Introduction

Vision impairments (VIs) and blindness are a major global public health issue [1]. In 2020, the estimated number of people with distance VI in the world was 596 million, including 43 million with blindness [2]. A large proportion of those affected (90%) live in low- and middle-income countries. VI can be preventable or treatable for approximately 90% of people with VI by using highly cost-effective interventions. In low-income countries, diagnosis, monitoring, and treatment of vision problems are challenging, largely attributable to insufficient eye care professionals [3]. In high-income countries, there are also barriers to eye screening and patient compliance. In particular, there is often time pressure in primary consultations for diagnosing ophthalmic problems [4]. Thus, there is a need for ubiquitous, self-manageable, and automated tools for visual acuity (VA) tests to increase early detection and timely assistance for people with VIs [5].

To address the lack of eye care professionals and reduce the cost for eye screening, an increasing amount of research effort has been dedicated for building efficient eye screening tools and methods by leveraging mobile devices (eg, smartphones) and technologies [6]. In both high- and low-income countries, the continuous growth of mobile device ownership has propelled mobile health (mHealth) interventions [7]. Mobile technologies provide point-of-care tools for real-time patient monitoring, patient data collection, health information delivery, and telemedicine throughout the world [8]. Free and paid mHealth apps have demonstrated notable success in detecting ophthalmic diseases [9], and the number of mobile apps intended to address eye care issues has been increasing. The VA test, a vision test often performed by an optometrist or ophthalmologist to measure a person's ability to see an object from 20 feet away, is one of the most crucial standard psychophysical tests for assessing visual function [10]. It is also a measurement of eye treatment effectiveness and changes in central vision over time in clinical settings [11].

Traditional clinical ophthalmic equipment is cumbersome and difficult to transport, lacks mobility, and requires trained ophthalmic professionals. Therefore, an automated, accurate, and user-friendly approach is needed for vision screening or self-monitoring. Some studies have proposed novel mobile device–based techniques for a VA test [12]. Bastawrous et al [13] conducted the Early Treatment Diabetic Retinopathy Study (ETDRS) that proposed the Peek Acuity mobile app, which was validated against Snellen charts. ETDRS charts were used as part of a survey about epidemiologic eyes among adults in central Kenya. Peek Acuity is a Logarithm of the Minimum Angle of Resolution (logMAR)-style smartphone-based vision test. Using a fast-testing algorithm, it is capable of measuring VA at a clinically acceptable time, with greater reliability and precision than those using Snellen charts [14]. It also allows individuals to choose from multiple types of visual charts for VA assessment, such as the Snellen [15], ETDRS [16], and Tumbling E [17] charts. Thus, the Peek Acuity app provides an advantage over traditional logMAR acuity measurement.

Although current modern mobile devices with high-resolution screens offer novel, ubiquitous, and portable vehicles for VA tests, it remains unclear whether existing VA test apps are effective for ophthalmic disease diagnosis and management. In this study, we conducted a comprehensive review of the research to investigate the diagnostic accuracy and limitations of existing mobile VA assessment apps for detecting VI. Based on age, test examiner, and type of mobile devices, we categorized existing studies into different categories and performed subgroup analysis. Furthermore, 4 variables (ie, publication year, sample size, mobile device, and examiner) were selected in the multivariate meta-regression.

## Methods

### Data Sources and Search Strategy

We searched PubMed, Embase, Cochrane Library, and Google Scholar for relevant articles on mobile apps for VA testing published between January 1, 2008, and July 1, 2020. The literature search used the following terms, as well as their different combinations, as the search keywords: smartphone, iPhone, iPad, phone, tablet, mobile devices, visual acuity, VA, eye screening, app, application, Snellen chart, Tumbling E chart, Early Treatment Diabetic Retinopathy Study chart, and ETDRS chart (Multimedia Appendix 1).

### Inclusion and Exclusion Criteria

We applied several inclusion criteria when identifying relevant studies for this research. A study would be considered relevant if it (1) was to evaluate VA via a smartphone or tablet app, (2) used an acceptable VA reference standard, (3) was written in English, and (4) was published after 2008. Exclusion criteria included (1) studies where the number of participants with VI was fewer than 10 and (2) literature review or commentary articles, short communications, or case reports.

### Data Collection

Two authors and a research assistant extracted information from the identified studies, including the study design, sample size, participant characteristics, nature of eye screening, mobile techniques (eg, functions and features of smartphones and tablet app), and main research results (eg, true positives, false positives, true negatives, and false negatives). The extracted information was reviewed and verified by other coauthors.

### Risk of Bias and Quality Assessment

Two researchers specializing in eye care independently reviewed each selected article and assessed its quality by using the Quality Assessment of Diagnostic Accuracy Studies–2 scores (QUADAS-2) tool. They discussed and resolved disagreements in their scores with other coauthors through a face-to-face meeting. Among included studies, the risk of bias was evaluated in 4 aspects by the QUADAS-2 tool: patient selection, index test, reference standard, and flow and timing. For applicability

concerns, we assessed patient selection, index test, and reference standard as low, high, or unclear.

### Statistical Analysis

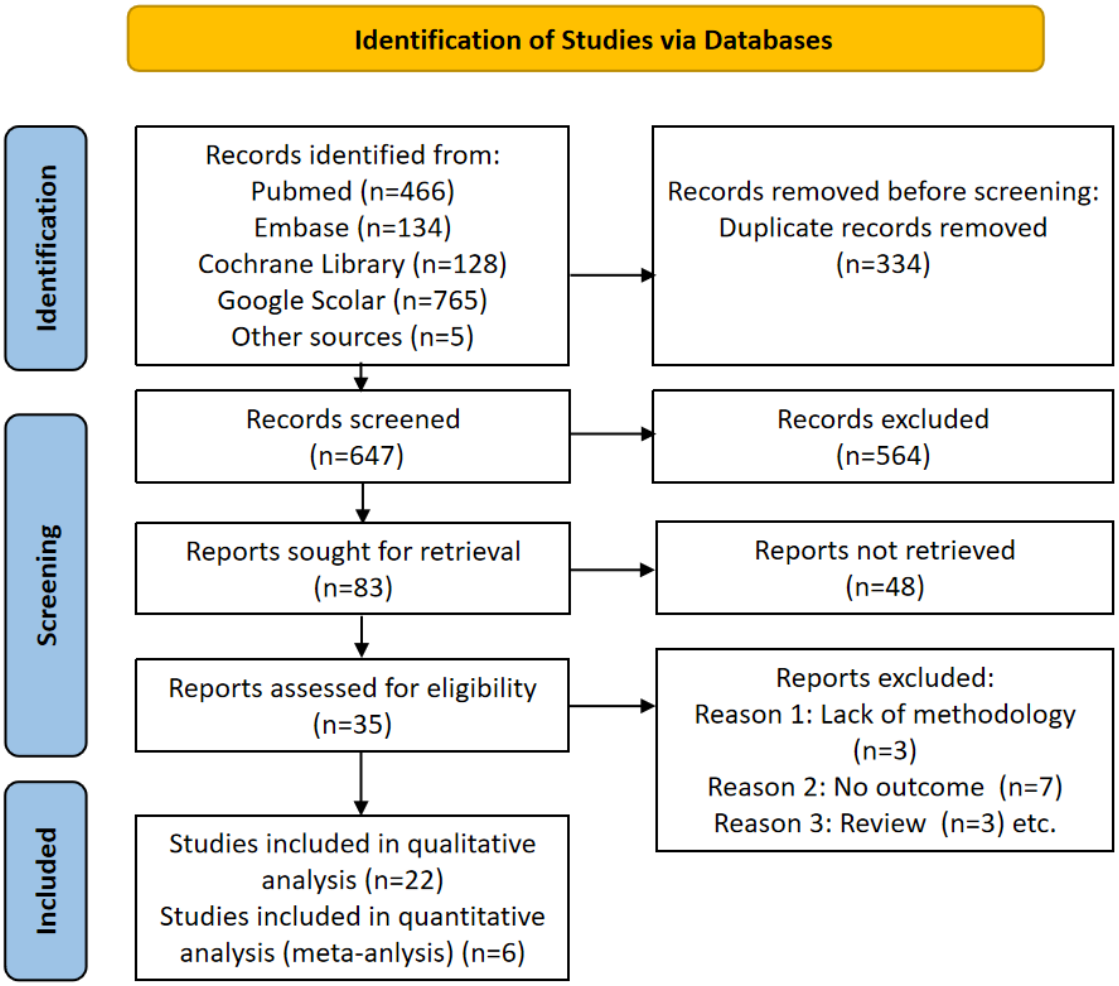
We used chi-square and $I^2$ values of sensitivity, specificity, likelihood ratio tests, and diagnostic odds ratio (DOR) to evaluate heterogeneity. Heterogeneity was evaluated by Cochrane Q-test ($I^2$ value); heterogeneity was considered to exist when $P$>.10. When $I^2$ results were ≤50%, a fixed effects model was used; otherwise, a random effects model was used. The values of DOR ranges from 0 to infinity, with 0 indicating no test discrimination. Higher scores indicate better discrimination. The sensitivity, specificity, likelihood ratio positive (LR+), likelihood ratio negative (LR–), and DOR of each age, examiner, and mobile device type subgroup were calculated using a random effects model given the high expected heterogeneity. We performed meta-regression to explore whether the sources of heterogeneity could be explained by some methodological factors (eg, year of publication, sample size, mobile device, examiner) and characteristics of study samples. We also constructed a summary receiver operating characteristic (SROC) curve using the Moses constant of linear mode model. We used Meta-DiSc software (version 1.4, Ramón y Cajal Hospital, Madrid, Spain) for meta-analysis and Review Manager (version 5.3, Cochrane Collaboration) for paper quality assessment. Extracted data were synthesized by creating forest plots of sensitivity and specificity.

## Results

### Search Results

Our literature search yielded a total of 981 papers. After our review of the titles and abstracts, 959 studies were excluded either because of duplication or lack of adherence to our topic, resulting in 22 full-text articles for quality assessment (Figure 1). We also checked their reference lists for further relevant studies and retrieved additional studies. Finally, 6 full-text studies met inclusion criteria for quantitative analysis (meta-analysis). Figure 1 presents the flowchart of our systematic literature search. The selected studies were conducted in more than 11 countries.

**Figure 1.** Flowchart of systematic literature search.



### Characteristics of the Studies

Among the 22 identified studies, 4 studies were population-based, 10 were observational, 3 were cross-sectional, 4 were prospective, and 1 was a validation study (Table S1 in Multimedia Appendix 2 and Multimedia Appendix 3). Nine studies assessed the performance of tablet-based (ie, iPad) VA measurement apps, and 13 studies assessed the performance of smartphone-based apps. A total of 25 mobile apps were evaluated in those 22 enrolled studies.

### Characteristics of Study Samples

Among the 22 identified studies, the average number of participants per study was 1648, ranging from 43 [18] to 10,579 participants [17]. The age of the participants in those studies varied significantly from 3 to 89 years. Five studies included children aged younger than 18 years [17,19-22], 10 studies included middle-aged (19-55 years) adults [5,9,12,18,23-28], and the remaining 7 studies included older adults (aged 55 years and older) [7,13,14,29-31]. The main demographic characteristics of the participants in each study are summarized in Table S1 in Multimedia Appendix 2 and Multimedia Appendix 3.

### Eye Screening with VA apps

Among the 22 selected papers, 9 studies used iPads for VA tests [5,18,21,23,28-30,32,33], and 13 used smartphones [7,9,12-14,17,19,20,22,24,25,27,31]. Six studies used the Peek Acuity app (Tumbling E chart) [7,13,17,19,20,22], 2 used the Sightbook app (Snellen chart) [31,32], 2 used the Eye Chart Pro app (Snellen chart) [5,28], and 1 [12] evaluated 11 different VA apps (eg, Eye Test app, OptOK app). The rest of the studies investigated other VA apps.

A total of 10 studies used the Tumbling E chart in mobile apps. A Tumbling E chart, also known as an E chart, is useful for patients who are unable to read the Latin alphabet (eg, very young children). The chart contains multiple rows of the letter E in various rotations and with decreasing sizes. Patients were asked to state where the limbs of the E were pointing (up, down, left, or right). Depending on how far a patient can see, his or her VA can be quantified. The Tumbling E chart shares the same principle as the Snellen distant vision chart [7,9,13,17-20,22,28,30].

Three studies used the ETDRS chart for VA measurement in mobile apps. A logMAR chart, also called a Bailey-Lovie chart or an ETDRS chart, is a chart consisting of rows of letters used by ophthalmologists, orthoptists, optometrists, and vision scientists to estimate VA. The chart was developed by the National Vision Research Institute of Australia in 1976 to enable a more accurate estimate of VA than other charts (eg, the Snellen chart). For this reason, the logMAR chart has been recommended, particularly in a research setting [14,23,29].

Five studies used the Snellen chart in VA measurement apps. A Snellen chart is another eye chart used to measure VA. Snellen charts are named after the Dutch ophthalmologist Herman Snellen, who developed the chart in 1862. The normal Snellen chart is printed with eleven lines of block letters, also known as optotypes. The first line consists of a single very large letter. Subsequent rows have increasing numbers of letters that decrease in size. A person taking the test covers one eye from 6 meters or 20 feet away and reads the letters of each row from the top to the bottom. The rest of the studies used other types of charts for VA measurement, such as Landolt C or Numbers [5,12,18,27,32].

The 22 selected papers measured VA at a range from 36 centimeters to 6 meters: 8 studies measured VA at a 2-meter testing distance [13,14,19,20,29]; 3 studies measured VA at a 3-meter testing distance [9,21,26]; and the remaining 11 studies measured VA at 36 cm, 40 cm, 1 m, 1.2 m, 2 m, 14 inches, 20 feet, 4 m, and 6 m.

### Meta-analysis

Table 1 shows the characteristics and findings of 6 meta-analyses on mobile apps for VA testing. Table 2 shows the summary of 6 meta-analyses that examined mobile apps for VA testing. We categorized studies based on the age of their participants, examiner, and the type of mobile devices. As summarized in Multimedia Appendix 4, 4 variables (ie, publication year, sample size, mobile device, examiner) were selected in the multivariate meta-regression (sensitivity); however, none of those variables was significantly associated with the detected heterogeneity, as shown in Multimedia Appendix 5.

**Table 1.** Main characteristics and findings of 6 meta-analyses.

| Source | Study design | Age, year | Sample size (P/E[a]) | Mobile device type | App name | App description | TD[b] | Main results |
|---|---|---|---|---|---|---|---|---|
| Nik Azis et al [21], Malaysia | Cross-sectional study | 5-6 | 195/290 | iPad mini | AAPOS[c] Vision Screening | Lea symbols chart | 3 m | Sensitivity: 82.1% (right vision), 82.1% (left vision); Specificity: 81.3% (right vision), 76.9% (left vision) |
| Rono et al [17], Kenyan | Population-based study | 11.5, 11.7 | S[d]:10,284/S:20,568, P[e]:10,579/P:21,158 | Samsung Galaxy S3 | Peek Acuity | Tumbling E chart | 2 m | Sensitivity: 76.9% (64.8%-86.5%), Specificity: 90.8% (89.3%-92.1%) |
| Zhao et al [22], US | Prospective study | 3-17 | 106/212 | Samsung Galaxy S3 SGH-i747 | Peek Acuity | Tumbling E chart | 2 m | Sensitivity: 83%-86% for decreased vision, Sensitivity: 69%-83% for referable ocular disease |
| de Venecia et al [20], US | Observational study | 6-17 | 393/190 | Samsung Galaxy A3 | Peek Acuity | Tumbling E chart | 2 m | Sensitivity: 48%, Specificity: 83% |
| Bastawrous et al [13], UK | Population-based study | 55-97 | 233/466 | Galaxy S3 GT-I9300 | Peek Acuity | Tumbling E chart | 2 m | Sensitivity: 84.6% (95% CI 54.5%-97.6%); Specificity: 97.7% (95% CI 94.8%-99.3%) |
| Andersen et al [19], Botswana | Population-based study | 6-22 | 12,877/– | Android phones | Peek Vision | Tumbling E chart | 2 m | Sensitivity: 91.6%; Specificity: 90.7% |

[a]P/E: participant/eye.

[b]TD: test distance.

[c]AAPOS: American Association for Pediatric Ophthalmology and Strabismus.

[d]S: standard group.

[e]P: peek group.

**Table 2.** A summary of mobile apps for evaluating visual acuity.

| Types | Sensitivity | Specificity | Positive LR[a] | Negative LR | Diagnostic OR[b] | AUC[c] |
|---|---|---|---|---|---|---|
| **Examiners** | | | | | | |
| Professional | 0.72 (0.66-0.79) | 0.80 (0.71-0.85) | 3.81 (2.87-5.06) | 0.30 (0.10-0.90) | 12.25 (4.33-34.71) | 0.87 (0.83-0.91) |
| Nonprofessional | 0.87 (0.85-0.89) | 0.91 (0.90-0.91) | 8.66 (8.62-10.98) | 0.17 (0.08-0.34) | 54.60 (21.98-135.59) | 0.93 (0.86-1.00) |
| **Patient age (years)** | | | | | | |
| 3-5 | 0.87 (0.79-0.93) | 0.78 (0.70-0.85) | 3.93 (2.82-5.46) | 0.17 (0.10-0.28) | 24.01 (11.95-48.22) | —[d] |
| 6-22 | 0.86 (0.84-0.87) | 0.91 (0.90-0.91) | 8.04 (6.49-9.98) | 0.25 (0.10-0.66) | 25.47 (9.02-71.94) | 0.96 (0.92-0.99) |
| ≥55 | 0.85 (0.55-0.98) | 0.98 (0.95-0.99) | 37.23 (15.18-91.30) | 0.16 (0.04-0.56) | 236.50 (41.17-1358.49) | — |
| **Mobile devices** | | | | | | |
| iPads | 0.86 (0.76-0.92) | 0.79 (0.71-0.86 | 4.14 (2.85-6.01) | 0.18 (0.11-0.31) | 22.96 (10.69-49.30) | — |
| Smartphones | 0.86 (0.84-0.87) | 0.91 (0.90-0.91) | 8.01 (6.21-10.32) | 0.23 (0.09-0.54) | 33.86 (13.02-88.06) | 0.92 (0.81-1.00) |

[a]LR: likelihood ratio.

[b]OR: odds ratio.

[c]AUC: area under the curve.

[d]Not applicable.

## Meta-analysis of Mobile Apps for VA Testing With Different Age Groups

The 6 meta-analyses involved 24,089 participants. For the group aged 3 to 5 years (230 participants), we used a fixed effects model. The pooled sensitivity was 0.87 (95% CI 0.79-0.93; *P*=.39), and the pooled specificity was 0.78 (95% CI 0.70-0.85; *P*=.37; Figure 2A). LR+ was 3.93 (95% CI 2.82-5.46), LR– was 0.17 (95% CI 0.10-0.28), and DOR was 24.01 (95% CI 11.95-48.22).

For the group aged 6 to 22 years (23,626 participants), a random effects model was chosen within these studies because of the significant heterogeneity ($P<.10$, $I^2>50\%$). The pooled

sensitivity was 0.86 (95% CI 0.84-0.87; *P*<.001), and the pooled specificity was 0.91 (95% CI 0.90-0.91; *P*=.27; Figure 2B). LR+ was 8.04 (95% CI 6.49-9.98), LR– was 0.25 (95% CI 0.10-0.66), DOR was 25.47 (95% CI 9.02-71.94), and AUC (area under the curve) was 0.96 (95% CI 0.92-0.99; Figure 3D).

In the group aged 55 years and older (233 participants), we used a random effects model. The pooled sensitivity was 0.85 (95% CI 0.55-0.98), and the pooled specificity was 0.98 (95% CI 0.95-0.99). LR+ was 37.23 (95% CI 15.18-91.30), LR– was 0.16 (95% CI 0.04-0.56), and DOR was 236.50 (95% CI 41.17-1358.49).

**Figure 2.** Summary of sensitivity and specificity of meta-analysis studies with different age groups: (A) age 3-5 years (230 participants) and (B) age 6-22 years (23,626 participants).

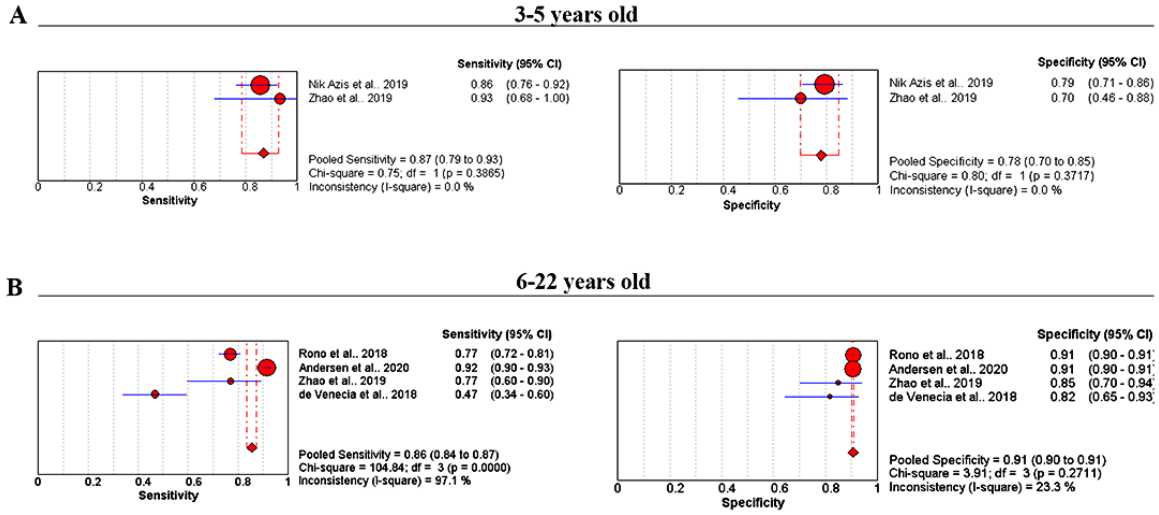
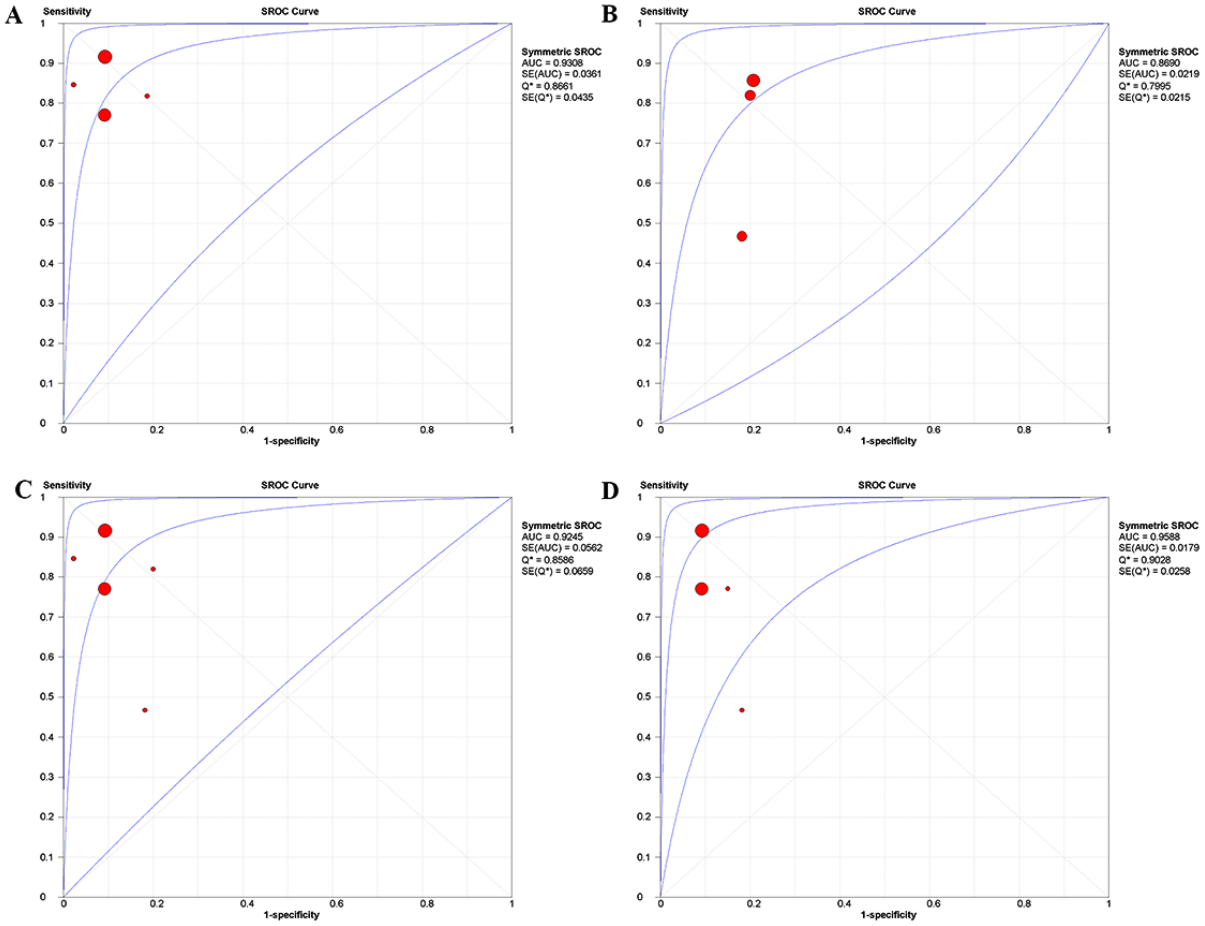



**Figure 3.** Summary receiver operating characteristic curves for study groups included in the meta-analysis: (A) nonprofessional examiners, (B) professional examiners, (C) smartphone-based, and (D) age 6-22 years.

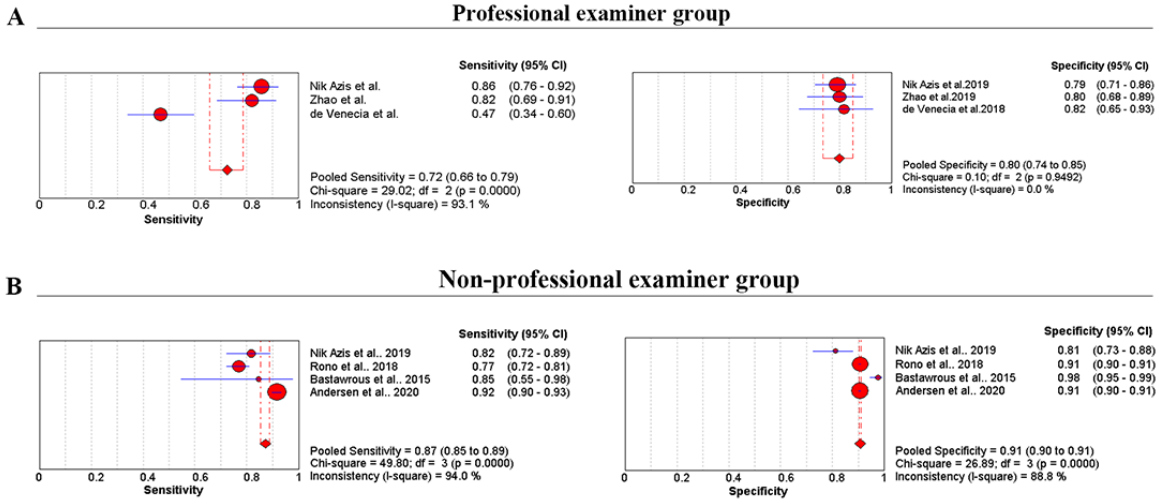


## Meta-analysis of Mobile Apps for VA Testing With Different Examiners

Data from 24,284 participants were analyzed in the 6 meta-analyses. For the professional examiner group (400 participants), we deployed a fixed effects model. The pooled sensitivity was 0.72 (95% CI 0.66-0.79; *P*<.001), and the pooled specificity was 0.80 (95% CI 0.71-0.85; *P*=.95; Figure 4A). LR+ was 3.81 (95% CI 2.87-5.06), LR– was 0.30 (95% CI 0.10-0.90), DOR was 12.25 (95% CI 4.33-34.71), and AUC was 0.87 (95% CI 0.83-0.91; Figure 3B).

In the nonprofessional examiner group (23,884 participants), we deployed a random effects model. The pooled sensitivity was 0.87 (95% CI 0.85-0.89; *P*<.001), and the pooled specificity was 0.91 (95% CI 0.90-0.91; *P*<.001; Figure 4B). LR+ was 8.66 (95% CI 8.62-10.98), LR– was 0.17 (95% CI 0.08-0.34), DOR was 54.60 (95% CI 21.98-135.59), and AUC was 0.93 (95% CI 0.86-1.00; Figure 3A).

**Figure 4.** Summary of sensitivity and specificity of the meta-analysis studies with different examiner groups: (A) professional examiners and (B) nonprofessional examiners.

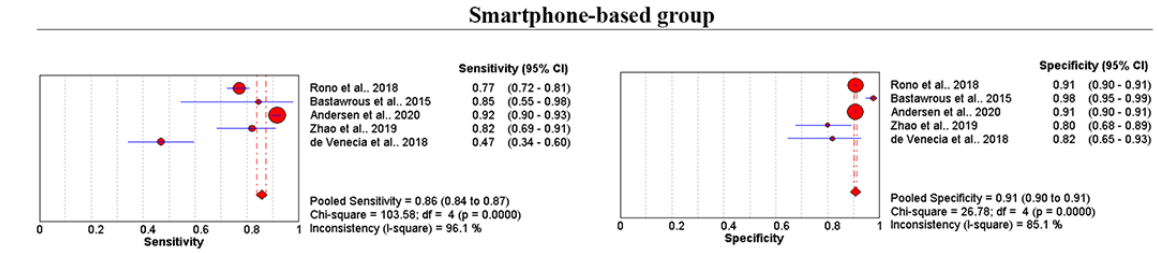



## Meta-analysis of Mobile Apps for VA Testing With Different Mobile Devices

Data from 24,284 participants were analyzed in the 6 studies, which used mobile apps either on iPads or smartphones. We used a random effects model for both iPad and smartphone groups. In the iPad-based group (195 participants), the pooled sensitivity of eyes was 0.86 (95% CI 0.76-0.92), and the pooled specificity was 0.79 (95% CI 0.71-0.86). LR+ was 4.14 (95% CI 2.85-6.01), LR– was 0.18 (95% CI 0.11-0.31), and DOR was 22.96 (95% CI 10.69-49.30).

In the smartphone-based group (24,089 participants), the pooled sensitivity was 0.86 (95% CI 0.84-0.87; *P*<.001), and the pooled specificity was 0.91 (95% CI 0.90-0.91; *P*<.001; Figure 5). LR+ was 8.01 (95% CI 6.21-10.32), LR– was 0.23 (95% CI 0.09-0.54), DOR was 33.86 (95% CI 13.02-88.06), and AUC was 0.92 (95% CI 0.81-1.00; Figure 3C).

**Figure 5.** Summary of sensitivity and specificity of the smartphone-based group included in the meta-analysis.




## Study Quality Assessment

We assessed the quality of the 22 included studies by using the QUADAS-2 tool (Figures 6 and 7). Most studies were of high quality with low risk of bias and applicability concerns. Because of the nature of mobile apps, participant blinding was not always feasible in trials. We considered studies at high risk of bias when they involved participants without clear diagnosis of VI or when there was no strict standardization in the VA examination process and examination conditions.

**Figure 6.** Risk of bias and applicability concerns of studies included in the literature review.

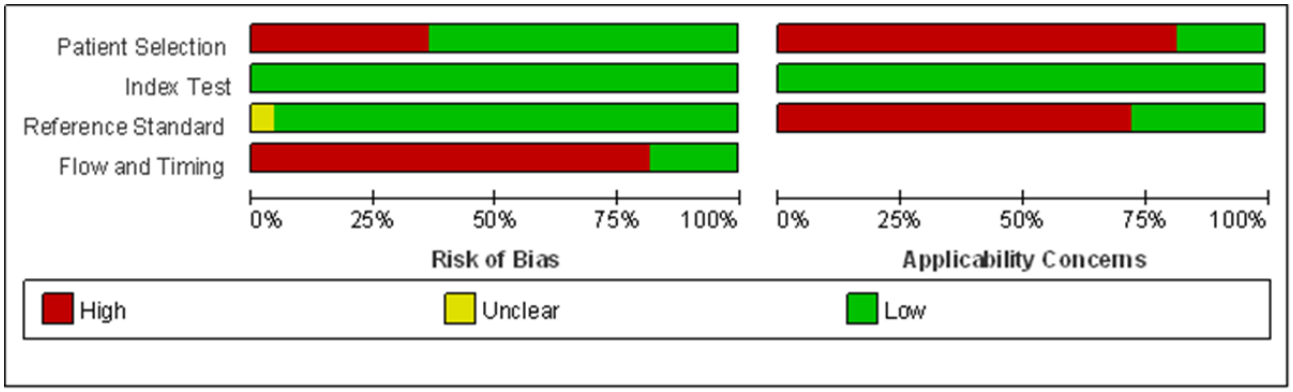
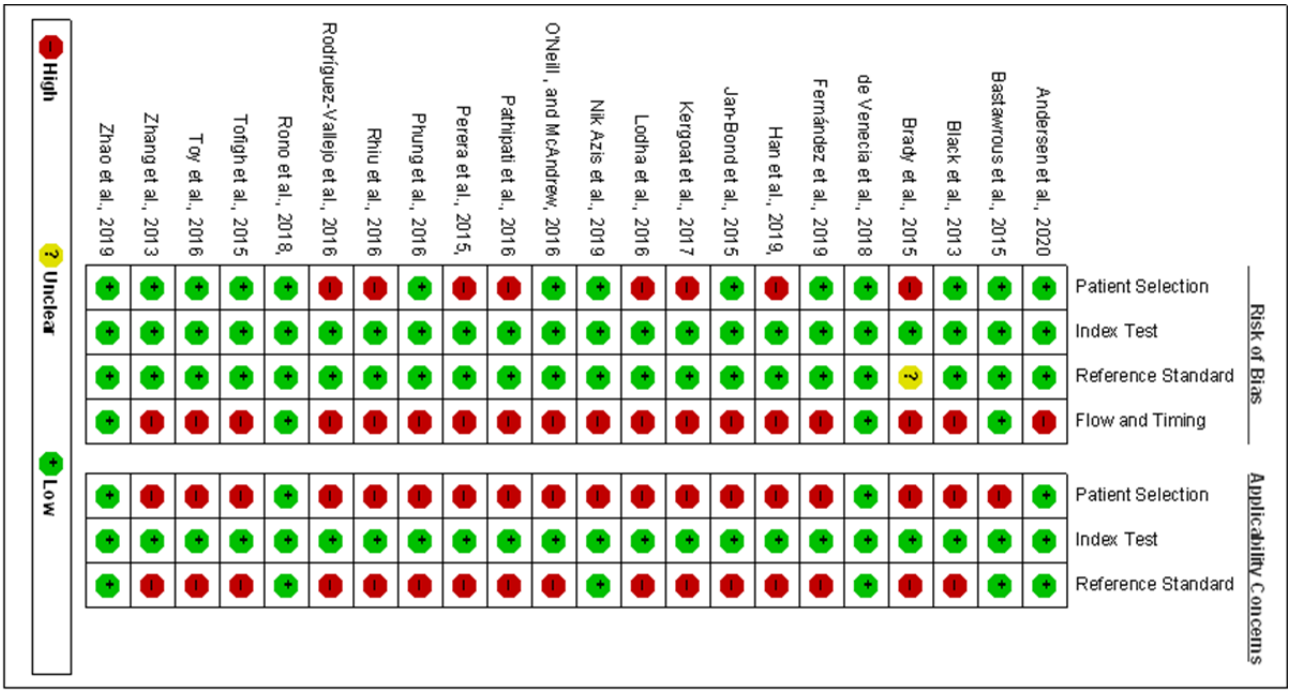



**Figure 7.** Quality of the included studies assessed via the Quality Assessment of Diagnostic Accuracy Studies–2 tool.




## Discussion

### Principal Findings

With increasing smartphone and tablet penetration, mobile VA apps provide a good quality, repeatable, objective, and cost-effective approach to vision test for eye screening. For low-income countries, with most of the world's blind people, the effective tools and techniques to improve early detection and appropriate referral are critical to reducing VI [19]. A VA test is fundamental to evaluating visual function. Accurate assessment of VA depends heavily on factors such as viewing distance, chart illumination, type of eye chart used, and scoring technique used [23].

Mobile devices, which are portable and equipped with a high-resolution screen, provide a novel platform for VA testing. An increasing number of mobile apps for VA testing have been developed that can be downloaded to different mobile platforms. Most of them, however, have not been evaluated for accuracy and reliability for following a reference standard. In our study, we aimed to systematically review and evaluate the accuracy of mobile apps for VA testing. In general, our analysis reveals that those apps investigated in the selected studies performed well in VA testing. They had different levels of accuracy for different participant age groups, between professional and nonprofessional examiners, and between apps for iPad and apps for smartphones. We observed increasing sensitivity, specificity, and DOR of those mobile apps as participant age increased (Table 2).

For the eye chart selection, based on our literature review, Peek Acuity (Tumbling E chart) [33,34], Sightbook app (Snellen chart), and Eye Chart Pro app (Snellen chart) are the most commonly used methods for testing VA. Notably, the apps with a fast-testing algorithm completed VA tests with greater reliability and precision. Examiners should select age-appropriate standardized charts, randomize letters or optotypes, vary screen illumination, adjust the size of letters or optotypes, and store or transmit the VA data collected. Compared with smartphones, tablets have larger screens with higher resolutions able to display an assortment of letters and optotype (test symbol) charts at both high and low contrasts. However, we found that VA testing via smartphones have better performance than those via iPads. Nik Azis et al [21] used iPads

to test VA of children aged 5 to 6 years with a 3-meter distance using the Lea symbols chart. The other 5 studies using smartphones [13,17,19,20,22] tested the participants with 2-meter distance using the Tumbling E chart. It is uncertain if the differences in test performance between apps for iPad and apps for smartphones are caused by the test distance. Therefore, using different VA charts with different testing distances may influence the performance of iPad-based apps. Further studies are needed to identify the cause of variable VA testing performance.

Early identification and management of children with VI is important [35] as nearly 19 million children in the world live with this condition. The World Health Organization suggests that refractive errors are one of the most common causes of these impairments, especially for children in low-income countries [4]. Mobile apps that provide VA tests can address this issue and make VA testing easier. In our 6 reviewed meta-analyses, 5 concentrated on children's vision screening [17,19-22]. Among those studies, 4 [17,19,20,22] tested VA using Android smartphones with the Peek Acuity app (Tumbling E chart). We found a good correlation between VA via Peek Acuity and clinical standard examination. However, our analysis results showed lower specificity for children aged 3 to 5 years than the other 2 age groups. Previous literature has reported that the test of resolution acuity (eg, Tumbling E) may overestimate VA and thus be less sensitive for ocular diseases than tests of recognition acuity (eg, Lea, HOTV, Snellen) [26]. The American Association for Pediatric Ophthalmology and Strabismus Vision Screening app (ie, Lea symbols chart) was evaluated by Nik Azis et al [21] via an iPad mini in Malaysia among children aged 5 to 6 years.

Accurate tests of VA can be performed by nonprofessional examiners using a mobile VA app. Professional training in vision screening may be one of the factors affecting the accuracy and reliability of VA testing. However, our meta-analysis reveals that the nonprofessional examiner group had higher accuracy (AUC 0.93) than the professional examiner group (AUC 0.87), especially for children. Given children's limited psychological and cognitive aptitude, parents or school teachers may better understand their children's responses, behavior, and mood than eye care professionals. Thus, parents or school teachers can potentially be eye screeners to test children's VA via mobile apps. For older adults, VA may be tested at patients' homes by an eye care worker with basic training or a field worker without formal training. For example, in Bastawrous et al [13], Peek Acuity was used at patients' homes by a community health care worker and achieved 84.6% sensitivity and 97.7% specificity in detecting eyes with severe VI.

### Limitations and Future Research

We recognize that this study has several limitations that provide opportunities for future research. First, the findings of this study should be interpreted with caution considering only a small number of studies were included in the meta-analysis. To increase the generalizability of our findings, more studies with longer follow-up periods are needed. Second, among the selected meta-analyses, there were no studies involving participants aged 22 to 54 years. Only one study involved participants aged 55 years and older [13], limiting the generalizability of the findings about that age group. Third, only one meta-analysis focused on VA tests using an iPad [21], which may not be representative. Thus, in subgroup analysis, we could not plot the summaries of sensitivity and specificity in Figure 5 and SROC curves in Figure 3. Fourth, some selected studies did not apply any strict standardized VA examination process and conditions. For instance, the brightness of a mobile device's screen could not be adjusted to be precisely the same as the light-box chart [9]. Some studies evaluated the sensitivity and specificity of VA of left and right eyes separately. For example, Nik Azis et al [21] demonstrated that right vision screening had a higher sensitivity and specificity than left vision screening.

Future research should explore how to determine abnormal changes in eye anatomy and functions. Multifunctional vision evaluation apps that integrate digital eye image recognition techniques may make smartphones and tablets more attractive for ophthalmic assessment. The modern artificial intelligence and mHealth techniques have great potential in improving timely VI detection and treatment.

### Conclusions

The findings of this study suggest that mobile VA test apps can play an important role in identifying VI by professional examiners as well as nonprofessionals, who can perform self-testing at a time and place convenient to them. Public awareness of the safety and benefits of VA tests should be promoted, and further research with a larger sample and longer follow-up are needed to evaluate the potential role of a mobile phone VA test app.

### Acknowledgments

This work was supported by grant CFH-2020-2-40911 from the Capital's Funds for Health Improvement and Research of China.

### Authors' Contributions

LGS designed the study and drafted the manuscript. XHK and XHC acquired data and undertook the statistical analysis and interpretation. XJQ and WWD drafted the manuscript. DZ and DFW undertook the statistical analysis. DSZ, YH, and CZ revised the manuscript and supervised the study.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategies.
[DOCX File , 15 KB - mhealth_v10i2e26275_app1.docx ]

Multimedia Appendix 2
Main study characteristics and findings from 8 studies that examined visual acuity by iPad apps.
[DOCX File , 24 KB - mhealth_v10i2e26275_app2.docx ]

Multimedia Appendix 3
Main study characteristics and findings from 8 studies that examined visual acuity by smartphone apps.
[DOCX File , 28 KB - mhealth_v10i2e26275_app3.docx ]

Multimedia Appendix 4
Multivariate meta-regression (sensitivity) for mobile device–based app in evaluating the visual acuity.
[DOCX File , 16 KB - mhealth_v10i2e26275_app4.docx ]

Multimedia Appendix 5
Multivariate meta-regression (specificity) for mobile device–based app in evaluating the visual acuity.
[DOCX File , 16 KB - mhealth_v10i2e26275_app5.docx ]

## References


1. Eckert KA, Carter MJ, Lansingh VC, Wilson DA, Furtado JM, Frick KD, et al. A simple method for estimating the economic cost of productivity loss due to blindness and moderate to severe visual impairment. Ophthalmic Epidemiol 2015 Sep 23;22(5):349-355. [doi: 10.3109/09286586.2015.1066394] [Medline: 26395661]
2. Burton M, Ramke J, Marques A, Bourne R, Congdon N, Jones I, et al. The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. Lancet Glob Health 2021 Apr;9(4):e489-e551 [FREE Full text] [doi: 10.1016/S2214-109X(20)30488-5] [Medline: 33607016]
3. Bourne R, Flaxman S, Braithwaite T, Cicinelli M, Das A, Jonas J, Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health 2017 Sep;5(9):e888-e897 [FREE Full text] [doi: 10.1016/S2214-109X(17)30293-0] [Medline: 28779882]
4. Solebo AL, Rahi J. Epidemiology, aetiology and management of visual impairment in children. Arch Dis Child 2014 Apr;99(4):375-379. [doi: 10.1136/archdischild-2012-303002] [Medline: 24148891]
5. O'Neill S, McAndrew D. The validity of visual acuity assessment using mobile technology devices in the primary care setting. Aust Fam Physician 2016 Apr;45(4):212-215 [FREE Full text] [Medline: 27052139]
6. Li B, Powell A, Hooper PL, Sheidow TG. Prospective evaluation of teleophthalmology in screening and recurrence monitoring of neovascular age-related macular degeneration. JAMA Ophthalmol 2015 Mar 01;133(3):276. [doi: 10.1001/jamaophthalmol.2014.5014] [Medline: 25473945]
7. Brady CJ, Eghrari AO, Labrique AB. Smartphone-based visual acuity measurement for screening and clinical assessment. JAMA 2015 Dec 22;314(24):2682. [doi: 10.1001/jama.2015.15855] [Medline: 26720028]
8. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of Peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR mHealth uHealth 2016 May 09;4(2):e30. [doi: 10.2196/mhealth.4746] [Medline: 27160779]
9. Jan-Bond C, Wee-Min T, Hong-Kee N, Zu-Quan I, Khairy-Shamel S, Zunaina E, et al. REST: an innovative rapid eye screening test. J Mobile Technol Med 2015 Oct 30;4(3):20-25. [doi: 10.7309/jmtm.4.3.4]
10. Black JM, Jacobs RJ, Phillips G, Chen L, Tan E, Tran A, et al. An assessment of the iPad as a testing platform for distance visual acuity in adults. BMJ Open 2013 Jun 20;3(6):e002730 [FREE Full text] [doi: 10.1136/bmjopen-2013-002730] [Medline: 23794568]
11. Shah N, Dakin SC, Whitaker HL, Anderson RS. Effect of scoring and termination rules on test-retest variability of a novel high-pass letter acuity chart. Invest Ophthalmol Vis Sci 2014 Mar 06;55(3):1386-1392. [doi: 10.1167/iovs.13-13340] [Medline: 24519424]
12. Perera C, Chakrabarti R, Islam FMA, Crowston J. The Eye Phone Study: reliability and accuracy of assessing Snellen visual acuity using smartphone technology. Eye (Lond) 2015 Jul 1;29(7):888-894 [FREE Full text] [doi: 10.1038/eye.2015.60] [Medline: 25931170]
13. Bastawrous A, Rono HK, Livingstone IAT, Weiss HA, Jordan S, Kuper H, et al. Development and validation of a smartphone-based visual acuity test (Peek Acuity) for clinical practice and community-based fieldwork. JAMA Ophthalmol 2015 Aug 01;133(8):930-937 [FREE Full text] [doi: 10.1001/jamaophthalmol.2015.1468] [Medline: 26022921]

14. Han X, Scheetz J, Keel S, Liao C, Liu C, Jiang Y, et al. Development and validation of a smartphone-based visual acuity test (Vision at Home). Transl Vis Sci Technol 2019 Jul 01;8(4):27 [FREE Full text] [doi: 10.1167/tvst.8.4.27] [Medline: 31440424]
15. Peters HB. Vision screening with a Snellen chart. Am J Optom Arch Am Acad Optom 1961 Sep;38(9):487-505. [doi: 10.1097/00006324-196109000-00001] [Medline: 13734772]
16. Shah N, Laidlaw D, Brown G, Robson C. Effect of letter separation on computerised visual acuity measurements: comparison with the gold standard Early Treatment Diabetic Retinopathy Study (ETDRS) chart. Ophthalmic Physiol Opt 2010 Mar;30(2):200-203. [doi: 10.1111/j.1475-1313.2009.00700.x] [Medline: 20444125]
17. Rono HK, Bastawrous A, Macleod D, Wanjala E, Di Tanna GL, Weiss HA, et al. Smartphone-based screening for visual impairment in Kenyan school children: a cluster randomised controlled trial. Lancet Global Health 2018 Aug;6(8):e924-e932. [doi: 10.1016/s2214-109x(18)30244-4]
18. Rhiu S, Lee HJ, Goo YS, Cho K, Kim J. Visual acuity testing using a random method visual acuity application. Telemed J E Health 2016 Mar;22(3):232-237. [doi: 10.1089/tmj.2015.0073] [Medline: 26280910]
19. Andersen T, Jeremiah M, Thamane K, Littman-Quinn R, Dikai Z, Kovarik C, et al. Implementing a school vision screening program in Botswana using smartphone technology. Telemed J E Health 2020 Feb 01;26(2):255-258. [doi: 10.1089/tmj.2018.0213] [Medline: 30794053]
20. de Venecia B, Bradfield Y, Trane R, Bareiro A, Scalamogna M. Validation of Peek Acuity application in pediatric screening programs in Paraguay. Int J Ophthalmol 2018 Aug 18;11(8):1384-1389 [FREE Full text] [doi: 10.18240/ijo.2018.08.21] [Medline: 30140645]
21. Nik Azis NN, Chew FLM, Rosland SF, Ramlee A, Che-Hamzah J. Parents' performance using the AAPOS Vision Screening App to test visual acuity in Malaysian preschoolers. J AAPOS 2019 Oct;23(5):268.e1-268.e6. [doi: 10.1016/j.jaapos.2019.01.019] [Medline: 31520719]
22. Zhao L, Stinnett SS, Prakalapakorn SG. Visual acuity assessment and vision screening using a novel smartphone application. J Pediatr 2019 Oct;213:203-210. [doi: 10.1016/j.jpeds.2019.06.021] [Medline: 31326117]
23. Black JM, Jacobs RJ, Phillips G, Chen L, Tan E, Tran A, et al. An assessment of the iPad as a testing platform for distance visual acuity in adults. BMJ Open 2013 Jun 20;3(6):e002730 [FREE Full text] [doi: 10.1136/bmjopen-2013-002730] [Medline: 23794568]
24. Lodha VS. Comparison of visual acuity measurement by smartphone based application vs. conventional Snellen visual acuity chart. Int J Enhanced Res Med Dental Care 2015;2(6):39-41.
25. Pathipati AS, Wood EH, Lam CK, Sáles C, Moshfeghi DM. Visual acuity measured with a smartphone app is more accurate than Snellen testing by emergency department providers. Graefes Arch Clin Exp Ophthalmol 2016 Jun 1;254(6):1175-1180. [doi: 10.1007/s00417-016-3291-4] [Medline: 26931323]
26. Rodríguez-Vallejo M, Llorens-Quintana C, Furlan WD, Monsoriu JA. Visual acuity and contrast sensitivity screening with a new iPad application. Displays 2016 Sep;44:15-20. [doi: 10.1016/j.displa.2016.06.001]
27. Tofigh S, Shortridge E, Elkeeb A, Godley BF. Effectiveness of a smartphone application for testing near visual acuity. Eye (Lond) 2015 Nov 24;29(11):1464-1458 [FREE Full text] [doi: 10.1038/eye.2015.138] [Medline: 26206531]
28. Zhang Z, Zhang S, Huang X, Liang L. A pilot trial of the iPad tablet computer as a portable device for visual acuity testing. J Telemed Telecare 2013 Feb 22;19(1):55-59. [doi: 10.1177/1357633x12474964]
29. Fernández J, Rodríguez-Vallejo M, Tauste A, Albarrán C, Basterra I, Piñero D. Fast measure of visual acuity and contrast sensitivity defocus curves with an iPad application. Open Ophthalmol J 2019 Mar 26;13(1):15-22. [doi: 10.2174/1874364101913010015]
30. Kergoat H, Law C, Chriqui E, Leclerc B, Kergoat M. Tool for screening visual acuity in older individuals with dementia. Am J Alzheimers Dis Other Demen 2017 Mar 24;32(2):96-100 [FREE Full text] [doi: 10.1177/1533317517689877] [Medline: 28116912]
31. Toy B, Myung D, He L, Pan C, Chang R, Polkinhorne A, et al. Smartphone-based dilated fundus photography and near visual acuity testing as inexpensive screening tools to detect referral warranted diabetic eye disease. Retina 2016 May;36(5):1000-1008. [doi: 10.1097/IAE.0000000000000955] [Medline: 26807627]
32. Phung L, Gregori N, Ortiz A, Shi W, Schiffman J. Reproducibility and comparison of visual acuity obtained with Sightbook mobile application to near card and Snellen chart. Retina 2016 May;36(5):1009-1020. [doi: 10.1097/IAE.0000000000000818] [Medline: 26509223]
33. Rono H, Bastawrous A, Macleod D, Bunywera C, Mamboleo R, Wanjala E, et al. Smartphone-guided algorithms for use by community volunteers to screen and refer people with eye problems in Trans Nzoia county, Kenya: development and validation study. JMIR Mhealth Uhealth 2020 Jun 19;8(6):e16345 [FREE Full text] [doi: 10.2196/16345] [Medline: 32558656]
34. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of Peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR Mhealth Uhealth 2016 May 09;4(2):e30 [FREE Full text] [doi: 10.2196/mhealth.4746] [Medline: 27160779]
35. Stewart-Brown S, Haslum M. Screening of vision in school: could we do better by doing less? BMJ 1988 Oct 29;297(6656):1111-1113 [FREE Full text] [doi: 10.1136/bmj.297.6656.1111] [Medline: 3143448]

## Abbreviations

**AUC:** area under the curve
**DOR:** diagnostic odds ratio
**ETDRS:** Early Treatment Diabetic Retinopathy Study
**logMAR:** Logarithm of the Minimum Angle of Resolution
**LR+:** likelihood ratio positive
**LR–:** likelihood ratio negative
**mHealth:** mobile health
**QUADAS-2:** Quality Assessment of Diagnostic Accuracy Studies–2
**SROC:** summary receiver operating characteristic
**VA:** visual acuity
**VI:** vision impairment

*Edited by L Buis; submitted 04.12.20; peer-reviewed by C Ochoa-Zezzatti, KL Mauco; comments to author 23.02.21; revised version received 20.06.21; accepted 01.10.21; published 14.02.22.*

*Please cite as:*
*Suo L, Ke X, Zhang D, Qin X, Chen X, Hong Y, Dai W, Wu D, Zhang C, Zhang D*
*Use of Mobile Apps for Visual Acuity Assessment: Systematic Review and Meta-analysis*
*JMIR Mhealth Uhealth 2022;10(2):e26275*
*URL: https://mhealth.jmir.org/2022/2/e26275*
*doi:10.2196/26275*
*PMID:35156935*

Review

# Measurement Properties of Smartphone Approaches to Assess Diet, Alcohol Use, and Tobacco Use: Systematic Review

Louise Thornton[1,2,3], BA, PhD; Bridie Osman[1], BSc; Katrina Champion[1], BA, PhD; Olivia Green[1], BA; Annie B Wescott[4], BA, PhD; Lauren A Gardner[1], BA, PhD; Courtney Stewart[5], BA; Rachel Visontay[1], BA; Jesse Whife[5], BA; Belinda Parmenter[6], BA, PhD; Louise Birrell[1], BA, PhD; Zachary Bryant[1], BA; Cath Chapman[1], BA, PhD; David Lubans[7], BA, PhD; Tim Slade[1], BA, PhD; John Torous[8], BA, PhD; Maree Teesson[1], BA, PhD; Pepijn Van de Ven[9], BA, PhD

[1]The Matilda Centre for Research in Mental Health and Substance Use, The University of Sydney, Sydney, Australia

[2]School of Medicine and Public Health, The University of Newcastle, Newcastle, Australia

[3]School of Public Health and Community Medicine, University of New South Wales, Kensington, Australia

[4]Galter Health Sciences Library & Learning Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, United States

[5]National Drug Research Institute, Curtin University, Perth, Australia

[6]School of Health Sciences, The University of New South Wales, Sydney, Australia

[7]Priority Research Centre for Physical Activity and Nutrition, The University of Newcastle, Newcastle, Australia

[8]Beth Israel Deaconness Medical Centre, Harvard Medical School, Boston, MA, United States

[9]Health Research Institute, University of Limerick, Limerick, Ireland

**Corresponding Author:**
Louise Thornton, BA, PhD
The Matilda Centre for Research in Mental Health and Substance Use
The University of Sydney
Level 6, Jane Foss Russel Building
Camperdown
Sydney, 2006
Australia
Phone: 61 0403744089
Email: louise.thornton@sydney.edu.au

## Abstract

**Background:** Poor diet, alcohol use, and tobacco smoking have been identified as strong determinants of chronic diseases, such as cardiovascular disease, diabetes, and cancer. Smartphones have the potential to provide a real-time, pervasive, unobtrusive, and cost-effective way to measure these health behaviors and deliver instant feedback to users. Despite this, the validity of using smartphones to measure these behaviors is largely unknown.

**Objective:** The aim of our review is to identify existing smartphone-based approaches to measure these health behaviors and critically appraise the quality of their measurement properties.

**Methods:** We conducted a systematic search of the Ovid MEDLINE, Embase (Elsevier), Cochrane Library (Wiley), PsycINFO (EBSCOhost), CINAHL (EBSCOHost), Web of Science (Clarivate), SPORTDiscus (EBSCOhost), and IEEE Xplore Digital Library databases in March 2020. Articles that were written in English; reported measuring diet, alcohol use, or tobacco use via a smartphone; and reported on at least one measurement property (eg, validity, reliability, and responsiveness) were eligible. The methodological quality of the included studies was assessed using the Consensus-Based Standards for the Selection of Health Measurement Instruments Risk of Bias checklist. Outcomes were summarized in a narrative synthesis. This systematic review was registered with PROSPERO, identifier CRD42019122242.

**Results:** Of 12,261 records, 72 studies describing the measurement properties of smartphone-based approaches to measure diet (48/72, 67%), alcohol use (16/72, 22%), and tobacco use (8/72, 11%) were identified and included in this review. Across the health behaviors, 18 different measurement techniques were used in smartphones. The measurement properties most commonly examined were construct validity, measurement error, and criterion validity. The results varied by behavior and measurement approach, and the methodological quality of the studies varied widely. Most studies investigating the measurement of diet and

alcohol received *very good* or *adequate* methodological quality ratings, that is, 73% (35/48) and 69% (11/16), respectively, whereas only 13% (1/8) investigating the measurement of tobacco use received a *very good* or *adequate* rating.

**Conclusions:** This review is the first to provide evidence regarding the different types of smartphone-based approaches currently used to measure key behavioral risk factors for chronic diseases (diet, alcohol use, and tobacco use) and the quality of their measurement properties. A total of 19 measurement techniques were identified, most of which assessed dietary behaviors (48/72, 67%). Some evidence exists to support the reliability and validity of using smartphones to assess these behaviors; however, the results varied by behavior and measurement approach. The methodological quality of the included studies also varied. Overall, more high-quality studies validating smartphone-based approaches against criterion measures are needed. Further research investigating the use of smartphones to assess alcohol and tobacco use and objective measurement approaches is also needed.

**International Registered Report Identifier (IRRID):** RR2-10.1186/s13643-020-01375-w

***(JMIR Mhealth Uhealth 2022;10(2):e27337)*** doi:10.2196/27337

**KEYWORDS**

smartphone; app; alcohol; smoking; diet; measurement; mobile phone

# Introduction

## Background

Traditional measurement techniques to assess health behaviors can be difficult and burdensome for individuals, clinicians, and researchers alike and are often subject to problems such as recall bias and forgotten information [1]. Novel measurement techniques are needed to increase compliance and accuracy with recording data, reduce respondent burden, and increase the quality and detail of health behavior information. Smartphones may present an opportunity to do just this.

Smartphones have become an integral part of the lives of many people [2], and users often use their smartphones and smartphone apps to record and measure a range of health behaviors [3]. In addition, the standard features of smartphones (ie, sensors, such as accelerometers, gyroscopes, and light sensors) allow these devices to continuously monitor contexts of users (eg, activity, location, and environment). Data from these sensors can be collected *passively*, without the active involvement of the user, and generate information about some behaviors with little burden [4]. Unfortunately, the ability to accurately measure key health behaviors using smartphones is currently hampered by a lack of understanding of the validity and reliability of the approaches used.

Consumption behaviors, such as dietary intake, alcohol use, and tobacco smoking, are typically measured using approaches prone to bias. For instance, diet is often assessed using food diaries that require participants to record everything they eat and drink for a period. This approach requires participants to be literate and highly motivated, and research has shown that the quality of food records declines considerably over time [5]. Retrospective recall methods are also commonly used for these behaviors. These methods often require multiple administrations to accurately capture variations in behavior over time [5,6], rely heavily on the memory of participants and interviewer training, and may be affected by social desirability bias, particularly for smoking and alcohol use. In addition, the accuracy of these self-report approaches is dependent on the ability of participants to accurately estimate portion sizes (or standard drinks) and, as such, often suffer from underreporting of behaviors [5,7,8].

Furthermore, traditional methods to objectively measure consumption behaviors are often burdensome and costly to administer. Weighed food records, for example, where food to be consumed and any waste left over are weighed and recorded, have been shown to be a valid method of recording dietary intake. However, outside of a laboratory setting, this approach is extremely burdensome and impractical [5]. In addition, although the *gold standard* doubly labeled water method (where isotopes in water provided to participants are used for tracing purposes) can accurately estimate the energy intake of participants, the approach requires multiple urine, saliva, or blood samples to be taken; is costly; requires sophisticated equipment; and is valid only among weight-stable participants. Therefore, it is only feasible within specialized research laboratories and not for use in clinical settings or by consumers themselves [5]. Although devices to objectively measure alcohol and tobacco use via expired breath ethanol and expired carbon monoxide (CO) are readily available for purchase, they must be regularly and properly calibrated to produce accurate results. Furthermore, as these behaviors often occur outside of the home and in social situations, their use may not be practical or acceptable in free-living conditions.

Given the ubiquitous and portable nature of smartphones, their powerful computing abilities, built-in cameras and sensors, and the social acceptance of their use in almost all situations, accurate smartphone measurement could offer solutions to many of the issues associated with traditional approaches to measure diet, alcohol, and tobacco use. Although several reviews of both published literature and mobile apps available in the marketplace have examined the efficacy of apps to help improve diet, alcohol use, and tobacco use, only 1 review to date has specifically focused on the measurement properties of smartphone-based approaches to measure any of these behaviors [9]. As such, there is a limited understanding of how these 3 behaviors might be validly and reliably measured using smartphones [10-13].

## Objectives

This study aims to systematically review the existing literature on the measurement properties of smartphone-based approaches to assess diet, alcohol use, and tobacco use. The specific objectives of this review are as follows:

1. To identify and describe the ways in which diet, alcohol use, and tobacco use have been measured using smartphones
2. To describe and critically evaluate the available evidence on the measurement properties of these approaches
3. To provide recommendations on the most suitable and effective ways of measuring diet, alcohol use, and tobacco use with smartphones

# Methods

## Overview

This review was conducted in accordance with the published review protocol [14] and the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines [15]. It is part of a larger systematic review that examines the measurement properties of smartphone approaches to assess 6 key health behaviors (physical activity, sedentary behavior, sleep, diet, alcohol use, and tobacco use). Owing to the large number of eligible studies identified in this larger review, only those studies that examined consumption behaviors (ie, diet, alcohol use, and tobacco use) were included in the current review to allow for adequate description and discussion of the approaches identified and their associated measurement properties.

## Search Strategy and Selection Criteria

A research librarian (ABW) searched Ovid MEDLINE, Embase (Elsevier), Cochrane Library (Wiley), PsycINFO (EBSCOhost), CINAHL (EBSCOHost), Web of Science (Clarivate), SPORTDiscus (EBSCOhost), and IEEE Xplore Digital Library for research describing the measurement properties of smartphone-based approaches to assess at least one of the 6 key health behaviors. All databases were searched on March 1, 2020. A date limit was applied from 2007 to present, as 2007 is the year in which the first *smartphones* (ie, mobile phones with large capacitive touchscreens using direct finger input, as opposed to a stylus or keypad) were released. An example search strategy developed for MEDLINE is shown in Multimedia Appendix 1. Published studies with any type of study design, involving participants of all ages, were eligible for inclusion. Included articles were required to be in English language, peer-reviewed studies of human participants, describe a smartphone-based approach to assess diet, alcohol use, and tobacco use and to report on at least one measurement property of this approach identified in the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) Taxonomy of Measurement Properties (Table 1).

**Table 1.** Consensus-Based Standards for the Selection of Health Measurement Instruments taxonomy of measurement properties[a].

| Domain | Domain description | Measurement properties |
|---|---|---|
| Reliability | Degree to which the measurement is free from measurement error | • Internal consistency<br>• Reliability<br>• Measurement error |
| Validity | Degree to which an outcome measure measures the construct it purports to measure | • Content validity (including face validity)<br>• Construct validity (including structural validity, hypotheses testing, and cross-cultural validity)<br>• Criterion validity |
| Responsiveness | Ability of an outcome measure to detect change over time | • Responsiveness |

[a]See Consensus-Based Standards for the Selection of Health Measurement Instruments definitions of domains, measurement properties, and aspects of measurement properties [16] for full descriptions and definitions of measurement properties.

Studies were excluded if they described the feasibility of the measurement approach only, described the measurement properties of using text messaging only to measure behaviors, or described the measurement properties of a wearable device (eg, Fitbit [Fitbit Inc]) alone.

## Data Extraction and Screening

All identified studies were exported into Endnote (version 8) to remove duplicates. Records were then uploaded to the Covidence Systematic Review software (Veritas Health Innovation) for screening. Authors participating in the screening, full-text review, and data extraction process participated in training sessions where multiple reviewers independently reviewed and discussed a selection of papers to ensure consistency across reviewers. Titles and abstracts were first screened by 1 reviewer (RV, JW, CS, LT, BO, LB, LG, OG, BP, or JT). Records were excluded if it was clear from the title and abstract that they did not examine the measurement properties of a smartphone-based approach to measure diet, tobacco, or alcohol. A total of 8 members of the research team (OG, CS, JW, LT, BO, ZB, KC, and RV) then participated in full-text screening of results, with the full text of potentially relevant studies independently assessed for eligibility by 2 members of this group, and any disagreements were resolved with the assistance of a third researcher. LT, BO, CS, or OG extracted data using a standardized form. Further details of the data extraction are included in the published protocol [14].

## Data Analysis

The primary outcomes of interest were the measurement properties of smartphone-based approaches to assess diet, alcohol use, and tobacco use. Specifically, we investigated, as reported, the internal consistency, reliability, measurement error, content validity, construct validity (including convergent validity), criterion validity, and responsiveness of the approaches identified. As there is currently no agreed-upon gold standard method for self-reported measurement of diet, alcohol use, or tobacco use, only studies in which the smartphone-based

approach was compared with an objective measure of the behavior (eg, weighed food records and observed number of drinks or cigarettes consumed) were classified as investigating criterion validity. The smartphone-based approach was compared with a self-report measure, even if it was described as a gold standard method by the authors, the paper was classified as an investigating construct, specifically convergent validity.

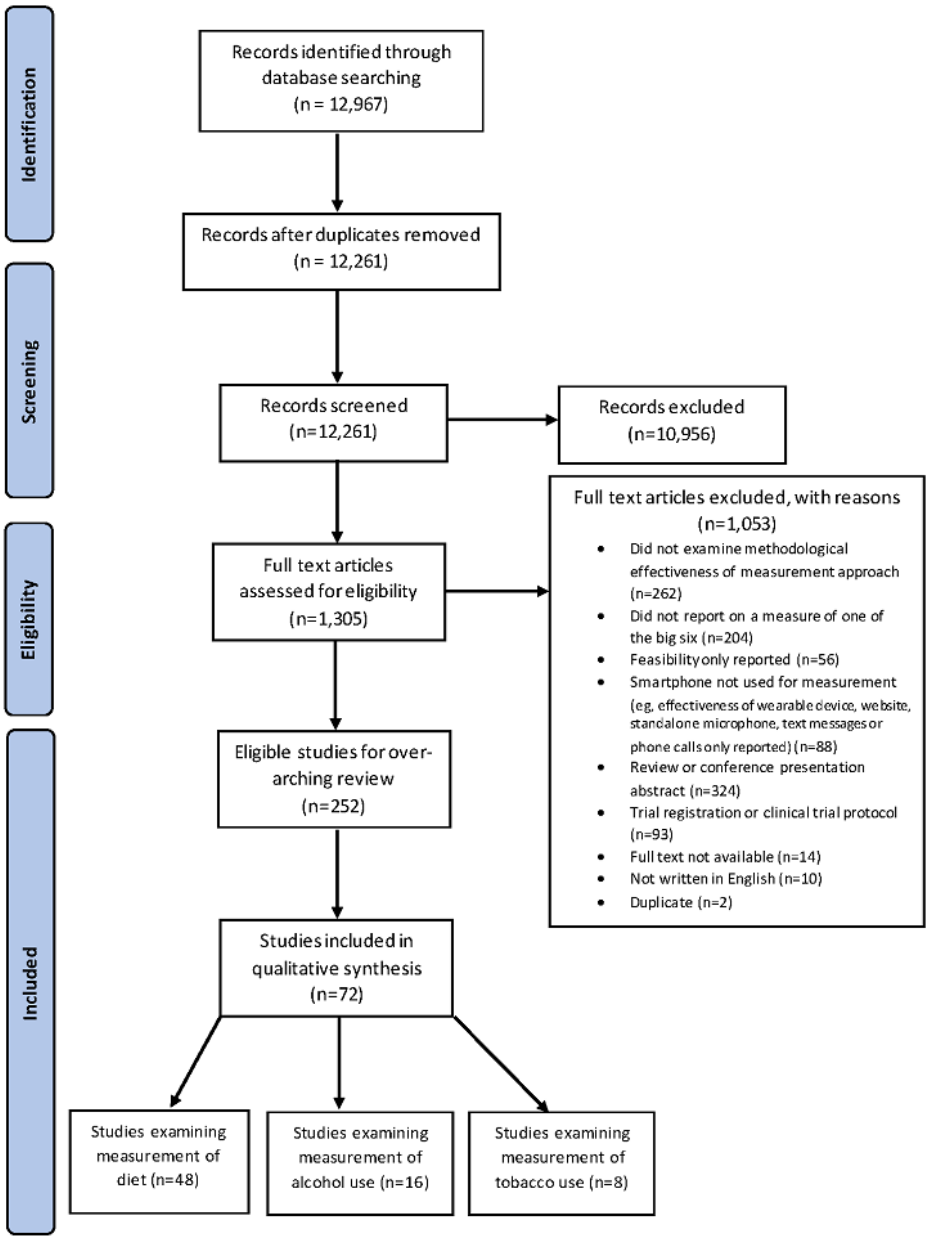
A narrative synthesis of the included studies was undertaken for diet, alcohol use, and tobacco use separately, grouped according to the type of measurement approach used, which included *self-report approaches*, where participants were asked to actively enter self-report information about their behaviors; *active objective approaches*, where participants were asked to actively provide an objective measure of their behavior (eg, taking a photo of their food); and *passive objective approaches*, where data generated by smartphone sensors were collected without the active involvement of the participant and used to generate information about behaviors. The methodological quality of the included studies was assessed using the COSMIN Risk of Bias checklist [17]. The COSMIN Risk of Bias checklist was designed to assess the methodological quality of studies investigating the measurement properties of patient-reported outcome measures. It specifies several standards for design requirements and preferred statistical methods when assessing different measurement properties. The methodological quality of each study was evaluated by rating all standards for each measurement property investigated on a 4-point Likert scale. A standard can be rated as *very good* (there is evidence that the standard is met or when a preferred method was optimally used), *adequate* (it can be assumed that the standard is met or when the preferred method was used, but it was not optimally applied) *doubtful* (it is unclear whether the standard is met or unclear if a preferred method was used), or *inadequate* (there is evidence that the standard is not met or when the preferred method was not used). The overall quality of a study is determined by taking the lowest rating of any standard [17].

## Results

### Overview

Of 12,967 identified records, 1305 (10.06%) were independently fully reviewed by 2 reviewers. Agreement between reviewers was 83.22% (1086/1305). A total of 72 studies were ultimately included in the current review. These 72 studies involved 4732 participants and were most commonly conducted in the United States (27/72, 38%), European countries (9/72, 13%), the United Kingdom (6/72, 8%), and Australia (13/72, 18%). As shown in Figure 1, up to 67% (48/72) papers examined the measurement of diet, 22% (16/72) examined alcohol use measurement, and 11% (8/72) examined measurement of tobacco use. The details of the identified smartphone-based measurement approaches are discussed below.

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




## Diet

### *Overview*

Overall, 67% (48/72) of the papers examined the measurement properties of a smartphone-based approach to assess diet (n range 0-203; 63.77% of participants in included studies were female; age range of participants 3-75 years). The key characteristics of these studies are detailed in Table 2 (for full study details, see Multimedia Appendix 2 [3,18-85]). Of the studies, 58% (28/48) described self-report approaches, whereas 42% (20/48) investigated active objective approaches. No studies have identified that used passive objective approaches to measure diet. The most commonly assessed measurement properties were construct validity, measurement error, and criterion validity.

**Table 2.** Key characteristics of studies examining the measurement of diet via a smartphone.

| Study | Country | App name | Publicly available | Measurement approach | Measurement properties assessed | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Reliability | Measurement error | Construct validity | Criterion validity | Responsiveness | Risk of bias |
| Ahmed et al [40] | Canada | MyFitnessPal | Yes | Self-report | __[a] | ✓ | — | ✓ | — | Very good |
| Ali et al [27] | United Kingdom | NR[b] | NR | Self-report | — | ✓ | ✓ | — | — | Doubtful |
| Ambrosini et al [26] | Australia | Easy diet diary | Yes | Self-report | — | ✓ | ✓ | — | — | Very good |
| Ashman et al [57] | Australia | DietBytes | NR | Active objective | — | ✓ | ✓ | — | — | Very good |
| Béjar [25] | Spain | e-EPIDEMIOLOGY | NR | Self-report | — | — | ✓ | — | — | Very good |
| Béjar [25] | Spain | e-12HR[c] | NR | Self-report | — | — | ✓ | — | — | Very good |
| Béjar et al [39] | Spain | e-12HR | NR | Self-report | — | — | ✓ | — | — | Very good |
| Béjar et al [36] | Spain | e-12HR | NR | Self-report | — | — | ✓ | — | — | Very good |
| Boushey et al [55] | United States | N/A[d] | N/A | Active objective | — | ✓ | — | ✓ | — | Adequate |
| Bruening et al [42] | United States | devilSPARC[e] | NR | Self-report | — | — | ✓ | — | — | Inadequate |
| Bucher et al [23] | Switzerland | e-CA[f] | NR | Self-report | — | ✓ | ✓ | ✓ | — | Very good |
| Carter et al [24] | United Kingdom | My Meal Mate | NR | Self-report | — | ✓ | ✓ | — | — | Inadequate |
| Chen et al [34] | Australia | MyFitnessPal | Yes | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Chmurzynska et al [37] | Poland | NR | NR | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Costello et al [56] | United Kingdom | N/A | N/A | Active objective | — | ✓ | ✓ | ✓ | — | Adequate |
| Delisle Nyström et al [54] | Sweden | N/A | N/A | Active objective | — | ✓ | ✓ | ✓ | — | Adequate |
| Fallaize et al [32] | United Kingdom | Samsung Health; MyFitnessPal; FatSecret; Noom Coach; Lose it! | Yes | Self-report | — | ✓ | — | ✓ | — | Adequate |
| Griffiths et al [22] | United States | MyFitnessPal; Fitbit; Lose it!; MyPlate; Lifesum | Yes | Self-report | — | — | ✓ | — | — | Very good |
| Hezarjaribi et al [33] | United States | EZNutriPal | Yes | Self-report | — | — | ✓ | ✓ | — | Very good |
| Hezarjaribi et al [38] | United States | Speech2Health | No | Self-report | — | — | ✓ | — | — | Inadequate |
| Huang et al [53] | Australia | NR | NR | Active objective | — | — | — | ✓ | — | Very good |

| Study | Country | App name | Publicly available | Measurement approach | Measurement properties assessed | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Reliability | Measurement error | Construct validity | Criterion validity | Responsiveness | Risk of bias |
| Hutchesson et al [45] | Australia | N/A | Yes | Self-report | — | ✓ | ✓ | — | — | Doubtful |
| Kato et al [62] | Japan | DialBetics | NR | Active objective | — | ✓ | — | ✓ | — | Very good |
| Kong et al [52] | China | N/A | N/A | Active objective | — | ✓ | — | ✓ | — | Doubtful |
| Lancaster et al [28] | Australia | Research Food Diary | Yes | Self-report | — | ✓ | — | — | — | Inadequate |
| Lemacks et al [31] | United States | Bridge2U | No | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Liu et al [49] | Taiwan | NR | No | Active objective | — | — | ✓ | — | — | Very good |
| Liu et al [30] | Taiwan | NR | NR | Self-report | — | ✓ | — | ✓ | — | Adequate |
| Martin et al [50] | United States | NR | No | Active objective | ✓ | — | — | ✓ | — | Very good |
| Martin et al [51] | United States | NR | No | Active objective | ✓ | ✓ | — | ✓ | — | Inadequate |
| Most et al [63] | United States | SmartIntake | Yes | Self-report | — | — | — | ✓ | — | Very good |
| Nicklas et al [48] | United States | NR | N/A | Active objective | — | ✓ | — | ✓ | — | Adequate |
| Pendergast et al [21] | Australia | FoodNow | NR | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Prinz et al [60] | Germany | N/A | N/A | Active objective | — | ✓ | — | ✓ | — | Very good |
| Rangan et al [20] | Australia | e-DIA[g] | NR | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Rangan et al [19] | Australia | e-DIA | NR | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Rhyner et al [47] | Switzerland | GoCARB | NR | Active objective | — | ✓ | — | — | — | Doubtful |
| Rodder et al [35] | United States | MyNetDiary | Yes | Self-report | — | — | ✓ | — | — | Very good |
| Rollo et al [59] | Australia | Nutricam dietary assessment method | NR | Self-report | — | ✓ | ✓ | — | — | Very good |
| Rollo et al [61] | Australia | Nutricam dietary assessment method | NR | Active objective | — | — | ✓ | ✓ | — | Very good |
| Schiel et al [64] | Germany | DiaTrace | Yes | Active objective | — | — | ✓ | — | — | Inadequate |
| Schiel et al [65] | Germany | DiaTrace | Yes | Active objective | — | — | ✓ | — | — | Inadequate |
| Smith et al [44] | China | SA-24R[h] | NR | Self-report | — | ✓ | ✓ | — | — | Adequate |
| Swendeman et al [43] | United States | Ohmage | No | Self-report | — | — | ✓ | ✓ | — | Very good |

| Study | Country | App name | Publicly available | Measurement approach | Measurement properties assessed | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Reliability | Measurement error | Construct validity | Criterion validity | Responsiveness | Risk of bias |
| Teixeira et al [18] | Brazil | MyFitnessPal | NR | Self-report | — | ✓ | — | — | — | Doubtful |
| Wellard-Cole et al [29] | Australia | Eat and Track app | NR | Self-report | — | ✓ | ✓ | — | — | Very good |
| Zhang et al [46] | United States | Snap-n-Eat | No | Active objective | — | — | — | ✓ | — | Very good |
| Zhu et al [58] | United States | NR | NR | Self-report | — | — | — | — | ✓ | Very good |

[a]Measurement property was either not assessed or not reported.

[b]NR: not reported.

[c]e-12HR: electronic 12-hour dietary recall.

[d]N/A: not applicable.

[e]devilSPARC: Social impact of Physical Activity and Nutrition in College.

[f]e-CA: electronic carnet alimentaire (“food record” in French).

[g]e-DIA: Electronic Dietary Intake Assessment.

[h]SA-24R: Smartphone Assisted 24 Hour Recall.

## *Self-report*

### Overview

Of the 28 studies that examined self-report methods for recording diet, 24 (86%) investigated food diary apps, 2 (7%) used ecological momentary assessment (EMA), 1 (4%) examined a smartphone-assisted 24-hour dietary recall tool, and 1 (4%) investigated the use of a web-based food database via a smartphone.

### Food Diary Apps

A total of 24 studies investigated food diary apps [18-40,86] designed to facilitate daily or real-time recording of dietary intake. Usually, these are linked to a large database containing preprogrammed information about the energy and nutrient content of popular foods. These apps allow users to select food and beverages they have consumed, and their energy and nutrient intake for the day is automatically calculated. A wide range of food diary apps were examined within the included studies, 12 of which (described across 9 studies) [18,21,22,26,32,34,35,40,87] were publicly available on the leading app stores (Google Play or iOS).

A total of 3 studies [18,34,40] exclusively examined the measurement properties of MyFitnessPal, a widely used commercially available app, and 2 studies examined MyFitnessPal along with another app [22,32]. Furthermore, 80% (4/5) of these studies found evidence to support the validity of MyFitnessPal. Teixeira et al [18] compared the energy intake generated by MyFitnessPal with estimates generated by a paper-based food record. Griffiths et al [22] compared the app with estimates generated by a dietary analysis program [87], and Ahmed et al [40] and Fallaize et al [32] compared the app with weighed food records. They found correlations between the energy intake estimated by MyFitnessPal and their comparison measure of 0.70-0.99. Falliaze et al [32], Griffiths et al [22], and Ahmed et al [40] found no significant differences among the estimation of energy and most nutrients; however, where differences did exist, MyFitnessPal was found to yield lower intakes. Chen et al [34], by contrast, found poor agreement between MyFitnessPal and energy intake estimated via a 24-hour recall measure, finding weak to moderate correlations (0.21-0.42) and significantly lower values for total energy and all macronutrients recorded via MyFitnessPal. They found no proportional bias for energy or any of the nutrients assessed; however, wide limits of agreement were observed.

Furthermore, 2 studies investigated top nutrition tracking apps, including Fitbit, Lose it!, MyPlate, Lifesum, Samsung Health, Fatsecret, Noom Coach, and MyFitnessPal [22,32]. Both studies found strong correlations among energy and nutrient estimations via the apps and their comparison measures (0.73-0.96 and 0.79-0.91, respectively). However, numerous significant differences among nutrient estimations generated by the apps and comparison measures were identified, particularly within the *Lose it!* app. Other publicly available nutrition apps were investigated in 2 studies, with moderate mean correlations between apps and their comparison measures found (mean 0.61, SD 0.11 [26] and mean 0.67, SD 0.14 [35]).

A total of 3 studies [21,33,38] investigated the use of unstructured data entry methods to self-report food intake, compared with structured forms of recording food intake information. The unstructured data entry methods examined information about food intake recorded via free-form speech and text descriptions. Food intake information was then extracted using manual coding or natural language processing (NLP) software. Pendergast et al [21] investigated the FoodNow app, which allowed diet information to be recorded via text descriptions, voice messages, and optional images. This

unstructured data were coded by trained nutritionists to match each food or beverage item described in the app to an appropriate item in a food and nutrient database [88]. They compared this approach to energy expenditure measured via the SenseWear Armband. Bland–Altman plots showed wide limits of agreement, indicating error at the individual level but no evidence of systematic bias among methods. The correlation among methods was strong (0.75), and an acceptable level of reliability among methods was found (intraclass correlation coefficient 0.75, 95% CI 0.61-0.84). Hezarjaribi et al [33,38] examined *EZNutriPal* and *Speech2Health*, 2 interactive diet monitoring systems that facilitate the collection of speech recordings and free-text data regarding dietary intake, real-time prompting, and personalized nutrition monitoring. In contrast to Pendergast et al [21] and the FoodNow app, the EZNutriPal and Speech2Health apps feature an NLP unit that allows automatic identification of food items described in the unstructured data provided. In the Speech2Health system, Hezarjaribi et al [38] used standard NLP techniques in combination with a bespoke pattern mapping technique to extract food names and portion sizes from spoken text. These data were then used to estimate the nutrient information. In EZNutriPal, Hezarjaribi et al [33] used an NLP framework based on named-entity recognition, where unrecognized entities were added to a training set to continuously update the ability of the NLP framework to correctly identify food items from an individual's speech. Individual recognized entities relating to food items, units, and quantities were then further processed to obtain an estimate of nutrient information. This methodology was tested using 13 participants across a 13-day period using EZNutriPal. The authors found that compared with labeling of the unstructured data by patients, EZNutriPal achieved an accuracy of 89.7% in calorie intake estimation [33], whereas Speech2Health achieved an accuracy of 92.2%. In their 2019 study, Hezarjabi et al [33,38] also compared the performance of these 2 apps and found that the Speech2Health app identified 3.4 times more than the actual number of food items contained in test sentences, whereas EZNutriPal identified 0.8 times less than the actual number of food items contained in test sentences. An interesting aspect of the 2019 study of Hezarjabi et al [33] was that it explicitly incorporated personalization of the food recognition system (from voice) by allowing users to provide labels for unrecognized voice inputs. These inputs were then used to further train the algorithm and thus improve the future performance of the app.

#### EMA Apps

EMA aims to maximize the ecological validity of data collected by repeatedly collecting information about the current behaviors of participants in real-time in their natural environment [41]. Overall, 2 studies investigated apps using EMA where participants were prompted multiple times throughout the day to record their food intake [42,43]. Bruening et al [42] compared smartphone-based EMAs with 24-hour dietary recalls, whereas Swendeman et al [43] examined the agreement among EMAs of self-reported diet quality and brief dietary recall measures, anthropometric measurements, and bloodspot biomarkers. Bruening et al [42] found good agreement between their methods, with 87% of food reported in both systems. Similarly, Swendeman et al [43] found that self-reported diet quality assessed via EMAs was moderately correlated with dietary recall measures for foods with high sugar content and fast food but weakly correlated with fruits and vegetables, anthropometric, and biomarker measures.

#### 24-Hour Dietary Recall

One study investigated the performance of a smartphone-assisted 24-hour dietary recall tool in measuring beverage intake among young Chinese adults [44], comparing it with a paper-based tool and 24-hour urine samples. Participants reported significantly reduced beverage intake via the smartphone-assisted 24-hour recall compared with that via the paper-based recall and fluid intake as assessed by the smartphone, and urine volume was moderately correlated (0.58). In addition, they found evidence of systematic measurement errors whereby the bias for smartphone and paper-based recall methods were not consistent across levels of intake, with the bias increasing with higher intake of beverages.

#### Web-Based Food Database

One study [45] evaluated the accuracy of 7-day food record methods accessed on the web via a smartphone, via a computer, and using pen and paper. They found no significant differences among total energy expenditure and energy intake reported for the 3 different methods; however, their examination of the measurement error of these approaches suggested that there may be greater underreporting of energy intake using paper-based diaries compared with computer- and smartphone-based methods.

### *Active Objective*

#### Overview

A total of 20 studies [46-65] examined apps that actively and objectively measured dietary intake. All studies used images of food to be consumed (and often also food waste) captured by the camera of a smartphone. Overall, 75% (15/20) of studies [48,50-52,54-57,59-65] investigated *manually analyzed food photography* methods where participants took photos of their food, which were then sent to researchers for analysis. Furthermore, 25% (5/20) of studies [46,47,49,53,58] used *automatically analyzed food photography* methods where images of food were captured by participants using specialized apps, which then analyzed images and calculated the energy and nutrient content of foods pictured automatically.

#### Manually Analyzed Food Photography

A total of 15 studies [48,50-52,54-57,59-65] used this method, of which 87% (13/15) demonstrated some evidence of its reliability and validity. Rollo et al [59,61], for example, conducted 2 studies to examine the performance of their Nutricam Dietary Assessment Method (NuDAM). NuDAM is an app that allows users to capture a photograph of food items before consumption and store a voice recording to explain the image contents before it is sent to a website for analysis by a dietitian. In their 2011 study [59], energy intake measured by the app was compared with a written food diary. Individual differences in energy intake between the 2 records varied from 6.7% to 29.7%, and on average, energy intake was

underrecorded using the app. In their 2015 study [61], energy intake assessed via NuDAM was compared with weighed food records and energy expenditure using the doubly labeled water method. Moderate to strong correlations between NuDAM and weighed food records were found for energy and nutrient intakes (0.57-0.85), and mean nutrient intakes were not significantly different. The overall mean energy intake calculated by the app and weighed food records were both significantly lower than the total energy expenditure calculated using the doubly labeled water method. Participants who were found to underreport using the app were also underreported via weighed food records.

Another 6 studies [50-52,56,60,62] compared manually analyzed food photography to weighed food records and found strong correlations among methods for energy (0.92-0.99) [56,60], carbohydrates (0.93-0.99), fat (0.84-0.99), and protein (0.94-0.99) estimates [52,60]. A total of 2 studies by Martin et al [50,51] also examined the reliability of food photography methods over time and found that the energy intake estimated using this method was reliable over 3 [50] and 6 days of testing [51]. Although there was good agreement among the methods for daily energy and macronutrient intakes in Kong et al [52] and Kato et al [62], in the study by Kong et al [52], as intake increased, underestimation by the app was identified, whereas Kato et al [62] found that images captured via the app generated higher values than the weighed food record for some macronutrients. Costello et al [56] also found evidence of a small standardized bias.

In addition, 2 studies [57,63] conducted among pregnant women generated limited evidence for the validity of food photography among this population. For example, Ashman et al [57] found moderate to strong correlations among food photography and 24-hour recall for energy and macronutrients (0.58-0.84) among this population. Three studies among children and adolescents found no significant differences among energy intake estimated via food images and self-reported energy intake [64,65] or energy intake estimated via the doubly labeled water method [54]. However, in their study of 3- to 5-year-old children, Niklas et al [48] found the remote food photography method to significantly underestimate the mean daily energy intake when using the doubly labeled water method.

Similarly, Boushey et al [55] found only moderate correlations (0.58) among dietary intake estimates using the doubly labeled water method and manually analyzed food photography. There was no evidence of a systematic bias. Energy intake calculated via their app was found to be significantly less than the estimates calculated via the doubly labeled water method, with differences more pronounced in men than in women.

#### Automatically Analyzed Food Photography

A total of 5 studies [46,47,49,53,58] used this method, all of which provided some evidence of its reliability or validity. In the study by Zhu et al [58], for example, images of meals captured using a smartphone camera were segmented and identified, and their volume was estimated. *Before* and *after* images were used to estimate food intake and determine energy and nutrients consumed. The app accurately identified between 84% and 96% of 19 different food items. The study also explored the estimation of volume using 7 food items and the estimation of weight using 2 food items. The mean percentage error of the volume estimates was 5.65%. To estimate the mass, the system had a percentage error between 3% and 56%.

Liu et al [49] examined two new methods to assist with the automatic analysis of food photography—an interactive photo interface (IPI) and a sketching-based interface (SBI). The IPI presented users with images of predetermined portion sizes of a specific food and allowed users to scan and select the most representative image matching the food that they were measuring. The SBI required users to relate the food shape to a readily available comparator (eg, credit card) and scribble to shade in the appropriate area. These were compared with traditional life-sized photos commonly used by dietitians to help people identify portion sizes. The overall accuracies of the IPI, SBI, and traditional life size photo method were 66.98%, 46.05%, and 72.06%, respectively, showing that the SBI method was significantly less accurate than the IPI and traditional life size photo methods. In another study [47] investigating the GoCARB app, participants were required to place a reference card next to their plate and take 2 images using a smartphone. A series of computer vision modules detected the plate and automatically segmented and recognized different food items into 9 broad food classes (pasta, potatoes, meat, breaded items, rice, green salad or vegetables, mashed potatoes, carrots, and beans) while their 3D shape was reconstructed. The carbohydrate content of foods was then calculated by combining the volume of each food item with the nutritional information provided by a nutrition database. GoCARB estimates were compared with participant estimates of carbohydrate content and the ground truth (measured by weighing the meals and calculating carbohydrates using the same nutrition database). The mean relative error in carbohydrate estimation was 54.8% (SD 72.3%) for the estimations of participants and 26.2% (SD 18.7%) for the GoCARB app.

### Alcohol

#### *Overview*

A total of 16 papers examined the measurement properties of a smartphone-based approach to assess alcohol use (Multimedia Appendix 2; Table 3 for full study details). A total of 1453 participants were included in these 16 studies (range 0-671; age range 16-74 years; 510/1453, 35.09% female). Moreover, 62% (10/16) of these studies described self-report approaches, 2% (2/16) described active objective approaches, and 25% (4/16) described passive objective approaches to measuring alcohol use. The most commonly assessed measurement properties were criterion and construct validity. Although numerous apps measuring alcohol use are described here, only 1 app (Intellidrink [66]) is currently accessible via the leading app stores for consumers to monitor their own alcohol use (3 other apps [67,68,89], although publicly available, are only available for use by researchers for data collection).

**Table 3.** Key characteristics of studies examining the measurement of alcohol via a smartphone.

| Study | Country | App name | Publicly available | Measurement approach | Measurement properties assessed | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Reliability | Measurement error | Content validity | Construct validity | Criterion validity | Responsiveness | Risk of bias |
| Arnold et al [3] | United States | AlcoGait | NR[a] | Passive objective | —[b] | — | — | — | ✓ | ✓ | Very good |
| Bae et al [67] | United States | AWARE | Yes | Passive objective | — | — | — | ✓ | — | ✓ | Very good |
| Barrio et al [74] | Spain | SIDEAL | No | Self-report | — | — | — | ✓ | — | — | Doubtful |
| Bernhardt et al [90] | United States | HAND | NR | Self-report | — | — | — | ✓ | — | — | Very good |
| Dulin et al [72] | United States | LBMI-A[c] | NR | Self-report | — | ✓ | — | ✓ | — | — | Adequate |
| Kim et al [76] | United States | SPAQ[d] | NR | Active objective | ✓ | ✓ | — | — | — | — | Inadequate |
| Kizakevich et al [73] | United States | PHIT[e] for duty | No | Self-report | — | — | — | ✓ | — | — | Very good |
| Luczak et al [66] | United States | Intellidrink | Yes | Self-report | — | — | — | ✓ | — | — | Very good |
| Matsumuraet al [77] | Japan | Spiral | NR | Active objective | ✓ | — | — | — | — | ✓ | Very good |
| McAfee et al [78] | United States | AlcoGait and AlcoWear Smartwatch app | N/A[f] | Passive objective | — | — | — | ✓ | — | ✓ | Doubtful |
| Monk et al [71] | United Kingdom | NR | NR | Self-report | — | — | — | ✓ | — | — | Very good |
| Paolillo et al [70] | United States | NR | N/A | Self-report | — | — | ✓ | ✓ | — | — | Very good |
| Poulton et al [89] | Australia | CNLab-A | Yes | Self-report | — | — | — | ✓ | — | — | Adequate |
| Santani et al [79] | Switzerland | Sensor logger and Drink logger | No | Passive objective | — | — | — | ✓ | — | ✓ | Doubtful |
| Swendeman et al [68] | United States | Ohmage | Yes | Self-report | — | ✓ | — | ✓ | — | — | Doubtful |
| Wray et al [69] | United States | Metricwire | No | Self-report | — | — | — | ✓ | — | — | Adequate |

[a]NR: not reported.

[b]No reporting of measurement property assessed.

[c]LBMIA: Location-Based Monitoring and Intervention for Alcohol Use Disorders.

[d]SPAQ: Smartphone Addiction Questionnaire.

[e]PHIT: Personal Health Intervention Toolkit.

[f]N/A: not applicable.

### *Self-report*

#### Overview

Overall, 62.5% studies [66,68-74,89,90] examined the measurement properties of apps that asked users to self-report alcohol use. Most studies (7/10, 70%) asked participants to report alcoholic beverage consumption daily [68,72,90] or in real-time (ie, using EMA) [70,71,74,89] using an app. Furthermore, 2 studies [69,73] examined the validity and reliability of completing standardized measures of alcohol use disorder via a smartphone, and 1 study [66] examined the ability of self-reported alcohol consumption via a commercially available app to accurately estimate breath alcohol concentrations (BrACs).

#### Daily Self-report

Recording alcohol consumption once a day via a smartphone app was investigated in 3 studies [68,72,90]. These studies demonstrated that this approach possesses good convergent validity when compared with traditional recall methods such as the Timeline Follow Back (TLFB) [91], which is a calendar-prompted, retrospective measure of alcohol consumption. Swendeman et al [68] found a moderate correlation (0.65) between daily self-reports and web-based 14-day recall surveys of alcohol use. However, this study also found significant differences in the mean percentage of days of alcohol use, with higher reports via daily app-based self-reports compared with 14-day recall. Similarly, Dulin et al [72] found moderate to strong correlations between data recorded through an app they tested and the TLFB for percentage of days abstinent (0.76-0.92), percentage of heavy drinking days (0.49-0.74), and the number of drinks consumed per drinking day (0.49-0.74). However, these correlations were found to diminish as more time elapsed between consumption and recall.

#### EMA Apps

A total of 4 studies [70,71,74,89] examined the measurement properties of self-reported alcohol use recorded via a smartphone using EMA, that is, as it occurred in real time or close to real time. Each of these studies employed smartphone apps that asked participants if they had consumed alcohol since the last prompt or last submission of data and the quantity consumed. Participants were often instructed to record their alcohol use as it occurred; however, Paolillo et al [70], Poulton et al [89], and Monk et al [71] also proactively prompted participants to report their alcohol use multiple times each day. These studies each demonstrated real-time, self-reports of alcohol use via a smartphone to have some convergence with retrospective reports of alcohol use, particularly the TLFB (correlations of 0.42-0.95). However, Monk et al [71] also found that participants reported consuming more drinks when reporting in real time compared with retrospective reporting. In addition, Monk et al [71] found that more drinks consumed were related to higher discrepancies between real-time and retrospective reports. Poulton et al [89] also found that participants reported a significantly faster rate of consumption when recording in real time via an app, compared with retrospective accounts.

#### Standardized Measures of Alcohol Use Disorders

A total of 2 studies [69,73] examined the measurement properties of administering the Alcohol Use Disorders Identification Test (AUDIT), a standard measure of alcohol use with established reliability and validity, via a smartphone [75]. In their study, Kizakevich et al [73] compared the AUDIT completed via their app with pen and paper administration of the measure. Wray et al [69] asked participants to complete the AUDIT once a day for 30 days and compared this with the TLFB. Both studies provided some evidence for the validity of completing the AUDIT via a smartphone app. Kizakevich et al [73] demonstrated that there was very good convergence between the AUDIT completed on paper and via the app (0.97), whereas Wray et al [69] found that the AUDIT and web-based TLFB were moderately correlated (0.55-0.88). Wray et al [69] found evidence of underreporting alcohol use on the TLFB.

#### BrAC Apps

Luczak et al [66] investigated the ability of a transdermal alcohol concentration (TAC) device in combination with a commercially available app *Intellidrink* to estimate BrAC. TAC devices measure the amount of alcohol diffusing through the skin at a particular time. As the raw TAC data are not directly related to blood alcohol concentration (BAC) or BrAC, further information on consumed alcoholic drinks is required to calibrate the models that convert TAC data to BrAC. The Intellidrink app was used to allow participants to self-report basic demographic data and data for each drinking episode. These data were combined with the TAC data in the authors' BrAC estimator software to accurately estimate BrAC. The authors found that the BrAC algorithm combined with the Intellidrink app had good convergent validity when compared with results generated by the previously validated breath alcohol estimator software developed by the authors. The combination of TAC device and Intellidrink app calculated peak BrAC estimates (eBrAC) to within 0.0003% of that calculated by the BrAC estimator software when using raw breadth analyzer data. The Intellidrink calculated time of peak eBrAC was within 18 minutes of the reference data, and the area under the eBrAC curve was within 0.025% for alcohol hours.

### *Active Objective*

Two studies [76,77] investigated approaches to actively and objectively measure alcohol use via smartphones. One [77] examined the potential of a mobile-based test of psychomotor performance to measure alcohol-induced impairment, whereas the other [76] described the validation of an optical attachment for smartphones to identify the results of saliva alcohol concentration test strips.

In their study, Matsumura et al [77] tested the performance of participants on a mobile-based test of psychomotor performance (Spiral for iPhone) and 3 computer-based tests assessing psychomotor and cognitive performance at predrink baseline (BAC of 0%) and after alcohol consumption. When participants had a BAC close to 0.1%, their performance on all tests, including the Spiral for iPhone, was found to be significantly worse than baseline and 0% BACs. Although significant decreases in performance accuracy for the 3 computer-based tests were also found when participants had BACs close to

0.06%, performance on the smartphone-based test (Spiral for iPhone) was not significantly worse than baseline.

Kim et al [76] examined a custom-built smartphone attachment and smartphone app to capture an image of saliva alcohol concentration test strips and identify their correct saliva alcohol concentration. Using their system, they inserted test strips into the custom-built smartphone attachment, and images were captured using the camera of the smartphone. Their smartphone app used machine learning techniques to calculate the estimated saliva alcohol concentration. The authors used test strips prepared with various concentrations of ethyl alcohol to generate the training data. A total of 14 images were recorded for each concentration, but the study by Kim et al [76] did not report how many of these were used for training, and how many were used for unbiased testing of the trained machine learning algorithms. The authors reported that this approach to analyzing saliva alcohol concentrations is valid and reliable across different types of smartphones, providing average classification rates of 100% accuracy for standard concentrations (0%, 0.02%, 0.04%, 0.08%, and 0.3%) and 80% accuracy for intermediate concentrations that required finer discrimination.

### *Passive Objective*

#### Overview

A total of 4 studies examined the measurement properties of passive objective approaches to measure alcohol use. Arnold et al [3] and McAfee et al [78] used smartphone sensors [3] and a combination of smartphone and smartwatch sensors [78] to measure gait as a proxy for alcohol-induced impairment, whereas Santani et al [79] and Bae et al [67] used phone sensor data to infer alcohol use (see Multimedia Appendix 2 for full study details).

#### AlcoGait App

A total of 2 studies [3,78] investigated whether a smartphone user's level of alcohol intoxication could be accurately inferred from their gait. Both studies used the AlcoGait app, which runs continuously in the background of smartphones of users. In the study by Arnold et al [3], accelerometer data were collected by the app, and information about users' gait generated. This information was then labeled the following day using an in-app survey that asked users to identify when they began drinking and finished drinking and how many drinks they had. Machine learning algorithms were trained with these data to infer BAC as membership of one of the three classes: 0 to 2 drinks, 3 to 6 drinks, or >6 drinks. In their study, McAfee et al [78] extended the AlcoGait app with the AlcoWatch to create the AlcoWear system, which also uses gyroscopes to capture information on the rotational velocity of the smartphone in response to the user's movement.

Both studies generated evidence for the validity of this approach. Arnold et al [3] found that after training the system on 209 data points, the AlcoGait app could classify the alcohol consumption of a user into 1 of the 3 classes with an accuracy of 56% (*F* score of 0.629 and area under the receiver operating characteristic curve [AROC] of 0.685) on training data. They reported a higher performance of 70% (*F* score of 0.786 and AROC of 0.825) on yet unseen data. McAfee et al [78] used 33 participants wearing sensor-impairment goggles to simulate the effects of alcohol consumption on the body. Training data were gathered by extracting features such as step count, cadence, and sway from 90-second walks with sensor-impairment goggles simulating BAC in 4 ranges (0.04-0.06, 0.08-0.15, 0.15-0.25, and 0.25-0.35). These training data were then used to train and validate several machine learning algorithms. They found that the AlcoGait app was able to infer the correct BAC range with an accuracy of 89.45% with 99% of the data used for training and 1% used for validation. The authors reported a maximum accuracy of 79.8% when using the smart watch to infer BAC as being higher or lower than 0.08.

#### Smartphone Sensors

Overall, 2 studies [67,79] examined the use of data from multiple smartphone sensors and machine learning to automatically recognize drinking behavior. They both used apps that run in the background on user's phones to collect sensor data from participants' phones. Bae et al [67] used the app AWARE to collect data continuously over 28 days from 38 young adults with hazardous drinking. They collected information relating to time (eg, day of week or time of day), movement (eg, accelerometer or gyroscope), communication (phone calls or texts), and psychomotor impairment (keystroke speed; available for Android phones only) and used these data to train random forests to predict periods of no drinking, low-risk drinking, and high-risk drinking from historic data (1- and 3-day history). Alcohol use information was collected via SMS text messages sent at 10 AM each day asking about the previous day. The performance of their algorithms was tested using 20% of the data not previously used for training.

Santani et al [79] used the Android app SensorLogger to collect information related to location (GPS or Wi-Fi), movement (accelerometer), social context (density of nearby Bluetooth devices), and phone use (battery, screen, and app use) on 10 weekend nights from 8 PM to 4 AM from 241 participants. The DrinkLogger app was then used to allow participants to report their alcohol consumption when it occurred. After preprocessing, 1011 user nights from 160 individuals were used to train a random forest algorithm with 500 trees to predict whether a user had consumed alcohol on a given night.

Bae et al [67] provided some evidence of the validity of using mobile phone sensors and machine learning algorithms to identify alcohol use among young people. They found that drinking categories were significantly correlated with time of day (0.11) and day of week (0.06), claiming that with time of day and day of week alone, they were able to detect low- and high-risk drinking with 90% accuracy. Their best-performing model to predict drinking used random forests and 3 days of historical data from multiple sensors. The model had a Cohen κ of 0.80 and an AROC of 0.96 and correctly classified 30-minute windows of time as nondrinking 98.5% of the time, low-risk drinking 70.2% of the time, and high-risk drinking 84.4% of the time. The AROC is a measure of the ability of an algorithm to achieve high sensitivity as well as high specificity and has a maximum value of 1, indicating a perfect classifier. Random predictions (of drinking in this case) would result in an AROC score of 0.5. In contrast, Santani et al [79] found that

even data from their most informative features (accelerometer data) could only identify drinking nights with 75.8% accuracy. This was followed by location, Wi-Fi, and Bluetooth logs with 68.5%, 65.2%, and 64.2% accuracy, respectively (note that a random guess would have resulted in an accuracy of 67%, as 67% of the data used reported alcohol consumption for that night).

## Tobacco

### *Overview*

A total of 8 studies [68,80-85,92] described the measurement properties of approaches assessing tobacco use. The number of participants involved in these studies ranged from 3 to 146 (N=363). The studies involved, on average, 34.7% female participants ranging in age from 18 to 64 years. The key characteristics of the 8 studies are shown in Table 4 (see Multimedia Appendix 2 for full study details). Furthermore, 50% (4/8) of the studies described active objective approaches to measure smoking, 12% (1/8) used a self-report method, and 37% (3/8) described passive objective approaches. All 8 studies assessed the construct validity (specifically convergent validity) of their approaches. Although several different apps are described in the included studies, only 4 apps are accessible via the leading app stores (Instant Heart Rate, Cardio [83], Smokerlyzer [82], and SmokeBeat [84]), and designed to help consumers monitor their own tobacco use.

**Table 4.** Key characteristics of studies examining the measurement of tobacco via a smartphone.

| Study | Country | App name | Publicly available | Measurement approach | Measurement properties assessed | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Reliability | Measurement error | Construct validity | Criterion validity | Responsiveness | Risk of bias |
| Dar [84] | Israel | SmokeBeat | Yes | Passive objective | __[a] | — | ✓ | — | — | Doubtful |
| Herbec et al [83] | United Kingdom | *Instant Heart Rate* or *Cardio* | Yes | Active objective | — | — | ✓ | — | ✓ | Doubtful |
| McClure et al [80] | United States | My Mobile Monitor | No | Active objective | — | — | ✓ | — | — | Doubtful |
| Meredith et al [81] | United States | NR[b] | NR | Passive objective | ✓ | — | ✓ | ✓ | ✓ | Doubtful |
| Qin et al [92] | Canada | NR | NR | Passive objective | — | — | ✓ | — | — | Doubtful |
| Shoaib et al [85] | Netherlands | NR | No | Passive objective | — | — | ✓ | — | — | Very good |
| Swendeman et al [68] | United States | Ohmage | Yes | Self-report | — | ✓ | ✓ | — | — | Doubtful |
| Wong et al [82] | Malaysia | Smokerlyzer and iCOSmokerlyzer | Yes | Active objective | ✓ | — | ✓ | ✓ | — | Doubtful |

[a]No reporting of measurement property assessed.

[b]NR: not reported.

### *Self-report*

One study [68] examined the measurement properties of a self-report method to assess tobacco use. In this study, HIV-positive adults were asked to complete daily mobile surveys when prompted by the app, and whenever they smoked, for 6 weeks. Participants were asked to indicate if they had smoked since the last time they self-reported via the app. This study demonstrated that there was very good convergent validity among daily mobile self-reports and web-based 14-day recall surveys of tobacco use, with a strong correlation between methods (0.92).

### *Active Objective*

#### Overview

Of the 4 studies that examined active objective measures of tobacco use, 3 (75%) [80-82] investigated the measurement of expired CO using a smartphone app in conjunction with expired CO monitors, and 1 (25%) [83] investigated whether heart rate measured by a smartphone could accurately identify smoking episodes.

#### Expired CO

A total of 3 studies used this methodology [80-82], and all found evidence to support its validity and reliability. Overall, 67% (2/3) of these studies [81,82] used expired CO monitors designed to attach directly to smartphone users. Meredith et al [81] described the use of a prototype CO monitor for smartphones, developed by the authors, and Wong et al [82] examined the first commercially available CO monitor for use with a smartphone and accompanying smartphone app (iCO Smokerlyzer and the Smokerlyzer app). Both studies found that the first and second CO measures collected via their smartphone CO monitors were strongly and significantly correlated with

each other (0.98 and 0.94, respectively). Both studies also found that measurements of expired CO collected via smartphone-attached CO monitors were strongly correlated with measurements collected via stand-alone CO monitors. A third study [80] described a protocol whereby young smokers (aged 15-25 years) used a smartphone app (MyMobile Monitor) and the camera of their smartphone to take time-stamped photographs of themselves exhaling into a stand-alone expired CO monitor (PiCo Smokerlyzer), and a photograph of the CO readings displayed by the monitor to be verified by the research staff. This study found a moderate agreement among the methods (0.49).

#### Heart Rate Apps

One study [83] used 2 publicly available heart rate apps (*Instant Heart Rate* and *Cardio*) to investigate whether resting heart rate measured using a smartphone could be used to verify smoking abstinence. The study by Herbec et al [83] of 18 adult daily smokers found some evidence to support this approach. Specifically, they found that lower heart rates were observed among all participants on days they did not smoke and did not use nicotine replacement products, compared with days on which they smoked as usual. Similarly, lower heart rates were also observed among 83% (15/18) of participants on days they were abstinent but used a nicotine replacement product compared with those on days when they smoked as usual.

### *Passive Objective*

#### Overview

A total of 3 studies [84,85,92] investigated passive objective approaches to measure tobacco use via smartphones. Overall, 67% (2/3) of these studies [84,85] used wrist-worn sensor devices (eg, smartwatches) in conjunction with smartphone apps to detect episodes of smoking. A third [92] used in-phone sensors only to recognize the occurrence of smoking might be taking place.

#### Wrist-Worn Sensors

Shoaib et al [85] used accelerometer and gyroscope data collected from smartwatches and smartphones to test a 2-layer hierarchical smoking detection algorithm. In their study, participants wore a smartwatch on their right wrist and a smartphone in their right pocket. A total of 11 participants performed 17 hours (230 cigarettes) of smoking while sitting, standing, walking, and in group conversation and 28 hours of other similar activities (eg, eating and drinking). Data were collected at 50 samples per second from these sensors. Dar [84] provided participants with smartwatches and instructed them to wear them on the hand that they used for smoking. Dar [84] then used the SmokeBeat app to process raw data from these devices, identify smoking episodes, and provide feedback to participants in real time. These 2 studies demonstrated very good convergent validity with self-reported smoking episodes. Shoaib et al [85] achieved a very high precision and recall for smoking in 83% to 97% *F*-measure, whereas Dar [84] detected 82.29% of smoking episodes, with a negligible frequency of erroneously detected episodes (2.85%).

#### In-Phone Sensors

One study [92] used data collected from the GPS, Wi-Fi, and accelerometer within the smartphones of participants and self-reported smoking behaviors, collected over 1 month to train and evaluate algorithms to accurately classify smoking and nonsmoking periods based on in-phone sensor data alone. First, each of the individual features extracted from the sensor data collected was used to train univariate hidden Markov models (HMMs), which were then evaluated. Next, multivariate HMMs using 3 features and 5 features were trained and evaluated. Qin et al [92] were able to detect smoking activity with an accuracy over 0.9, and an AROC of >0.8. HMMs with a single feature were found less accurate than multivariate HMMs.

### Risk of Bias

Most studies (35/48, 73%) investigating the measurement of diet received *very good* (23/48, 48%) or *adequate* (12/48, 25%) methodological quality ratings on the COSMIN Risk of Bias tool (Table 2). As Table 3 shows, most studies (11/16, 69%) investigating the measurement of alcohol use were of at least adequate methodological quality (*very good*, 8/16, 50%; and *adequate*, 3/16, 19%), whereas only 13% (1/8) of the studies investigating the measurement of tobacco use received a *very good* rating (Table 4).

Across the 3 behaviors, 73% (29/40) of studies investigating self-report measurement approaches received *very good* or *adequate* quality ratings. This is compared with 60% (15/25) of studies investigating active objective approaches and 38% (3/8) of studies investigating passive objective approaches. Ratings of *adequate* or *very good* were achieved by 72% (21/40) of diet self-report studies, 74% (14/19) diet active objective studies, and 63% (5/8) alcohol self-report studies. By contrast, *adequate* or *very good* ratings were achieved by only 50% (1/2) of alcohol active objective and 50% (2/4) alcohol passive objective studies, by 25% (1/4) of tobacco passive objective studies, and by none of the tobacco self-report and active objective studies.

The 3 most commonly investigated measurement properties were also found to be the most rigorously examined. Overall, 71% (22/31) of studies examining measurement error, 96% (26/27) of studies examining criterion validity, and 76% (35/46) of studies examining construct validity received very good or adequate ratings for their examination of these measurement properties. In contrast, reliability and responsiveness were only examined with very good or adequate methodological quality in 29% (2/7) and 40% (2/5) of the studies, respectively.

## Discussion

### Principal Findings

#### *Overview*

This systematic review is the first to bring together the existing evidence of the measurement properties of smartphone-based approaches to measure three key lifestyle behaviors—diet, alcohol use, and tobacco use. Overall, there was some evidence to support the reliability and validity of using smartphones to assess these behaviors. However, results varied by behavior and

measurement approach, and the methodological quality of studies ranged from inadequate to very good. To an extent this large range can be attributed to the significant number of new technologies and methodologies being designed and tested for automated measurements of the behaviors of interest. These methods are not yet mature enough but may provide exciting opportunities as they develop further.

Studies most commonly focused on approaches to assess dietary behaviors (48/72, 67%), with only 22% (16/72) and 11% (8/72) measuring alcohol and tobacco use, respectively. Across the health behaviors, 19 different smartphone-based measurement techniques were used, the most commonly examined approach being food diary apps. Most studies investigated the construct validity, criterion validity, or measurement error associated with their measurement technique, whereas few examined their reliability or responsiveness. Although a wide range of smartphone apps were described in the included studies, most of these apps are not currently publicly available or are designed as research data collection tools, rather than tools that can be easily used by clinicians or consumers to monitor diet, alcohol use, or tobacco use behaviors.

The highest quality evidence was found for diet, with most studies examining diet being rated as very good or adequate (35/48, 73%). This was compared with those assessing alcohol use (11/16, 69%) and tobacco use (1/8, 13%). In light of the fewer number of studies and poorer quality of evidence available, conclusions and recommendations drawn from the existing literature regarding smartphone-based measurement of alcohol and tobacco use should be interpreted with caution.

### *Diet*

Diet was most commonly examined using self-report methods. These studies indicated that food diary apps, in particular MyFitnessPal, can be a reliable and valid method of measuring energy intake. Individual studies investigating other smartphone-based self-report methods, including 24-hour recall, web-based food databases, EMA, and food diary apps using unstructured data entry methods, have demonstrated promising results. However, there is currently insufficient evidence supporting the reliability or validity of these approaches, and further research is required. A growing body of literature suggests that manually analyzed food photography may be valid and reliable in a general adult population. However, because of the need for highly trained individuals to analyze every captured image, this approach is unlikely to be scalable or sustainable outside of a research context.

This review identified a small body of literature investigating the novel approach of using smartphones to capture images and voice, extract food intake information from these data, and access external databases to retrieve nutrient information. These studies relied on spoken reports by users or on the use of machine learning to automatically recognize food items (and their size) in photographs. Although the results are encouraging, most of these studies confined their investigations to a small number of food items, with tests performed on a small number of participants. Hence, the generalizability of these results cannot be assessed. Given the large variation in the appearance of food in a global society, such approaches will likely require vast amounts of varied training data to be of general applicability. Promisingly, the lower burden this automatic analysis approach places on users and administrators and its potential to provide real-time feedback to users mean that this approach is potentially scalable and could be a powerful addition to eating behavior interventions. Both manually and automatically analyzed food photography methods address some of the key issues associated with traditional methods for measuring diet behavior. As long as users remember to take a photograph of their food, these methods provide an objective record of their food intake reduction issues associated with recall bias, and the use of fiducial markers (as was common in included studies) reduces the reliance on users to be able to accurately estimate portion sizes.

An important limitation of most of the studies that investigated the measurement of diet behaviors was small sample sizes, with 77% (37/48) of the studies involving under 100 participants, including 50% (24/48) of the studies with <50 participants.

### *Alcohol*

As with diet, most studies assessing alcohol use used self-report methods. A strength of this literature is the frequent (though not universal) use of a common comparison measure, the TLFB [91]. The included studies provided good evidence for the validity of daily and real-time self-reporting of alcohol use via smartphones, with moderate to strong correlations with retrospective reports of alcohol use found across studies. However, several of these studies [68,71,72,89] also found that participants reported greater alcohol use via smartphone-based reporting compared with retrospective reports of alcohol use, such as the TLFB. These discrepancies were interpreted by a number of authors [71,72,89] as evidence that underreporting of alcohol use occurs when using recall methods and that app-based self-reports of alcohol use may be able to provide a better understanding of alcohol intake [89]. This interpretation is problematic, as the TLFB is widely acknowledged as the *gold standard* measure for self-reported alcohol use. Unfortunately, none of the included studies also used an objective measure of alcohol use, which may have elucidated this finding and would have allowed a comparison of the accuracy between app-based self-report and the TLFB.

Insufficient evidence for the smartphone administration of AUDIT [93], a standardized measure of alcohol use disorder, has been generated to date. Similarly, although the results of a study examining if BrACs could be accurately estimated via a smartphone app were positive, there is insufficient evidence at this stage supporting this approach. Two studies assessed novel active objective methods for measuring alcohol use. For example, the study by Matsumura et al [77] suggests that smartphone-based measures of psychomotor performance may be able to validate alcohol-induced impairment. However, more studies involving larger sample sizes are required before it can be determined if initial promising results are representative of the true measurement properties of these approaches.

Although only 4 studies focused on passive objective measurement of alcohol use, these findings suggest that using in-built phone sensors to infer and even predict alcohol use may be a promising assessment method. However, some

methodological issues are worth noting. For example, McAfee [78] likely severely *overfit* their data by training their algorithm on 99% of their data, meaning that their approach is unlikely to be able to generalize to a new data set. Further research is needed before the validity and reliability of these types of methods can be established and will likely include the gathering of large amounts of data. To be of most use to clinicians, consumers, and researchers who are interested in passively measuring alcohol use, easy-to-use interfaces that automatically process sensor data (preferably in real time) and feedback results are needed.

A limitation of most studies that investigated measurement of alcohol use was the small sample size, with 81% (13/16) of the studies involving samples of <100 participants (11/16, 69%, with <50 participants). In addition, many included studies were conducted within specific populations, such as college students, people with HIV, and military personnel, which may limit the generalizability of results to the broader population.

### *Tobacco*

Measuring tobacco use with smartphones has been examined by the fewest number of studies. Unlike other behaviors, most tobacco use studies have focused on objective measurement techniques rather than self-reports. A total of 3 studies supported the methodological soundness of measuring expired CO using smartphones (and expired CO monitors). Using apps that measure users' heart rate was also found to be a promising way to quickly and easily verify smoking abstinence. Passive measurement approaches using wrist-worn and in-phone sensors also show promise.

Although the results from these individual studies are promising, further research is needed to establish the validity and reliability of these types of objective approaches. In addition, most studies involved very small samples (7/8, 88%, involved <100 participants; and 6/8, 75%, involved <50 participants).

## Strengths and Weaknesses of Measurement Approaches

This study identified three key approaches used to measure diet, alcohol use, and tobacco use via smartphones: self-report, active objective, and passive objective approaches. Across behaviors, several key strengths and weaknesses associated with these approaches have emerged. To date, most evidence has been generated for self-report and smartphone-based measurement approaches. These approaches are most similar to traditionally used measurement approaches and often involve simply asking users to complete existing validated measures of behaviors by interacting with the touchscreen of their smartphone, rather than completing them using a pen and paper survey or an interview. Moving self-report measures onto a smartphone, a device that many people carry with them and that can automatically calculate summary information, improves upon traditional measurement approaches by facilitating real-time recording of health behaviors and providing feedback to users—a potentially powerful intervention tool [94]. Although many self-report methods have been shown to be reliable and valid, particularly for diet, these approaches remain burdensome and require considerable input from the user. It is also likely that smartphone-based self-report measures continue to suffer from biases similar to traditional self-report systems, for example, as response bias and declining accuracy over time.

This review identified several novel approaches to objectively measure diet, alcohol use, and tobacco use, both with and without the active involvement of participants. Fewer studies that used objective approaches (active and passive) received a quality rating of *very good*, compared with self-report approaches. This perhaps is not so much a criticism of these approaches as an acknowledgment that many of these studies used innovative machine learning methods with limited data sets and require further investigation before they can be considered mature. The obvious strength of these approaches is that they have the potential to provide objective measurements of consumption behaviors, which have traditionally been primarily assessed using self-report measures. These approaches can address issues with reporting accuracy, recall bias, and memory. However, for alcohol and tobacco use, the objective, smartphone-based measurement approaches developed to date do not directly assess these behaviors (as is the case for diet with food photography methods). Rather, these approaches use proxy measures related to the physiological response to the behaviors (eg, measuring CO content or BAC, or measuring gait to infer alcohol intoxication) or infer the physical movements associated with the behaviors (eg, hand movements to infer cigarette smoking). For all 3 behaviors, the results of active objective measurements suggested that, although these methods have good potential to significantly reduce the user burden and recall bias, they can still be quite burdensome for users and may not be particularly scalable, as for manually analyzed food photography methods.

Although passive objective approaches may address the issue of participant burden by collecting information from smartphones without the involvement of users, this continuous collection and storage of sensor data from phones is associated with privacy and data security issues, which may mean that these powerful approaches are not acceptable to many people. However, previous research indicates that when employed for health purposes (eg, sharing sleep, mood, or physical activity information with a physician), most people are comfortable sharing passively sensed information, and characteristics such as age may not influence the comfort of individuals by sharing this sort of information [95]. With only 7 studies that described the measurement properties of passive objective approaches, more research is needed to establish the validity and reliability of these approaches. Although sufficient evidence may not yet exist to recommend the use of passive objective measurement approaches, these types of approaches have huge potential to augment health behavior change interventions. Information gathered in this way can potentially be used to provide tailored support *in the moment* to users, allowing relevant support to be delivered during a time and context when it is most salient [67].

## Recommendations and Future Directions

Although 72 studies were identified that aimed to describe novel smartphone-based approaches to measuring diet, alcohol use, and tobacco use, a major issue identified within this literature is the extreme heterogeneity in approaches and evaluation

methods investigated. Nineteen broad measurement techniques were described in studies included in the current review, and within these groups, almost every individual study described a different specific technique. For example, each food diary app described used a different way to record diet information, used different food databases to provide nutritional information of recorded food items and different methods for entering data. Similarly, the algorithms used to automatically analyze food images or indicate that alcohol and tobacco use from sensor data differed. The relatively short time smartphones have been available (approximately 13 years) and the relatively early stage of research in this area may explain the lack of homogeneity in the types of specific techniques and methods investigated.

Noting the above limitations of current knowledge in this area, clinicians and consumers looking for valid, reliable, and publicly available ways to assess diet and alcohol use behaviors might consider using food diary apps such as *MyFitnessPal* [18,22,32,34,40] and apps that assess alcohol use via daily or real-time self-reports (eg, *Intellidrink*) [66]. Smartphone-compatible CO monitors such as the iCoSmokerlyzer and Smkerlyzer apps are also promising ways to assess tobacco use [82]; however, further research in this area in particular is needed.

Although it is important to continue moving the field forward and investigate if new and better ways to measure consumption behaviors using smartphones can be developed, it is strongly recommended that researchers first look to the existing literature described here (and in other fields) to determine if, in the search for a way to measure diet, alcohol use, or tobacco use, using a smartphone, an existing technique, or an app may be appropriate for their purposes before considering development of yet another app. Agreed-upon standards for capturing the data and extracting higher-level information (such as nutrient information) would be a constructive way of ensuring that the data collected can be pooled with similar data from other initiatives, thus providing a larger and more robust data set for algorithm development.

Only 4 studies [42,43,64,65] described apps that assessed >1 behavior (specifically diet and physical activity behaviors). Building or identifying systems that allow easy and accurate measurement of multiple health behaviors would be a useful addition to the field as we know that health risk behaviors such as poor diet, substance use, physical inactivity, and poor sleep [96].

The heterogeneity of methods used to evaluate the measurement properties of techniques is another weakness of the current literature. Again, it is recommended that researchers examine the existing literature closely when designing their own studies. For example, it is suggested that the TLFB be considered as a comparison measure for smartphone-based approaches to measure alcohol use, as it has been most frequently used in the current literature. However, no such common comparison measure of diet and tobacco use has emerged from the literature to date. In addition, the accuracy of self-reported measures of these consumption behaviors has been questioned [68,71,72,89], and it has been suggested that newer measurement approaches, such as the smartphone-based approaches discussed here, may in fact provide data closer to the actual behaviors under investigation and may eventually be themselves considered the gold standard in the measurement of these behaviors. In the meantime, it is recommended that researchers consider investigating the validity of smartphone-based approaches in comparison with objective measurements of these behaviors. Indeed, this review identifies a lack of objective comparisons as a key weakness, with few studies (particularly for alcohol and tobacco use) investigating the criterion validity of these approaches. Similarly, other measurement properties, such as reliability and responsiveness, have rarely been investigated. To take full advantage of smartphones in research, in clinical settings, and within consumers' everyday lives, the full variety of measurement properties of these different approaches needs to be better understood.

To have the biggest impact on chronic disease, we need to make valid and reliable tools easily available to clinicians and consumers to allow for the collection of quality and detailed health behavior information. There is also a need for easy-to-use interfaces to facilitate the use of these passive sensing systems by clinicians and consumers. Quality and detailed information regarding diet, alcohol use, and tobacco use can be leveraged to help individual consumers acquire better insights into their own behaviors and inform tailored support. In other words, it is important that the apps used to measure diet, alcohol use, and tobacco use are publicly and freely available.

## Limitations

An important limitation of this review is that it included only studies published up to March 2020. In this rapidly growing area, there are likely to be recent and ongoing studies that have also investigated the measurement properties of smartphone-based approaches to measuring health behaviors. In addition, this review only captures approaches whose measurement properties have been examined and discussed in the published literature. It is likely that other novel and potentially effective approaches to measure diet, alcohol use, and tobacco use have been developed and are currently in use but that they have been developed outside of academia, their measurement properties have not been not specifically assessed, or they simply have not been published.

## Conclusions

Accurate measurement of diet, alcohol use, and tobacco use is central to successful chronic disease risk reduction interventions [8-11]. Therefore, identifying new and valid ways to measure these behaviors could have major public health implications. This review highlights measurement approaches that clinicians and researchers may want to consider implementing to help clients better measure and manage their health behaviors and improve the measurement of these behaviors in research settings. The results suggest that food diary apps, particularly the commercially available app MyFitnessPal, may be appropriate tools to measure diet. The review also highlights approaches with growing bodies of promising evidence but where more research is needed before their use might be recommended (eg, food photography methods and CO monitor smartphone attachments). Finally, the review highlights several measurement approaches with great potential but where only mixed evidence or evidence from 1 or 2 studies is available (eg,

smartphone-based measurement of psychomotor performance to infer alcohol intoxication; the use of smartphone and wrist-worn sensors to infer alcohol intoxication and detect or predict alcohol and tobacco use; and the use of heart rate monitor apps to infer smoking abstinence). These conclusions should not be interpreted as a criticism of these approaches but rather as an acknowledgment that many of these approaches use cutting-edge technologies, which require further research (and data) before they can be expected to yield accurate and generalizable results.

## Acknowledgments

This study was funded by the Paul Ramsay Foundation and the Australian National Health and Medical Research Council (fellowship to KC [APP1120641] and MT [APP1078407]). The funders of the study had no role in study design, data collection, data analysis, or data interpretation.

## Authors' Contributions

LT, BO, and KC led the development of the manuscript. MT, KC, TS, CC, BP, DL, and LT secured funding for the study. ABW led the literature search. LT, LG, BO, OG, CS, RV, KC, JW, ZB, LB, and JT conducted the data extraction. PVV assisted with the interpretation and write-up of the results and discussion. All authors contributed to the development of protocols for the study and reviewed, edited, and approved the final version of the paper.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[PDF File (Adobe PDF File), 67 KB - mhealth_v10i2e27337_app1.pdf ]

Multimedia Appendix 2
Additional details and results of included studies.
[PDF File (Adobe PDF File), 398 KB - mhealth_v10i2e27337_app2.pdf ]

## References

1. Rabbi M, Philyaw-Kotov M, Lee J, Mansour A, Dent L, Wang X, et al. SARA: a mobile app to engage users in health data collection. In: Proc ACM Int Conf Ubiquitous Comput. 2017 Sep Presented at: UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 11 - 15, 2017; Maui Hawaii p. 781-789 URL: http://europepmc.org/abstract/MED/29503985 [doi: 10.1145/3123024.3125611]
2. Mobile fact sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-01-17]
3. Arnold Z, Larose D, Agu E. Smartphone inference of alcohol consumption levels from gait. In: Proceedings of the International Conference on Healthcare Informatics. 2015 Presented at: International Conference on Healthcare Informatics; Oct. 21-23, 2015; Dallas, TX, USA. [doi: 10.1109/ichi.2015.59]
4. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
5. Johnson RK, Coward-McKenzie D. Energy requirement methodology. In: Coulston AM, Boushey CJ, Ferruzzi MG, editors. Nutrition in the Prevention and Treatment of Disease. Amsterdam: Elsevier; 2001:31-42.
6. Raina S. Limitations of 24-hour recall method: micronutrient intake and the presence of the metabolic syndrome. N Am J Med Sci 2013 Aug;5(8):498 [FREE Full text] [doi: 10.4103/1947-2714.117329] [Medline: 24083229]
7. Basiotis P, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr 1987 Sep;117(9):1638-1641. [doi: 10.1093/jn/117.9.1638] [Medline: 3655942]
8. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population. Am J Prevent Med 1999 Oct;17(3):211-229. [doi: 10.1016/s0749-3797(99)00069-0]
9. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
10. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]

11. Jenkins RJ, McAlaney J, McCambridge J. Change over time in alcohol consumption in control groups in brief intervention studies: systematic review and meta-regression study. Drug Alcohol Depend 2009 Feb 01;100(1-2):107-114. [doi: 10.1016/j.drugalcdep.2008.09.016] [Medline: 19041196]
12. McCambridge J. [Commentary] Research assessments: instruments of bias and brief interventions of the future? Addiction 2009 Aug;104(8):1311-1312. [doi: 10.1111/j.1360-0443.2009.02684.x] [Medline: 19624324]
13. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. Diabetes Care 2001 Mar;24(3):561-587. [doi: 10.2337/diacare.24.3.561] [Medline: 11289485]
14. Thornton L, Osman B, Wescott AB, Sunderland M, Champion K, Green O, Health4Life Team. Measurement properties of smartphone approaches to assess key lifestyle behaviours: protocol of a systematic review. Syst Rev 2020 Jun 03;9(1):127 [FREE Full text] [doi: 10.1186/s13643-020-01375-w] [Medline: 32493467]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339(jul21 1):b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
16. COSMIN definitions of domains, measurement properties, and aspects of measurement properties. COSMIN. URL: https://cosmin.nl/wp-content/uploads/COSMIN-definitions-domains-measurement-properties.pdf [accessed 2022-01-17]
17. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018 May 19;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
18. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet 2018 Apr 27;75(2):219-225. [doi: 10.1111/1747-0080.12401] [Medline: 29280547]
19. Rangan AM, Tieleman L, Louie JC, Tang LM, Hebden L, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): relative validity of a mobile phone application to measure intake of food groups. Br J Nutr 2016 Apr 28;115(12):2219-2226. [doi: 10.1017/s0007114516001525]
20. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): comparison of a mobile phone digital entry app for dietary data collection with 24-hour dietary recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98 [FREE Full text] [doi: 10.2196/mhealth.4613] [Medline: 26508282]
21. Pendergast FJ, Ridgers ND, Worsley A, McNaughton SA. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act 2017 Mar 14;14(1):30 [FREE Full text] [doi: 10.1186/s12966-017-0488-9] [Medline: 28288657]
22. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Mar 14;21(8):1495-1502. [doi: 10.1017/s1368980018000393]
23. Torre SB, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1):76 [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
24. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2012 May 03;109(3):539-546. [doi: 10.1017/s0007114512001353]
25. Béjar LM. First evaluation steps of a new method for dietary intake estimation regarding a list of key food groups in adults and in different sociodemographic and health-related behaviour strata. Public Health Nutr 2017 Aug 09;20(15):2660-2669. [doi: 10.1017/s1368980017001641]
26. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 09;17(1):5 [FREE Full text] [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]
27. Ali ZC, Silvioli R, Rajai A, Aslam TM. Feasibility of use of a mobile application for nutrition assessment pertinent to age-related macular degeneration. Transl Vis Sci Technol 2017 Jan 01;6(1):4 [FREE Full text] [doi: 10.1167/tvst.6.1.4] [Medline: 28138414]
28. Lancaster R, Radd-Vagenas S, Singh MF, Noble Y, Daniel K, Mavros Y, et al. Electronic food records among middle-aged and older people: a comparison of self-reported and dietitian-assisted information. Nutr Diet 2021 Apr;78(2):145-153. [doi: 10.1111/1747-0080.12606] [Medline: 32037628]
29. Wellard-Cole L, Chen J, Davies A, Wong A, Huynh S, Rangan A, et al. Relative validity of the Eat and Track (EaT) smartphone app for collection of dietary intake data in 18-to-30-year olds. Nutrients 2019 Mar 14;11(3):621 [FREE Full text] [doi: 10.3390/nu11030621] [Medline: 30875772]
30. Liu Y, Chen C, Tsou Y, Lin Y, Chen H, Yeh J, et al. Evaluating mobile health apps for customized dietary recording for young adults and seniors: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e10931 [FREE Full text] [doi: 10.2196/10931] [Medline: 30767906]
31. Lemacks J, Adams K, Lovetere A. Dietary intake reporting accuracy of the Bridge2U mobile application food log compared to control meal and dietary recall methods. Nutrients 2019 Jan 19;11(1):199 [FREE Full text] [doi: 10.3390/nu11010199] [Medline: 30669430]
32. Fallaize R, Franco RZ, 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]

33. Hezarjaribi N, Mazrouee S, Hemati S, Chaytor NS, Perrigue M, Ghasemzadeh H. Human-in-the-loop learning for personalized diet monitoring from unstructured mobile data. ACM Trans Interact Intell Syst 2019 Dec 05;9(4):1-24. [doi: 10.1145/3319370]
34. Chen J, Berkman W, Bardouh M, Ng CY, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
35. Rodder S, Kindratt T, Xiao C, Orcutt V, Koch C, Mcilvaine K, et al. Teaching and evaluating smartphone applications: the effectiveness of a curriculum expansion. Educ Health 2018;31(2):95. [doi: 10.4103/efh.efh_322_16]
36. Béjar LM, García-Perea MD, Reyes OA, Vázquez-Limón E. Relative validity of a method based on a smartphone app (Electronic 12-hour Dietary Recall) to estimate habitual dietary intake in adults. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11531 [FREE Full text] [doi: 10.2196/11531] [Medline: 30973343]
37. Chmurzynska A, Mlodzik-Czyzewska M, Malinowska A, Czarnocinska J, Wiebe D. Use of a smartphone application can improve assessment of high-fat food consumption in overweight individuals. Nutrients 2018 Nov 06;10(11):1692 [FREE Full text] [doi: 10.3390/nu10111692] [Medline: 30404207]
38. Hezarjaribi N, Mazrouee S, Ghasemzadeh H. Speech2Health: a mobile framework for monitoring dietary composition from spoken data. IEEE J Biomed Health Inform 2018 Jan;22(1):252-264. [doi: 10.1109/jbhi.2017.2709333]
39. Béjar LM, Reyes OA, García-Perea MD. Electronic 12-hour dietary recall (e-12HR): comparison of a mobile phone app for dietary intake assessment with a food frequency questionnaire and four dietary records. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10409 [FREE Full text] [doi: 10.2196/10409] [Medline: 29907555]
40. Ahmed M, Mandic I, Lou W, Goodman L, Jacobs I, L'Abbé MR. Validation of a tablet application for assessing dietary intakes compared with the measured food intake/food waste method in military personnel consuming field rations. Nutrients 2017 Feb 27;9(3):200 [FREE Full text] [doi: 10.3390/nu9030200] [Medline: 28264428]
41. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
42. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A Mobile Ecological Momentary Assessment Tool (devilSPARC) for Nutrition and Physical Activity Behaviors in College Students: A Validation Study. J Med Internet Res 2016 Jul 27;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
43. Swendeman D, Comulada WS, Koussa M, Worthman CM, Estrin D, Rotheram-Borus MJ, et al. Longitudinal validity and reliability of brief smartphone self-monitoring of diet, stress, and physical activity in a diverse sample of mothers. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e176 [FREE Full text] [doi: 10.2196/mhealth.9378] [Medline: 30249576]
44. Smith L, Hua J, Seto E, Du S, Zang J, Zou S, et al. Development and validity of a 3-day smartphone assisted 24-hour recall to assess beverage consumption in a Chinese population: a randomized cross-over study. Asia Pac J Clin Nutr 2014;23(4):678-690 [FREE Full text] [doi: 10.6133/apjcn.2014.23.4.10] [Medline: 25516327]
45. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
46. Zhang W, Yu Q, Siddiquie B, Divakaran A, Sawhney H. "Snap-n-Eat": food recognition and nutrition estimation on a smartphone. J Diabetes Sci Technol 2015 May;9(3):525-533 [FREE Full text] [doi: 10.1177/1932296815582222] [Medline: 25901024]
47. Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey RH, et al. Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. J Med Internet Res 2016 May 11;18(5):e101 [FREE Full text] [doi: 10.2196/jmir.5567] [Medline: 27170498]
48. Nicklas T, Saab R, Islam NG, Wong W, Butte N, Schulin R, et al. Validity of the remote food photography method against doubly labeled water among minority preschoolers. Obesity (Silver Spring) 2017 Sep 31;25(9):1633-1638 [FREE Full text] [doi: 10.1002/oby.21931] [Medline: 28758370]
49. Liu Y, Chen C, Lee C, Lin Y, Chen H, Yeh J, et al. Design and usability evaluation of user-centered and visual-based aids for dietary food measurement on mobile devices in a randomized controlled trial. J Biomed Inform 2016 Dec;64:122-130 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.001] [Medline: 27720984]
50. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 2008 Jul 11;101(3):446-456. [doi: 10.1017/s0007114508027438]
51. Martin C, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
52. Kong K, Zhang L, Huang L, Tao Y. Validity and practicability of smartphone-based photographic food records for estimating energy and nutrient intake. Asia Pac J Clin Nutr 2017 May;26(3):396-401 [FREE Full text] [doi: 10.6133/apjcn.042016.05] [Medline: 28429902]
53. Huang J, Ding H, McBride S, Ireland D, Karunanithi M. Use of smartphones to estimate carbohydrates in foods for diabetes management. Stud Health Technol Inform 2015;214:121-127. [Medline: 26210428]

54. Nyström CD, Forsum E, Henriksson H, Trolle-Lagerros Y, Larsson C, Maddison R, et al. A mobile phone based method to assess energy and food intake in young children: a validation study against the doubly labelled water method and 24 h dietary recalls. Nutrients 2016 Jan 15;8(1):50 [FREE Full text] [doi: 10.3390/nu8010050] [Medline: 26784226]
55. Boushey C, Spoden M, Delp E, Zhu F, Bosch M, Ahmad Z, et al. Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients 2017 Mar 22;9(3):312 [FREE Full text] [doi: 10.3390/nu9030312] [Medline: 28327502]
56. Costello N, Deighton K, Dyson J, Mckenna J, Jones B. Snap-N-Send: a valid and reliable method for assessing the energy intake of elite adolescent athletes. Eur J Sport Sci 2017 Sep 17;17(8):1044-1055. [doi: 10.1080/17461391.2017.1337815] [Medline: 28627289]
57. Ashman A, Collins C, Brown L, Rae K, Rollo M. Validation of a smartphone image-based dietary assessment method for pregnant women. Nutrients 2017 Jan 18;9(1):73 [FREE Full text] [doi: 10.3390/nu9010073] [Medline: 28106758]
58. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The use of mobile devices in aiding dietary assessment and evaluation. IEEE J Sel Top Signal Process 2010 Aug;4(4):756-766. [doi: 10.1109/jstsp.2010.2051471]
59. Rollo ME, Ash S, Lyons-Wall P, Russell A. Trial of a mobile phone method for recording dietary intake in adults with type 2 diabetes: evaluation and implications for future applications. J Telemed Telecare 2011 Aug 15;17(6):318-323. [doi: 10.1258/jtt.2011.100906]
60. Prinz N, Bohn B, Kern A, Püngel D, Pollatos O, Holl RW. Feasibility and relative validity of a digital photo-based dietary assessment: results from the Nutris-Phone study. Public Health Nutr 2018 Mar 6:1-8. [doi: 10.1017/s1368980018000344]
61. Rollo M, Ash S, Lyons-Wall P, Russell A. Evaluation of a mobile phone image-based dietary assessment method in adults with type 2 diabetes. Nutrients 2015 Jun 17;7(6):4897-4910. [doi: 10.3390/nu7064897]
62. Kato S, Waki K, Nakamura S, Osada S, Kobayashi H, Fujita H, et al. Validating the use of photos to measure dietary intake: the method used by DialBetics, a smartphone-based self-management system for diabetes patients. Diabetol Int 2015 Oct 14;7(3):244-251. [doi: 10.1007/s13340-015-0240-0]
63. Most J, Vallo PM, Altazan AD, Gilmore LA, Sutton EF, Cain LE, et al. Food photography is not an accurate measure of energy intake in obese, pregnant women. J Nutr 2018 Apr 01;148(4):658-663 [FREE Full text] [doi: 10.1093/jn/nxy009] [Medline: 29659958]
64. Schiel R, Kaps A, Bieber G. Electronic health technology for the assessment of physical activity and eating habits in children and adolescents with overweight and obesity IDA. Appetite 2012 Apr;58(2):432-437. [doi: 10.1016/j.appet.2011.11.021] [Medline: 22155072]
65. Schiel R, Kaps A, Bieber G, Kramer G, Seebach H, Hoffmeyer A. Identification of determinants for weight reduction in overweight and obese children and adolescents. J Telemed Telecare 2010 Aug 02;16(7):368-373. [doi: 10.1258/jtt.2010.091005] [Medline: 20679405]
66. Luczak SE, Hawkins AL, Dai Z, Wichmann R, Wang C, Rosen I. Obtaining continuous BrAC/BAC estimates in the field: a hybrid system integrating transdermal alcohol biosensor, Intellidrink smartphone app, and BrAC Estimator software tools. Addict Behav 2018 Aug;83:48-55 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.038] [Medline: 29233567]
67. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
68. Swendeman D, Comulada WS, Ramanathan N, Lazar M, Estrin D. Reliability and validity of daily self-monitoring by smartphone application for health-related quality-of-life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV. AIDS Behav 2015 Feb 21;19(2):330-340 [FREE Full text] [doi: 10.1007/s10461-014-0923-8] [Medline: 25331266]
69. Wray TB, Adia AC, Pérez AE, Simpanen EM, Woods L, Celio MA, et al. Timeline: a web application for assessing the timing and details of health behaviors. Am J Drug Alcohol Abuse 2019 May 14;45(2):141-150 [FREE Full text] [doi: 10.1080/00952990.2018.1469138] [Medline: 29757671]
70. Paolillo EW, Obermeit LC, Tang B, Depp CA, Vaida F, Moore DJ, et al. Smartphone-based ecological momentary assessment (EMA) of alcohol and cannabis use in older adults with and without HIV infection. Addict Behav 2018 Aug;83:102-108 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.016] [Medline: 29126667]
71. Monk RL, Heim D, Qureshi A, Price A. "I have no clue what I drunk last night" using smartphone technology to compare in-vivo and retrospective self-reports of alcohol consumption. PLoS One 2015 May 19;10(5):e0126209 [FREE Full text] [doi: 10.1371/journal.pone.0126209] [Medline: 25992573]
72. Dulin PL, Alvarado CE, Fitterling JM, Gonzalez VM. Comparisons of alcohol consumption by time-line follow back vs. Smartphone-based daily interviews. Addict Res Theory 2017 Nov 07;25(3):195-200 [FREE Full text] [doi: 10.1080/16066359.2016.1239081] [Medline: 29170622]
73. Kizakevich P, Eckhoff R, Brown J, Tueller SJ, Weimer B, Bell S, et al. PHIT for duty, a mobile application for stress reduction, sleep improvement, and alcohol moderation. Mil Med 2018 Mar 01;183(suppl_1):353-363. [doi: 10.1093/milmed/usx157] [Medline: 29635566]
74. Barrio P, Ortega L, López H, Gual A. Self-management and shared decision-making in alcohol dependence via a mobile app: a pilot study. Int J Behav Med 2017 Oct;24(5):722-727. [doi: 10.1007/s12529-017-9643-6] [Medline: 28236288]

75. Saunders J, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
76. Kim H, Awofeso O, Choi S, Jung Y, Bae E. Colorimetric analysis of saliva–alcohol test strips by smartphone-based instruments using machine-learning algorithms. Appl Opt 2016 Dec 23;56(1):84. [doi: 10.1364/ao.56.000084]
77. Matsumura K, Yamakoshi T, Ida T. Performance measures of alcohol-induced impairment: towards a practical ignition-interlock system for motor vehicles. Percept Mot Skills 2009 Dec 01;109(3):841-850. [doi: 10.2466/pms.109.3.841-850] [Medline: 20178284]
78. McAfee A, Watson J, Bianchi B, Aiello C, Agu E. AlcoWear: detecting blood alcohol levels from wearables. In: Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2017 Presented at: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); Aug. 4-8, 2017; San Francisco, CA, USA. [doi: 10.1109/uic-atc.2017.8397486]
79. Santani D, Do T, Labhart F, Landolt S, Kuntsche E, Gatica-Perez D. DrinkSense: characterizing youth drinking behavior using smartphones. IEEE Trans on Mobile Comput 2018 Oct 1;17(10):2279-2292. [doi: 10.1109/tmc.2018.2797901]
80. McClure EA, Tomko RL, Carpenter MJ, Treiber FA, Gray KM. Acceptability and compliance with a remote monitoring system to track smoking and abstinence among young smokers. Am J Drug Alcohol Abuse 2018 May 08;44(5):561-570 [FREE Full text] [doi: 10.1080/00952990.2018.1467431] [Medline: 29737885]
81. Meredith S, Robinson A, Erb P, Spieler CA, Klugman N, Dutta P, et al. A mobile-phone-based breath carbon monoxide meter to detect cigarette smoking. Nicotine Tob Res 2014 Jun;16(6):766-773 [FREE Full text] [doi: 10.1093/ntr/ntt275] [Medline: 24470633]
82. Wong HY, Subramaniyan M, Bullen C, Siddiq AN, Danaee M, Yee A. The mobile-phone-based iCO Smokerlyzer: comparison with the piCO Smokerlyzer among smokers undergoing methadone-maintained therapy. Tob Induc Dis 2019 Sep 9;17(September):65 [FREE Full text] [doi: 10.18332/tid/111355] [Medline: 31582954]
83. Herbec A, Parker E, Ubhi HK, Raupach T, West R. Decrease in resting heart rate measured using smartphone apps to verify abstinence from smoking: an exploratory study. Nicotine Tob Res 2020 Jul 16;22(8):1424-1427 [FREE Full text] [doi: 10.1093/ntr/ntaa021] [Medline: 31971595]
84. Dar R. Effect of real-time monitoring and notification of smoking episodes on smoking reduction: a pilot study of a novel smoking cessation app. Nicotine Tob Res 2018 Nov 15;20(12):1515-1518. [doi: 10.1093/ntr/ntx223] [Medline: 29126209]
85. Shoaib M, Scholten H, Havinga P, Incel O. A hierarchical lazy smoking detection algorithm using smartwatch sensors. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom); Sept. 14-16, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749439]
86. Béjar LM, Vázquez-Limón E. Is there any alternative to traditional food frequency questionnaire for evaluating habitual dietary intake? Nutr Hosp 2017 Jul 28;34(4) [FREE Full text] [doi: 10.20960/nh.650] [Medline: 29095013]
87. Harnack L. Nutrition Data System for Research (NDSR). In: Gellman TJ, editor. Encyclopedia of Behavioral Medicine. Cham: Springer; 2020.
88. AUSNUT 2011-13 - Australian food, supplement and nutritient database for estimation of population nutrient intakes 2014. Food Standards Australia New Zealand. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/foodnutrient/Pages/default.aspx [accessed 2022-01-17]
89. Poulton A, Pan J, Bruns LR, Sinnott RO, Hester R. Assessment of alcohol intake: retrospective measures versus a smartphone application. Addict Behav 2018 Aug;83:35-41. [doi: 10.1016/j.addbeh.2017.11.003] [Medline: 29128148]
90. Bernhardt JM, Usdan S, Mays D, Martin R, Cremeens J, Arriola KJ. Alcohol assessment among college students using wireless mobile technology. J Stud Alcohol Drugs 2009 Sep;70(5):771-775 [FREE Full text] [doi: 10.15288/jsad.2009.70.771] [Medline: 19737502]
91. Sobell LC, Sobell MB. Timeline follow-back. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption. Totowa, NJ: Humana Press; 1992:41-72.
92. Qin Y, Qian W, Shojaati N, Osgood N. Identifying smoking from smartphone sensor data and multivariate hidden Markov models. In: Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. 2017 Presented at: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation; July 5-8, 2017; Washington, USA. [doi: 10.1007/978-3-319-60240-0_27]
93. World Health Organization. Alcohol use disorders identification test. APA PsycTests - Database record. 1993. URL: https://doi.apa.org/doiLanding?doi=10.1037%2Ft01528-000 [accessed 2022-01-17]
94. 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 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

95. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e12578 [FREE Full text] [doi: 10.2196/12578] [Medline: 30950799]
96. Spring B, Moller AC, Coons MJ. Multiple health behaviours: overview and implications. J Public Health (Oxf) 2012 Mar 23;34(suppl 1):3-10 [FREE Full text] [doi: 10.1093/pubmed/fdr111] [Medline: 22363028]

## Abbreviations

**AROC:** area under the receiver operating characteristic curve
**AUDIT:** Alcohol Use Disorders Identification Test
**BAC:** blood alcohol concentration
**BrAC:** breath alcohol concentration
**CO:** carbon monoxide
**COSMIN:** Consensus-Based Standards for the Selection of Health Measurement Instruments
**eBrAC:** breath alcohol concentration estimates
**EMA:** ecological momentary assessment
**HMM:** hidden Markov model
**IPI:** interactive photo interface
**NLP:** natural language processing
**NuDAM:** Nutricam Dietary Assessment Method
**PRISMA:** Preferred Reporting Items for Systematic Review and Meta-Analyses
**SBI:** sketching-based interface
**TAC:** transdermal alcohol concentration
**TLFB:** Timeline Follow Back

*Edited by L Buis; submitted 21.01.21; peer-reviewed by P Markkanen, P Atorkey, K Bold, C Carrion; comments to author 22.03.21; revised version received 23.06.21; accepted 16.09.21; published 17.02.22.*

*Please cite as:*
*Thornton L, Osman B, Champion K, Green O, Wescott AB, Gardner LA, Stewart C, Visontay R, Whife J, Parmenter B, Birrell L, Bryant Z, Chapman C, Lubans D, Slade T, Torous J, Teesson M, Van de Ven P*
*Measurement Properties of Smartphone Approaches to Assess Diet, Alcohol Use, and Tobacco Use: Systematic Review*
*JMIR Mhealth Uhealth 2022;10(2):e27337*
*URL: https://mhealth.jmir.org/2022/2/e27337*
*doi:10.2196/27337*
*PMID:35175212*

Review

# The Use of Gamification and Incentives in Mobile Health Apps to Improve Medication Adherence: Scoping Review

Steven Tran[1], BPharm; Lorraine Smith[1], BA, PhD, Prof Dr; Sarira El-Den[1], BPharm, MIPH, PhD; Stephen Carter[1], BPharm, MSc, PhD

School of Pharmacy, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

**Corresponding Author:**
Steven Tran, BPharm
School of Pharmacy
Faculty of Medicine and Health
University of Sydney
Camperdown
Sydney, 2006
Australia
Phone: 61 93512222
Email: steventran@hotmail.com.au

## Abstract

**Background:** Emerging health care strategies addressing medication adherence include the use of direct-to-patient incentives or elements adapted from computer games. However, there is currently no published evidence synthesis on the use of gamification or financial incentives in mobile apps to improve medication adherence.

**Objective:** The aim of this scoping review is to synthesize and appraise the literature pertaining to the use of mobile apps containing gamification or financial incentives for medication adherence. There were two objectives: to explore the reported effectiveness of these features and to describe and appraise the design and development process, including patient involvement.

**Methods:** The following databases were searched for relevant articles published in English from database inception to September 24, 2020: Embase, MEDLINE, PsycINFO, CINAHL, and Web of Science. The framework by Arksey and O'Malley and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist guided this scoping review. Using a systematic screening process, studies were included if incentives or game features were used within mobile apps to specifically address medication adherence. An appraisal using risk of bias tools was also applied to their respective study design.

**Results:** A total of 11 studies from the initial 691 retrieved articles were included in this review. Across the studies, gamification alone (9/11, 82%) was used more than financial incentives (1/11, 9%) alone or a combination of the two (1/11, 9%). The studies generally reported improved or sustained optimal medication adherence outcomes; however, there was significant heterogeneity in the patient population, methodology such as outcome measures, and reporting of these studies. There was considerable variability in the development process and evaluation of the apps, with authors opting for either the waterfall or agile methodology. App development was often guided by a theory, but across the reviewed studies, there were no common theories used. Patient involvement was not commonly evident in predevelopment phases but were generally reserved for evaluations of feasibility, acceptance, and effectiveness. Patient perspectives on gamified app features indicated a potential to motivate positive health behaviors such as medication adherence along with critical themes of repetitiveness and irrelevance of certain features. The appraisal indicated a low risk of bias in most studies, although concerns were identified in potential confounding.

**Conclusions:** To effectively address medication adherence via gamified and incentivized mobile apps, an evidence-based co-design approach and agile methodology should be used. This review indicates some adoption of an agile approach in app development; however, patient involvement is lacking in earlier stages. Further research in a generalized cohort of patients living with chronic conditions would facilitate the identification of barriers, potential opportunities, and the justification for the use of gamification and financial incentives in mobile apps for medication adherence.

***(JMIR Mhealth Uhealth 2022;10(2):e30671)*** doi:10.2196/30671

**KEYWORDS**

gamification; incentives; mobile application; mHealth; medication adherence; mobile phone

## Introduction

### Background

Medication nonadherence, defined simply as failure to take medications as prescribed, is prevalent throughout all parts of the world [1]. It is estimated that the prevalence of nonadherence in high-income countries, such as Australia, is approximately 50% in patients living with chronic conditions [2]. This results in substantial economic and social costs to the patient and the country [3]. There are many interventions aimed at addressing nonadherence with some notable examples being reminders and increased health care professional contact points for dosing supervision and dosing administration aids, for example, Webster-pak (Webstercare) [4].

More recently, with the increasing penetration rate of smartphones and digital literacy globally, there has been rapid progress by the public and private health sectors to take advantage of mobile apps to address public health concerns [5]. The use of mobile health (mHealth) apps has predominately focused on physical activity and health tracking as companies such as *Fitbit* and *Niantic* can profit from commercializing wearables that integrate with the app or *in-app* currency [6,7]. In addition to generating substantial profits for the company, evaluations of these products demonstrate that their use leads to significant improvements in physical activity [8,9]. A key characteristic of mHealth apps such as *Pokémon Go* (Niantic, Inc) is the use of gamification [10].

Gamification is defined as the use of game elements in activities that are not commonly associated with games [11]. These game elements include but are not limited to colorful aesthetics, point systems, social competitions (ie, leaderboard), avatars, in-game rewards, and storyline quests [11]. Although rewards and incentives are a subset and element of gamification generally, they are limited to within the intervention and have no tangible or real-world economic value [12]. For this review, financial incentives are defined as a separate feature having a financial or tangible value that can be provided to users and used outside the system of the app, for example, accruing points in an app that can be redeemed for a shopping voucher at a physical store.

Approximately 8% of Australians delay or decide not obtaining a prescription because of cost [4]. Hence, cost not only presents a barrier to medication adherence but is also an opportunity for interventions in this area [4]. The concept of financial incentives tries to mitigate the cost associated with medications and reinforces positive behavior [13,14]. Multiple studies and trials from as early as 2008 suggest that financial reinforcement to medication adherence results in lower rates of treatment failure and higher rates of medication adherence across various patient populations [15-17]. However, this effect is dwarfed by concerns regarding the sustainability of funding for such interventions [18]. An intervention with a positive cost-benefit ratio may help justify funding from public health systems such as Australia's universal health insurance scheme *Medicare* or private health funders where spending upfront through financial rewards results in more savings through prevented medical expenses [19].

Understanding patients' perspectives may also provide some insight into the minimal standard of reward or frequency of prize required to balance intervention uptake and cost-effectiveness [20]. In addition, considering the *users* before development and implementation ensures that gamified interventions are designed to be compatible with the target audience, which ultimately determines the intervention's effectiveness [11].

A recent systematic review [21] on the general use of mobile apps for medication adherence reported that although empirical results indicate that mobile apps may improve medication adherence, it is ultimately unclear whether they are effective or what makes them effective because of the high degree of heterogeneity in study design and features included in the various apps identified in included studies. An analysis of the specific features such as gamification and incentives was not included in that review. Another review [22] noted that game elements and app features such as rewards can be used as tools to support basic psychological needs that align with the self-determination theory of Desi and Ryan [23] for behavior change in various health areas such as medication adherence. Although these features can be applied to a behavior change theory, the efficacy or application of these features has not been evaluated in medication adherence.

Results from gamified apps [6-8,10], such as the above-mentioned Pokémon Go, and financial incentive programs [15-17] in health areas justify exploring mobile app interventions that use gamification techniques: to encourage use and uptake, facilitate medical education on the benefits of medication treatment [24], and promote long-term positive behavior, that is, medication adherence, through financial rewards.

### Objectives

As there is no synthesized literature on the efficacy or use of gamification and incentives in mobile apps for medication adherence, the aim of this review is to explore the current use of gamification or financial incentives in mobile apps to address medication adherence and help identify best practices for future applications. Specifically, the objectives of this scoping review are as follows:

1. To explore the reported effectiveness of gamification or financial incentives in improving medication adherence
2. To identify, describe, and appraise the design and development processes (including patient involvement) used when developing mobile apps, which include gamification or financial incentives for the purpose of improving medication adherence

# Methods

## Overview

A scoping review maintains the ability to review this digital health care topic at a high level, identify gaps in the literature, and synthesize possible avenues for future research [25,26]. The framework proposed by Arksey and O'Malley [27] and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [25] guided this scoping review.

## Search Strategy

A search strategy was formulated by selecting only critical keywords in the objectives to retrieve a broad search (ie, *medication adherence*, *mobile apps*, and *[gamification* or *incentives]*). Each keyword was expanded with relevant synonyms and Medical Subject Headings (MeSH) terms relevant to each database. The full search term strategy for the Embase database is available in Multimedia Appendix 1.

The following databases were searched for relevant articles published in English from database inception to September 24, 2020 (search date): Embase, MEDLINE, PsycINFO, CINAHL, and Web of Science. The selection of the databases was decided by the coauthors and an academic librarian.

## Inclusion and Exclusion Criteria

To ensure that all potentially relevant articles were identified, the inclusion criteria include primary studies irrespective of study design. An article was included in the review if the study reported on the effectiveness of a mobile app for medication adherence containing financial incentives or game features (objective 1) or if the study discussed the use or development of a mobile app for medication adherence containing financial incentives or game features (objective 2).

Studies were excluded if health care professionals were the recipients or target audience of the financial incentives or gamified app instead of patients. Studies were also excluded if a full article was not accessible or could not be retrieved. Conference abstracts, nonprimary data sources, books, and book chapters were also excluded.

## Study Selection and Data Extraction

Articles identified through database searching were filtered for duplications using reference management software (EndNote). After duplicates were removed, the abstracts of articles were checked simultaneously with their titles for appropriateness to the research topic before full-text screening using the inclusion and exclusion criteria. Both title and abstract screening and full-text screening were conducted independently by 2 reviewers (ST and SC). A third independent reviewer (LS) was consulted to resolve disagreements regarding the eligibility of articles, when needed.

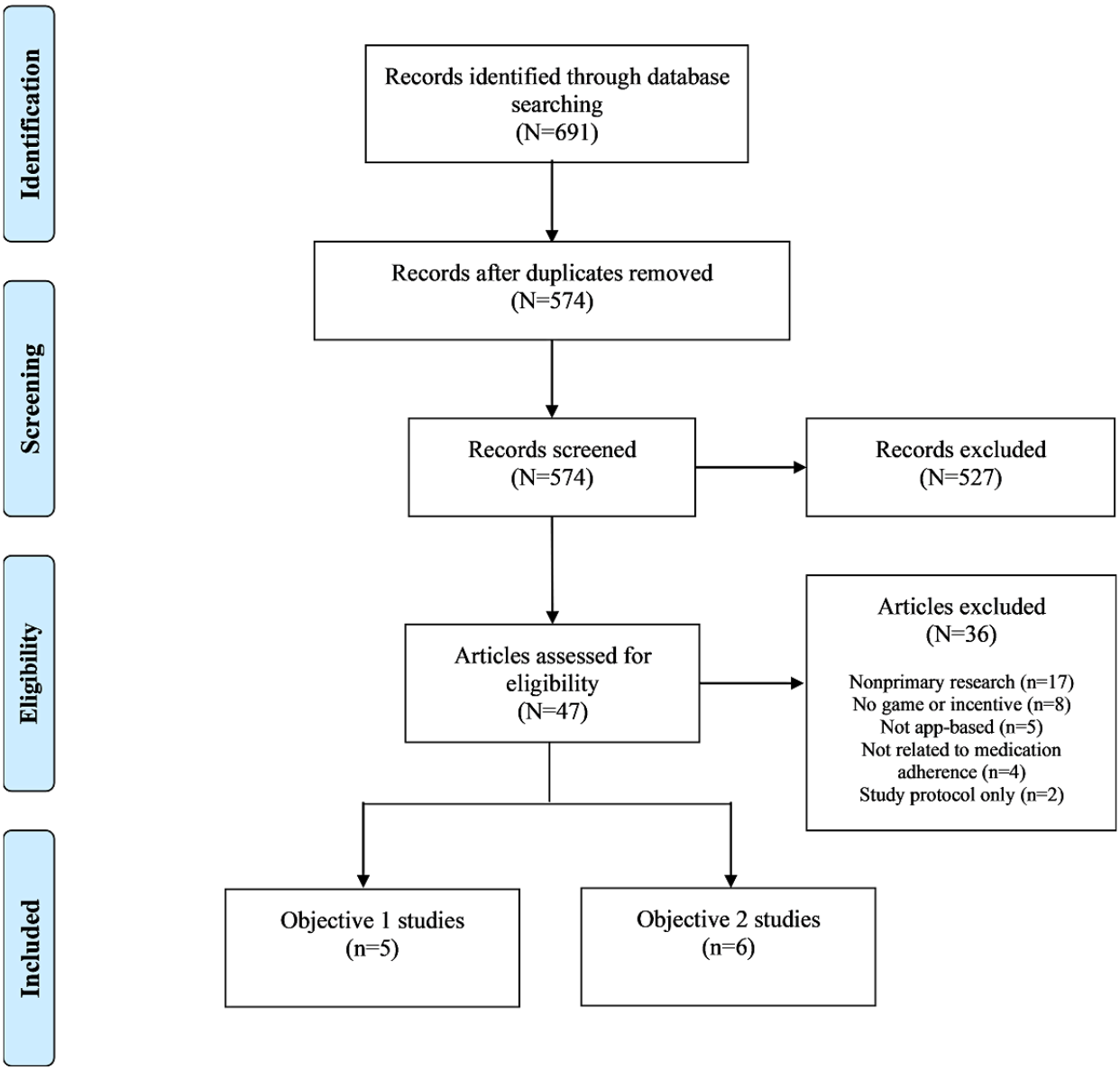
The included articles were reviewed by 3 researchers (ST, SC, and LS) during regular alignment meetings. The alignment meetings were used as a platform for data extraction to mitigate any discrepancy or bias in extraction and documentation. During the review process, the following attributes were recorded: the location of study, objective or aim of the study, short description of the study, patient population, sample size, and main results pertaining to the review objectives. Prespecified parameters were also recorded for analysis. These parameters included whether the study used gamification, financial incentives, or both; the underpinning theory or rationale to use gamification or financial incentives; and whether patient involvement or feedback was used in the development or testing of the app. In addition to the above-mentioned parameters, the studies were subject to an appraisal using 3 risk of bias tools depending on the study design, namely, the Cochrane Risk of Bias 2.0 Tool [28] for randomized trials, Risk of Bias in Nonrandomized Studies of Interventions [29] for nonrandomized studies of interventions, and the Joanna Briggs Institute checklist for qualitative research [30], where thematic analysis was reported.

# Results

## Total Reviewed Articles

A total of 691 articles were retrieved from the 5 databases. After duplicates were manually removed, 83.1% (574/691) of the articles underwent title and abstract screening. The title and abstract screening resulted in the exclusion of 91.8% (527/574) of the articles that were not relevant to the search. The remaining 8.2% (47/574) full-text articles were reviewed against the inclusion and exclusion criteria, of which 23% (11/47) articles were eligible for inclusion in the review. The PRISMA-ScR [25] flow diagram (Figure 1) provides further details on the screening process and the reasons for exclusion.

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




## Objective 1 Results

### *Article Characteristics*

Of the 11 studies, 5 (45%) fulfilled objective 1. Of the total 5 studies, 4 (80%) [31-34] were published after 2017 and 1 (20%) [35] was published in 2010. The studies were conducted in the following countries: Spain [35], South Korea [31], the United Kingdom and Scotland [32], Australia [34], and the United States [33].

Of the 5 studies, 3 (60%) were randomized studies [31-33], 1 (20%) was a cross-over study [35], and 1 (20%) was a retrospective observational study [34]. Moreover, of the 5 studies, 4 (80%) aimed to examine or evaluate the use of an app on medication adherence (among other outcomes) against the standard of care [31], against a negative control [33,35], or over time [34]. Finally, of the 5 studies, 1 (20%) [32] aimed to assess an app designed to promote disease and treatment management and reported medication adherence as a primary outcome.

Multimedia Appendix 2 [31-35] presents a summary of each of the 5 studies pertaining to objective 1.

### *Intervention Type, Period, and Theory*

Of the 5 studies, 2 (40%) [31,34] included an app that had more than 1 gamified or incentivized element. The most prevalent game elements used in the apps were point-based systems (ie, leaderboard [34,35], leveling up [31,33,34], quests [31,33], or in-game rewards [31]), which were used in 80% (4/5) of the studies. This was followed by gamified aesthetics or interface in 40% (2/5) of the studies [31,34] and the inclusion of mini games in 20% (1/5) of the studies [32]. Excluding payments for participating in the studies, financial incentive elements were only used in 1 app [34] in the form of a lottery or chance system to receive gift cards.

The target populations for the medication adherence apps were patients with Parkinson disease [32], youth with HIV [33], patients with cancer [31], the older adults [35], and a general group of patients living with chronic conditions [34]. Sample

sizes ranged from 18 [35] to 243 [34] participants, and the data set periods ranged from 3 weeks [31] to 6 months [34]. The largest sample size was from the retrospective chronic conditions study [34] (n=243 with 3 months of app use data available). The longest data set period was from the same study [34]; however, in another cohort (n=130 with 6 months of app use data available).

Across the 5 studies, only 3 (60%) mentioned an underpinning theory or framework for their intervention. The theories used to guide the development and evaluation of the interventions were the goal-setting theory and transtheoretical model [35], self-determination theory [34], social learning theory, and information-motivation-behavioral skills model of behavior change [33].

### *Effectiveness of Intervention*

Of the 5 studies, 2 (40%) [33,35] measured medication adherence with an independent pill box, whereas another 2 (40%) used self-reporting rating scales, namely, the Korean–Medication Adherence Rating Scale [31] and the Morsiky Medication Adherence Scale-8 [32]. The retrospective study [34] measured medication adherence through mobile direct observation of therapy in the app (in the form of taking a photo of the prescribed medication on the participant's hand or table). The retrospective study [34] aimed to analyze the impact of the app on medication adherence over time and excluded participants if the app was used for less than 30% of the analysis period.

Of the 5 studies, 3 (60%) [31,32,35] each showed statistically significant improvement in medication adherence in their respective intervention groups using the apps compared with that in the control or comparator groups. The study by Whiteley et al [33] reported no significant improvement in their *BattleViro* app compared with the control, but a significant improvement in adherence was observed in a patient subgroup analysis consisting of patients who had the newly initiated (within the past 3 months) antiretroviral therapy.

Overall, the studies represent varying degrees of evidence in support of the use of gamified interventions and a rationale for exploring further the potential of financially incentivized apps in improving medication adherence.

### *Patient Involvement*

Of the 5 studies, 4 (80%) [31-33,35] mentioned that their app was designed for their target population. Moreover, of the 5 studies, only 1 (20%) [33] specified the involvement of the target patient population in the development of the intervention. Finally, of the 5 studies, 1 (20%) [35] indicated that clinicians were involved in the design phase, and another (1/5, 20%) study [32] stated that the evaluation study also collected feedback on the app design from the users for future use.

### *Appraisal of Studies Pertaining to Objective 1*

A summary of the risk of bias appraisal for the studies pertaining to objective 1 is shown in Multimedia Appendix 3 [31-35]. Of the 5 studies in objective 1, 3 (60%) studies [31,32,35] were assessed as having a low risk of bias with no notable comments. The study by Whiteley et al [33] was assessed to have concerns relating to bias due to the selection of the reported results, specifically in the abstract. The study reported significant effectiveness of the intervention in a subgroup population despite finding nonsignificant changes in the total cohort of patients living with HIV and in the same subgroup using another outcome measure for medication adherence (ie, self-reported). It is important to consider that the aim of the study was to examine the preliminary effects of an app on several outcomes. The above-mentioned findings were discussed further by the authors as opportunities for furthering their research, and they noted that the study was limited by the small sample size and use of self-reporting to measure medication adherence, which is generally overreported.

Another study by Wiecek et al [34] was assessed to have serious risk in relation to possible confounding, selection of participants into the study, and possible bias due to missing data. These factors were identified by the authors as limitations in their study. In this retrospective observational study, baseline adherence and demographic data were not provided for all patients, and thus, the ability to control for confounding between the cohorts was not possible. In addition, the classification of the participants in the study was dependent on the duration of the study follow-up, which may have a direct link to the outcome measure. In addition, it was unclear if all recruited patients were included in the study; however, the exclusion criteria indicated that there were patients who were removed from the analysis to reach the study objective of assessing the impact of the intervention on medication adherence over time and not assessing adherence to the intervention over time. This is a serious concern, as the medication adherence outcome was measured via the intervention.

## Objective 2 Results

### *Article Characteristics*

Of the 11 studies, 6 (55%) studies did not address objective 1 but were included in the review as they pertained to objective 2 of this scoping review. All studies used either descriptive or qualitative methods; however, of the 6 studies, 4 (67%) [36-39] resembled a preliminary or pilot study. The earliest study [40] was published in 2013, whereas the other 83% (5/6) of the studies [36-39,41] were published after 2016. Of these 6 studies, 3 (50%) studies [38-40] were conducted in the United States, followed by 1 (17%) study in each of Spain [36], China [37], and Switzerland [41].

All studies included patients living with either cardiovascular disease [36-38,41] or HIV [36,39,40]. Of the 6 studies, 1 (17%) [40] also included young mothers in addition to patients with HIV. The study method varied, with 33% (2/6) of the studies using focus groups [40,41], 33% (2/6) using surveys [36,38], 17% (1/6) using individual interviews [39], and another (1/6, 17%) using a combination of focus groups and questionnaires [37]. The studies used a range of analysis techniques, including content analysis [36,41], clustering analysis [37], and thematic analysis [38-40]. Multimedia Appendix 4 [36-41] presents a summary of each of the 6 studies pertaining to objective 2.

### *Design and Development*

All 6 included studies underwent a design phase for their medication adherence or management app. Of the 6 studies, 5 (83%) studies [36-39,41] proceeded to develop their designed app, and 4 (67%) studies [36-39] further implemented their app among their target population for usability and feedback. In addition, of the 6 studies, 1 (17%) study [38] evaluated patients' perceived usefulness of the app for health-related measures such as medication management.

All authors [36-41] adopted a user-centered design for their app and focused on gathering information from their target audience or from a source relevant to their target audience. Across the included 6 studies, the authors explored the available literature or referred to external companies and existing apps to identify app features before validating them with a sample that represented their desired target audience through various methods. This indicates that the authors placed a high level of importance on the design of their app as opposed to other stages of app development. Among the 5 studies [36-39,41] that progressed from designing to developing an app, 3 (60%) studies [36,38,41] released only 1 build of the app after a linear development process (waterfall method), whereas 2 (40%) studies [37,39] decided to stagger the features in multiple separate builds (known as version or minimum viable product) and assess user uptake after each release.

### *Intervention Type and Theory*

Gamified elements and features were used in 83% (5/6) of the studies [36-39,41], whereas financial incentives were only mentioned but not used in 33% (2/6) of the studies [38,40]. Of the 6 studies, 1 (17%) [41] used the health access process approach model and required patients to match game elements, such as quests and a storyboard, to the model. Similarly, another study [37] used goal-directed design to identify game elements such as social leaderboards and in-game rewards. However, the feedback provided by participants following implementation of the leaderboard feature was that although it was easy to understand and use, it was too simple and users lost interest after a while. An existing game app was used in 17% (1/6) of the studies [39] as the basis for the mHealth game app by inserting health information and tailoring certain features as per feedback from patient interviews. Although 90% (10/11) of the participants were satisfied with the activities in the gamified app, 45% (5/11) of the participants did not find the game topics to be relevant to their lives, indicating a gap between what is fun or satisfying and what is useful or educational. Casino-style slot game features were used for an app in older patients following advice from nurses; however, older users testing the app expressed a desire to earn real money [38]. Similarly, in another study [34], patients expressed that they were more receptive to tasks or surveys in apps and the sensitivity of data privacy if there were financial incentives. However, there was no mention of what form of financial incentives would be preferred or what amount would be enough to entice user participation.

Owing to the variability of game features and lack of financial incentives used in the interventions, there is a lack of consensus as to the specific features that are suitable or desirable for a medication adherence health app.

### *Patient Perspectives and Voices*

Patient feedback and perspectives were either used for the requirements analysis or during feasibility and acceptance testing. Gamification or incentives were not the primary focus of the patient discussions in more than half of the included studies. Of the 6 studies, 2 (33%) [38,39] conducted a thematic analysis focusing on gamified apps, and 1 (17%) [40] mentioned financial incentives as an emerging theme. The latter [40] did not proceed into app development, and thus, the findings and patient preferences were not applied.

Of the 6 studies, 2 (33%) [39,41] gathered game features that were generally desired by their respective patient population and implemented them in their intervention. In contrast, in 33% (2/6) of the studies [36,37], the developers chose to implement a game feature without taking into account patient feedback or preferences.

Owing to time constraints, the study by Radhakrishnan et al [38] used nurses instead of older patients, the target patient population, to capture patient preference, including preferred game elements. Although the authors of this study [38] did not use the target patient population during development, they did ask older patients whether they thought the gamified app was or would be useful for medication adherence after testing. Approximately 80% (16/19) of the participants felt that the game motivated the user to adopt healthy behavior, such as salt restriction and medication management. Similarly, 80% (21/26) of the participants found gameplay and the content or information satisfying as it was *easy to play* and *informative*. Additional critical themes identified through the patients' responses were that the app was repetitive, lacked content, or did not interest users.

### *Appraisal of Studies Pertaining to Objective 2*

Table 1 provides the assessment of 3 studies where thematic analysis was reported using the Joanna Briggs Institute checklist for qualitative research. Ramanathan et al [40] represented a high-quality qualitative study focusing specifically on the thematic analysis of patient preferences for mHealth apps. Similarly, Whiteley et al [39] adequately represented the patients' voices; however, it is not mentioned where the researchers stand culturally or theoretically and if the researchers had any influence on the results. Radhakrishnan et al [38] also did not address the researchers' influence on the result or have congruity between the research methodology and research objective. The authors did not mention their intent to thematically analyze the comments from the patients but reported on a range of positive and critical themes. Ultimately, the authors identified that the results and themes were exploratory and require further investigation.

**Table 1.** JBI[a] checklist for qualitative research.

| JBI checklist for qualitative research | Radhakrishnan et al [38] | Ramanathan et al [40] | Whiteley et al [39] |
|---|---|---|---|
| 1. Is there congruity between the stated philosophical perspective and the research methodology? | Unclear | Yes | Yes |
| 2. Is there congruity between the research methodology and the research question or objectives? | No | Yes | Yes |
| 3. Is there congruity between the research methodology and the methods used to collect data? | Yes | Yes | Yes |
| 4. Is there congruity between the research methodology and the representation and analysis of data? | Yes | Yes | Yes |
| 5. Is there congruity between the research methodology and the interpretation of results? | Yes | Yes | Yes |
| 6. Is there a statement locating the researcher culturally or theoretically? | Yes | Yes | Unclear |
| 7. Is the influence of the researcher on the research, and vice versa, addressed? | No | Yes | Unclear |
| 8. Are participants, and their voices, adequately represented? | Yes | Yes | Yes |
| 9. Is the research ethical according to current criteria or, for recent studies, and is there evidence of ethical approval by an appropriate body? | Unclear | Yes | Yes |
| 10. Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data? | Yes | Yes | Yes |
| Overall and comments | Include | Include | Include |

[a]JBI: Joanna Briggs Institute.

# Discussion

## Principal Findings

This review explored the use of gamification or financial incentives in mobile apps to improve medication adherence. The findings indicate that gamification has been more widely studied than financial incentives in mobile apps for medication adherence. This review identified 1 study [34] that reported the use of gamification and financial incentives concomitantly to improve medication adherence. Although the study reported sustained optimal medication adherence for 6 months, it is unclear if the results were attributed to a single feature or the synergistic effects of the multiple components. There was an expectation that this review would identify more than 1 article that used both types of features based on the available articles relating to incentive programs or gamified interventions for other health outcomes such as physical activity. There was a wide variety of gamified features used as these were often specific and tailored to the studied patient population. The most prevalent type of gamified features observed across the reviewed studies were points-based features such as leaderboards and character leveling; however, it is unclear if such features are desirable to the general patient population as there was no analysis in a generalized population. It may be worth exploring preferred gamified and incentivized features for medication adherence in a generalized population taking medications for chronic conditions as this would increase the scope and reach of the app. A recent systematic review [42] supports this with their finding of a strong correlation between habit strength and medication adherence irrespective of patients' medical condition indicating that a habit-based intervention such as a financial incentive program [43] has the potential to increase medication adherence covering a wide audience. Generalized content can also be supplemented with condition-specific or population-specific content for those at higher risk of medication nonadherence such as people living with mental illness or HIV and AIDS [44].

In one of the included studies [33] for a gamified app, the authors did not observe a significant medication adherence improvement in the intervention group compared with control. This may be because patients who have lived with the condition (HIV) may not find gamified or educational apps as helpful or insightful compared with newly initiated or diagnosed patients owing to different challenges to medication adherence and the perception of an intentionally nonstigmatizing game as superficial [33]. This gap may be bridged with the use of financial incentives, as patients with HIV expressed that they were more inclined to record and partake in adherent behaviors with financial incentives provided to them, further supporting the concomitant use of gamification and financial incentives [40].

Owing to the limited published data, the effectiveness of financial incentives alone in mobile apps to improve medication adherence is unclear. Gamification alone may be effective for medication adherence; however, concerns arise from the heterogeneity in the intervention features, patient population, duration of the intervention, and outcome measures. In addition, it is unclear if any monetary or financial payments made to the patients for their participation in the study had any impact on the study outcomes. The use of a gamified intervention with financial incentives may eliminate the need for a study participation payment and would also represent the true effect of the intervention.

The retrospective study [34] that incorporated both games and financial incentives indicated sustained optimal medication

adherence over 3 and 6 months. However, the clinical question remains as to whether this effect is sustained beyond the 6-month follow-up period and whether this result is inflated because of the exclusion of participants who ceased using the app given that the medication adherence outcome was measured via the app.

Studies that use independent measures for medication adherence instead of self-reporting on the app represent the gold standard for measuring the true effect of the intervention by taking into account the patients' acceptance and use of the app [45]. In addition, more invasive methods of measuring such as direct observed therapy, pill counts, and electronic monitoring are more accurate compared with patient interviews and questionnaires [46]. Of the 5 studies pertaining to objective 1, 2 (40%) [31,34] reported on medication adherence measures opted for the more accurate but invasive independent pill count boxes.

There were various development methodologies undertaken by the included studies; however, they all followed three general stages: user-centered design (requirement analysis), intervention development, and testing. Where the differences can be seen are the theories and frameworks used, release phases, and the degree of patient involvement. Although each study used a different theory or framework, they were all able to achieve a functional app with satisfactory feedback from the participants. This supports the findings of a prior review [42] that indicates that the theoretical model or guiding framework may be of less importance when it comes to habit-based mHealth interventions. In addition, a recent review [47] found significant discrepancies within the conceptualization of gamification in several health behavior change theories, including the transtheoretical model and information-motivation-behavioral skills model, which were identified in 2 studies in this review. This indicates a poor understanding of the circumstances that allow gamification to support health behavior change [47]. Despite this unknown, the use of a behavior change theory, regardless of which one is used, helps inform design by considering the most relevant game or reward feature to the chosen theory [22].

This scoping review also revealed that the majority (3/5, 60%) of the identified apps had only 1 iteration or build before feasibility and acceptance testing. Of the 5 studies, only 2 (40%) [37,39] followed a more rigorous app development process that involved multiple iterations by upgrading the app based on feedback, as well as evaluation after each new version release. This approach of releasing and testing an intervention at multiple stages of the app development stage represents one of the more efficient and effective methods allowing for superior resource management, stakeholder or patient engagement, and product quality compared with the conventional waterfall method and is commonly referred to as the agile methodology [48].

Patient involvement was present in the user-centered design analysis and testing phases but was rarely seen in the app development stage. In the studies that did not include patients in the app development stage, agile methodology [48] was also not used. The benefit of having patients involved, particularly throughout the app development stage, ensures that the desired features are implemented as intended and that additional features that were initially missed in the design analysis can be incorporated more rapidly. In the included studies, there was little to no consideration for patients' perspectives and preferences regarding the use of gamification or incentives before app development, as often these features were selected by the developers or researchers or feature requirements were obtained from sources other than the intended target audience. By not consulting the desired audience directly, the potential for misinterpretation or bias in the selection of the gamified or incentive features is introduced [49]. Thus, there is a need to conduct high-quality qualitative studies such as that conducted by Ramanathan et al [40] but exploring gamification and incentives in mobile apps for medication adherence in patients with chronic conditions. This would provide the foundations for development by identifying perspectives of and receptiveness to gamification and incentives, the desirable features, cost-effective incentive prizes, barriers, and limitations and facilitate co-design.

## Limitations

This scoping review was guided by the PRISMA-ScR [25] methodology; however, limitations were identified during the systematic process. The high volume of articles obtained from the broad search meant that an abstract review was appropriate before a full-paper screening. However, this introduces the risk of accidentally excluding studies that are relevant. To minimize this, 2 independent reviewers (ST and SC) identified relevant studies, and a third reviewer (LS) adjudicated any discrepancies. In addition, because of the ever-changing digital landscape, these findings are bound by the search period and should be interpreted with caution as newer articles become available. These findings should serve as a summary snapshot of the historical data in this field.

Another limitation is that gray material or unpublished studies were not included in this review. The implementation of our broad search strategy in the gray literature would retrieve search results in the 100,000 range and thus was not included in the scope of this review because of pragmatic reasons. Many health apps are privately operated, and information pertaining to their development and evaluation is often not published in peer-reviewed journals or the public domain. This may lead to a substantial knowledge gap that cannot be mitigated because of the potential classification of the information as proprietary data. However, there is a trend for private companies to voluntarily publish these data to promote their intervention for transparency, marketing, or funding reasons [50].

## Conclusions

This scoping review highlights that gamification is more prevalent than financial incentives in mobile apps for medication adherence. The concurrent use of gamification and financial incentives is rare. Gamification alone may be effective for medication adherence; however, there are many knowledge gaps and inconsistencies in evidence, data generation, and development. In addition, the variability of features across identified apps indicates the lack of consensus as to which features are most desirable or effective. Features that are preferred by a generalized cohort of patients with chronic conditions should be explored in future research before further

personalization can be applied for specific patient populations. In addition, the development stages would benefit greatly from more patient involvement and contribution. This can be facilitated by applying a co-design and agile methodology.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search term strategy for the Embase database.
[PDF File (Adobe PDF File), 56 KB - mhealth_v10i2e30671_app1.pdf ]

Multimedia Appendix 2
Characteristics and findings of the studies pertaining to Objective 1.
[PDF File (Adobe PDF File), 166 KB - mhealth_v10i2e30671_app2.pdf ]

Multimedia Appendix 3
Summary of the risk of bias appraisal for the studies pertaining to Objective 1.
[PDF File (Adobe PDF File), 102 KB - mhealth_v10i2e30671_app3.pdf ]

Multimedia Appendix 4
Characteristics and findings of the studies pertaining to Objective 2.
[PDF File (Adobe PDF File), 138 KB - mhealth_v10i2e30671_app4.pdf ]

## References


1. Hugtenburg JG, Timmers L, Elders PJ, Vervloet M, van DL. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. Patient Prefer Adher 2013;7:675-682 [FREE Full text] [doi: 10.2147/PPA.S29549] [Medline: 23874088]
2. Usherwood T. Encouraging adherence to long-term medication. Aust Prescr 2017 Aug;40(4):147-150 [FREE Full text] [doi: 10.18773/austprescr.2017.050] [Medline: 28947853]
3. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018 Jan 21;8(1):e016982 [FREE Full text] [doi: 10.1136/bmjopen-2017-016982] [Medline: 29358417]
4. Underwood B, Williams N, Kinnard Jr HW. APhA2016 abstracts of contributed papers: medication adherence rates utilizing a community pharmacy-based mobile application. J Am Pharm Assoc 2016;56(3):1-141 [FREE Full text]
5. Martin W, Sarro F, Jia Y, Zhang Y, Harman M. A survey of app store analysis for software engineering. IIEEE Trans Software Eng 2017 Sep 1;43(9):817-847. [doi: 10.1109/tse.2016.2630689]
6. Mobile Consumer Survey 2019. Deloitte. 2019. URL: https://tinyurl.com/3purtufw [accessed 2020-05-29]
7. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on physical activity: study and implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
8. Khamzina M, Parab KV, An R, Bullard T, Grigsby-Toussaint DS. Impact of Pokémon Go on physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Feb;58(2):270-282. [doi: 10.1016/j.amepre.2019.09.005] [Medline: 31836333]
9. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable technology and physical activity behavior change in adults with chronic cardiometabolic disease: a systematic review and meta-analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
10. Militello LK, Hanna N, Nigg CR. Pokémon Go within the context of family health: retrospective study. JMIR Pediatr Parent 2018 Oct 03;1(2):e10679 [FREE Full text] [doi: 10.2196/10679] [Medline: 31518294]
11. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
12. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
13. Alipanah N, Jarlsberg L, Miller C, Linh NN, Falzon D, Jaramillo E, et al. Adherence interventions and outcomes of tuberculosis treatment: a systematic review and meta-analysis of trials and observational studies. PLoS Med 2018 Jul 3;15(7):e1002595 [FREE Full text] [doi: 10.1371/journal.pmed.1002595] [Medline: 29969463]
14. Willey C. Behavior-changing methods for improving adherence to medication. Curr Hypertens Rep 1999 Dec;1(6):477-481. [doi: 10.1007/s11906-996-0019-6] [Medline: 10981109]

15. Volpp KG, Loewenstein G, Troxel AB, Doshi J, Price M, Laskin M, et al. A test of financial incentives to improve warfarin adherence. BMC Health Serv Res 2008;8:272 [FREE Full text] [doi: 10.1186/1472-6963-8-272] [Medline: 19102784]
16. Noordraven EL, Wierdsma AI, Blanken P, Bloemendaal AF, Staring AB, Mulder CL. Financial incentives for improving adherence to maintenance treatment in patients with psychotic disorders (Money for Medication): a multicentre, open-label, randomised controlled trial. Lancet Psychiatry 2017 Mar;4(3):199-207. [doi: 10.1016/S2215-0366(17)30045-7] [Medline: 28236956]
17. Kurti AN, Davis DR, Redner R, Jarvis BP, Zvorsky I, Keith DR, et al. A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change. Transl Issues Psychol Sci 2016 Jun;2(2):128-152 [FREE Full text] [doi: 10.1037/tps0000067] [Medline: 27777964]
18. Noordraven EL, Wierdsma AI, Blanken P, Bloemendaal AF, Mulder CL. Medical and social costs after using financial incentives to improve medication adherence: results of a 1 year randomised controlled trial. BMC Res Notes 2018 Sep 10;11(1):655 [FREE Full text] [doi: 10.1186/s13104-018-3747-1] [Medline: 30201054]
19. Prakash AM, He Q, Zhong X. Incentive-driven post-discharge compliance management for chronic disease patients in healthcare service operations. IISE Transactions on Healthcare Systems Engineering 2019 Apr 19;9(1):71-82. [doi: 10.1080/24725579.2019.1567630]
20. Bosworth HB, Ngouyombo B, Liska J, Zullig LL, Atlani C, Beal AC. The importance of cholesterol medication adherence: the need for behavioral change intervention programs. Patient Prefer Adher 2018;12:341-348 [FREE Full text] [doi: 10.2147/PPA.S153766] [Medline: 29563777]
21. Ng R, Carter SR, El-Den S. The impact of mobile applications on medication adherence: a systematic review. Transl Behav Med 2019 Aug 06. [doi: 10.1093/tbm/ibz125] [Medline: 31384950]
22. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum-Comput Stud 2020 Aug;140:102449 [FREE Full text] [doi: 10.1016/j.ijhcs.2020.102449]
23. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
24. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
29. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
30. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
31. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]
32. Lakshminarayana R, Wang D, Burn D, Chaudhuri KR, Galtrey C, Guzman NV, et al. Erratum: Author correction: using a smartphone-based self-management platform to support medication adherence and clinical consultation in Parkinson's disease. NPJ Parkinsons Dis 2017 Nov 13;3(1):32 [FREE Full text] [doi: 10.1038/s41531-017-0034-0] [Medline: 29152558]
33. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018;30(sup4):21-33 [FREE Full text] [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]
34. Wiecek E, Torres-Robles A, Cutler RL, Benrimoj SI, Garcia-Cardenas V. Impact of a multicomponent digital therapeutic mobile app on medication adherence in patients with chronic conditions: retrospective analysis. J Med Internet Res 2020 Aug 12;22(8):e17834 [FREE Full text] [doi: 10.2196/17834] [Medline: 32784183]
35. de Oliveira R, Cherubini M, Oliver N. MoviPill: improving medication compliance for elders using a mobile persuasive social game. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing. New York: Assoc Computing Machinery; 2010 Presented at: Ubicomp '10: The 2010 ACM Conference on Ubiquitous Computing; September 26 - 29, 2010; Copenhagen Denmark p. 251-260. [doi: 10.1145/1864349.1864371]

36. Anglada-Martínez H, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Sequeira E, Aragunde V, et al. An interactive mobile phone-website platform to facilitate real-time management of medication in chronically ill patients. J Med Syst 2017 Aug 8;41(8):122. [doi: 10.1007/s10916-017-0767-7] [Medline: 28689310]
37. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466. [doi: 10.2196/14466]
38. Radhakrishnan K, Toprac P, O'Hair M, Bias R, Kim MT, Bradley P, et al. Interactive digital e-health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374 [FREE Full text] [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
39. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A mobile gaming intervention to increase adherence to antiretroviral treatment for youth living with HIV: development guided by the information, motivation, and behavioral skills model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e96 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
40. Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform 2013 Apr;82(4):38-46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.05.009] [Medline: 22704234]
41. Ehrler F, Gschwind L, Meyer P, Christian L, Blondon K. SMART-MEDS: Development of a medication adherence app for acute coronary syndrome patients based on a gamified behaviour change model. AMIA Annu Symp Proc 2018;2018:413-421 [FREE Full text] [Medline: 30815081]
42. Badawy SM, Shah R, Beg U, Heneghan MB. Habit strength, medication adherence, and habit-based mobile health interventions across chronic medical conditions: systematic review. J Med Internet Res 2020 Apr 28;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]
43. Gneezy U, Meier S, Rey-Biel P. When and why incentives (Don't) work to modify behavior. J Econ Perspect 2011 Nov 01;25(4):191-210. [doi: 10.1257/jep.25.4.191]
44. Addo B, Sencherey S, Babayara MN. Medication noncompliance among patients with chronic diseases attending a primary health facility in a periurban district in Ghana. Int J Chronic Dis 2018;2018:7187284 [FREE Full text] [doi: 10.1155/2018/7187284] [Medline: 29977904]
45. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
46. Vrijens B, Antoniou S, Burnier M, de la Sierra A, Volpe M. Current situation of medication adherence in hypertension. Front Pharmacol 2017;8:100 [FREE Full text] [doi: 10.3389/fphar.2017.00100] [Medline: 28298894]
47. Schmidt-Kraepelin M, Warsinsky S, Thiebes S, Sunyaev A. The role of gamification in health behavior change: a review of theory-driven studies. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020), Maui, Hawaii; Grand Wailea, Maui, Hawaii2020. 2020 Presented at: 53rd Hawaii International Conference on System Sciences (HICSS 2020), Maui, Hawaii; Grand Wailea, Maui, Hawaii2020; 2020; Grand Wailea, Maui, Hawaii. [doi: 10.24251/hicss.2020.155]
48. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018 Sep 13;1(1):46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
49. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016 Jul 4;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
50. Joshi M, Bhardwaj P. Impact of data transparency: scientific publications. Perspect Clin Res 2018;9(1):31-36 [FREE Full text] [doi: 10.4103/picr.PICR_104_17] [Medline: 29430415]


## Abbreviations

**MeSH:** Medical Subject Headings
**mHealth:** mobile health
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

*Edited by L Buis; submitted 24.05.21; peer-reviewed by I Shubina, B Chaudhry, S Bhattacharjya; comments to author 17.08.21; revised version received 19.09.21; accepted 13.12.21; published 21.02.22.*

*Please cite as:*
*Tran S, Smith L, El-Den S, Carter S*
*The Use of Gamification and Incentives in Mobile Health Apps to Improve Medication Adherence: Scoping Review*
*JMIR Mhealth Uhealth 2022;10(2):e30671*
*URL: https://mhealth.jmir.org/2022/2/e30671*
*doi:10.2196/30671*
*PMID:35188475*

Review

# Personalization of Intervention Timing for Physical Activity: Scoping Review

Saurabh Chaudhari[1], BE; Suparna Ghanvatkar[1], MTech; Atreyi Kankanhalli[1], PhD

Department of Information Systems and Analytics, School of Computing, National University of Singapore, Singapore, Singapore

**Corresponding Author:**
Saurabh Chaudhari, BE
Department of Information Systems and Analytics
School of Computing
National University of Singapore
Computing 1, 13 Computing Drive
Singapore, 117417
Singapore
Phone: 65 86153503
Email: sach16795@gmail.com

## Abstract

**Background:** The use of sensors in smartphones, smartwatches, and wearable devices has facilitated the personalization of interventions to increase users' physical activity (PA). Recent research has focused on evaluating the effects of personalized interventions in improving PA among users. However, it is critical to deliver the intervention at an appropriate time to each user to increase the likelihood of adoption of the intervention. Earlier review studies have not focused on the personalization of intervention timing for increasing PA.

**Objective:** This review aims to examine studies of information technology–based PA interventions with personalized intervention timing (PIT); identify inputs (eg, user location) used by the system for generating the PIT, the techniques and methods used for generating the PIT, the content of the PA intervention, and delivery mode of the intervention; and identify gaps in existing literature and suggest future research directions.

**Methods:** A scoping review was undertaken using PsycINFO, PubMed, Scopus, and Web of Science databases based on a structured search query. The main inclusion criteria were as follows: the study aimed to promote PA, included some form of PIT, and used some form of information technology for delivery of the intervention to the user. If deemed relevant, articles were included in this review after removing duplicates and examining the title, abstract, and full text of the shortlisted articles.

**Results:** The literature search resulted in 18 eligible studies. In this review, 72% (13/18) of the studies focused on increasing PA as the primary objective, whereas it was the secondary focus in the remaining studies. The inputs used to generate the PIT were categorized as user preference, activity level, schedule, location, and predicted patterns. On the basis of the intervention technique, studies were classified as manual, semiautomated, or automated. Of these, the automated interventions were either knowledge based (based on rules or guidelines) or data driven. Of the 18 studies, only 6 (33%) evaluated the effectiveness of the intervention and reported positive outcomes.

**Conclusions:** This work reviewed studies on PIT for PA interventions and identified several aspects of the interventions, that is, inputs, techniques, contents, and delivery mode. The reviewed studies evaluated PIT in conjunction with other personalization approaches such as activity recommendation, with no study evaluating the effectiveness of PIT alone. On the basis of the findings, several important directions for future research are also highlighted in this review.

***(JMIR Mhealth Uhealth 2022;10(2):e31327)*** doi:10.2196/31327

**KEYWORDS**

review; physical activity; personalized intervention; intervention timing; mobile apps; fitness tracker; mobile phone

## Introduction

### Background

The increase in people's sedentary lifestyle is strongly correlated to the rise in chronic diseases [1]. The American Heart Association recommends at least 150-300 minutes of moderate-intensity aerobic physical activity (PA) a week to reduce the risk of heart disease and stroke [2]. However, in the United States, for example, an estimated 36.5% of the adults aged 18-44 years did not meet the recommended PA levels [3], leading to a call for approaches to increase PA levels.

In this regard, information technology (IT) advances have allowed for development of, and widespread access to, fitness apps and trackers. Here, IT refers to technologies used for the collection, communication, retrieval, storage, presentation, and processing of information in all its forms [4]. The availability of sensors in smartphones, smartwatches, and wearable devices allows PA monitoring of individuals in an increasingly accurate [5] and cost-effective manner [6]. In addition, fitness trackers such as Fitbit provide real-time personalized insights [7] regarding PA to users through fitness apps. However, despite the availability of PA insights and PA guideline levels, the lack of clear and actionable feedback or a recommendation tailored to the user often results in failure to achieve the recommended PA levels [8,9]. For example, an intervention with a goal recommendation for achieving weekly 150 minutes of moderate to vigorous PA (MVPA) does not provide users with actionable recommendations on achieving the goals.

The availability of fitness trackers coupled with the increase in the number of fitness apps has provided novel research opportunities to design, investigate, and assess interventions—defined as the messages or elements through which the apps aim to improve health behaviors [10]. Thus, IT-based interventions refer to interventions using IT (defined earlier) for their delivery. Initial intervention studies aimed at increasing PA levels typically delivered the interventions through web portals and relied on self-reported data by the user [11,12]. A key barrier in such intervention studies is the irregularity in the user's reporting and bias regarding self-reported data. The use of fitness trackers to monitor PA levels of users has allowed for in-depth analysis of PA at a more personal level [13] and can improve the effectiveness of the PA interventions [14].

However, increasing PA often requires a change in the lifestyle or behavior of the user. Users are motivated by varied reasons such as health benefits, hedonic motivations, and social rewards [15,16] to increase their PA. In addition, temporal and environmental factors such as time of day, day of the week, and weather often influence the user's decision to exercise or not [17]. Therefore, the *one size fits all* approach does not serve the diversity of users, thus creating a need for personalized PA interventions to promote adherence to the app and PA [18].

Several studies evaluating the effectiveness of personalized interventions have reported an increase in the PA levels of users [19-21]. However, the effectiveness of the interventions is adversely affected by users' poor adherence to the app and the PA guidelines [22], leading to short-lived lifestyle changes. A relevant recommendation delivered at an irrelevant time is one of the reasons reported as a cause for this behavior [23]. The timing of the intervention can be irrelevant to the user because of differences in schedules, preferences, or other factors influencing individuals' choices [24]. For instance, an intervention message, although personalized, is likely to be ignored if it is delivered when the user is otherwise busy in other activities. In addition, the appropriate time to deliver an intervention might depend on the type of intervention. For instance, some interventions, such as those for PA goal planning, might need the user to self-reflect instead of performing the PA itself. Thus, it becomes essential for interventions to be delivered at the time when the user can engage in the target activity of the intervention.

In this review, we adapt the definition of personalized intervention timing (PIT) for PA from the study by Ghanvatkar et al [25] to define it as a personalization that takes the user and context into account and determines the appropriate time to deliver the intervention (eg, message) regarding PA or recommends the time to the user. Users in various studies mentioned that an intervention delivered considering the individual's schedule, circadian rhythm, and lifestyle could increase the likelihood of the individual adopting the intervention's recommendations [26], thereby increasing the adherence rate [27]. Although studies have reiterated the importance of PIT to increase the effectiveness of a PA intervention, the specifics of how to achieve PIT are still unclear, which requires further investigation.

In this regard, we found 2 previous reviews on personalized interventions for increasing PA, which focused on their classification or evaluated the effectiveness of the interventions. First, a recent study by Ghanvatkar et al [25] broadly classified personalized PA interventions into six categories, that is, goal recommendation, activity recommendation, fitness partner recommendation, educational content, motivational content, and intervention timing. Second, the study by Aldenaini et al [28] further assessed various implementation methods and evaluated the effectiveness of the personalized intervention categories defined by Ghanvatkar et al [25]. In addition, a review study by Tong et al [29] aimed more broadly at evaluating the effectiveness of a personalized mobile intervention in promoting lifestyle behavior change (ie, in PA, diet, smoking, and alcohol consumption). Finally, a review by op den Akker et al [30] focused more narrowly on studies with personalization for PA coaching systems before 2013. Few of the studies in their review examined personalization of intervention timing, and these were mainly about personalizing music or vibrations based on the user's gait or personalizing the mobile app display based on user preferences [30].

However, we did not find any review focused on PIT for PA or other health behaviors. Therefore, despite the importance of PIT in the effectiveness of interventions, it is unclear what the existing knowledge is regarding the design and effectiveness of PIT. Motivated by the literature gap and the role of PIT in the effectiveness of interventions, this review primarily focuses on providing an overview of PIT research for PA improvement and suggesting directions for future research that remain

unexplored. The results from this review expand our current knowledge and help obtain insights that can lead to more effective personalized PA interventions with PIT.

Thus, this review examines the studies that provided PIT to increase PA and identifies and categorizes types of inputs, intervention techniques, intervention content, and mode of delivery for the interventions. An intervention with PIT, like other systems, can be viewed through an input-process-output model [31]. The components of the system according to this model are (1) inputs, defined as the requirements from the environment; (2) process, defined as the computation based on the inputs; and (3) outputs, which refers to the results or outcomes provided by the system. We adopt this model to define the system components that produce IT-based PA interventions for PIT. The inputs to the system include user and contextual characteristics to design the intervention. The personalization process uses a method or technique to create the intervention. The output of the system is the intervention with PIT received by the user, based on the processing. In this review, we identify the content of the output or intervention as well as its mode of delivery to the user (eg, email, SMS text message, and mobile app notification). Furthermore, theories used to design the intervention and the results of the intervention studies are explored. Finally, we identify the research gaps in existing literature and outline directions for future research.

### Objectives

This review aims to (1) examine recent studies of IT-based PA interventions with PIT; (2) identify inputs used by the system for creating the personalized intervention, techniques used to process the inputs and create the intervention, content of the intervention, delivery mode of the intervention, theories used in the intervention design, and effectiveness of providing PIT; and (3) identify gaps in existing literature and suggest future research directions.

## Methods

### The Scoping Review

This scoping review aims to identify and summarize prior studies that examined IT-based interventions with PIT to increase PA levels, as per the aims of scoping reviews [32]. To ensure the quality of the included studies, we only selected peer-reviewed articles, including research-in-progress articles, for which the full text was available. This review follows the scoping review methodology [33] of identifying the research objective (previous section); identifying relevant studies (search strategy); study selection; charting or extracting the data; and collating, summarizing, and reporting the results.

### Search Strategy

This review included relevant articles from the PsycINFO, PubMed, Scopus, and Web of Science databases published from January 1, 2013, to March 30, 2021. These databases were chosen because they cover the relevant studies in the medical and health informatics domains. Fitness trackers and mobile apps have been widely adopted for PA promotion only in recent years; hence, older studies might not be relevant for our review. Furthermore, the studies for personalization of PA coaching systems before 2013 have been reviewed by op den Akker et al [30], including the few studies that personalized the timing of the intervention. In addition, prior review studies [25] of user models for personalizing PA interventions have also considered articles published since 2013. Thus, only articles published after 2013 were considered in this review.

The constructed search query for shortlisting studies from the databases was as follows: *((fitness OR exercise OR physical activity OR activity level OR active living) AND (intervention OR recommend* OR prescribe OR prescription OR feedback OR message) AND (tailor* OR personaliz* OR personalis*) AND (mobile OR internet OR computer OR device OR fitness trackers OR website OR online) AND (time* OR timing* OR temporal))*. This should ensure that all the studies that satisfied a semantic similarity to the following are shortlisted: {*physical activity*} {*interventions*} having {*personalization*} provided through some form of {*information technology*} and containing {*temporal*} analysis in some aspect. In addition, the publication must be available in English. Furthermore, this review included cross-referenced articles that were relevant and met the selection criteria.

### Selection Criteria

Studies were considered eligible if all the following inclusion criteria were met: (1) either the primary or secondary objective of the study was to increase PA among its users; (2) the study included some form of PIT; (3) the study used some form of IT for delivery of the intervention to the user; (4) the study article was available in English and published between January 1, 2013, and March 30, 2021; and (5) it was not a review article or dissertation and was published through a peer-reviewed process. The following exclusion criteria were used for our review: (1) personalization not aimed at increasing PA; (2) intervention delivered to the user without any timing personalization, that is, intervention delivered to all users at the same time; and (3) intervention delivered without using any IT and delivered in face-to-face sessions.

The inclusion criteria for this review did not impose any restrictions on the group or category of participants, technology platform, study design, or study setting. Consequently, the studies included in this review have varied groups of participants, use any type of IT to deliver the intervention, adopt varied study designs, and even include hybrid human–digital intervention studies. Furthermore, because our focus was on PIT interventions increasing PA as either the primary or secondary objective of the study, we did not include studies that provided PIT to reduce sedentary behavior (SB), unless the intervention was also designed to increase PA.

### Screening and Study Selection

The screening and study selection were undertaken by 1 researcher (SC) and subsequently verified independently by another researcher (SG) for adherence to the selection criteria. The second researcher (SG) was not blinded and had access to the first researcher's (SC) findings. Disagreements between the 2 researchers were resolved through discussion and consensus with the third researcher (AK).

The article selection process comprised 2 search and filtering phases. The first phase involved assessing the title, abstract, and keywords of the articles obtained from the databases to see whether they should be included based on the inclusion criteria. Mendeley Reference Manager was used to organize and merge duplicate articles from the various databases. The second phase involved a full-text review of the articles that satisfied the inclusion criteria and did not meet the exclusion criteria. In this scoping review, only articles deemed relevant after the second phase were included.

Initially, in conjunction with the date range and language filters, the search query yielded 1955 studies. In addition, 10 relevant studies were identified by hand searching and cross-references. Next, these 1965 studies were scanned for duplicates, which resulted in a total of 1154 (58.73%) unique studies. The abstracts of the 1154 unique studies were assessed for the inclusion and exclusion criteria, resulting in a shortlist of 281 (24.35%) articles. The further assessment of these 281 shortlisted articles for full text resulted in 28 (9.9%) articles included in this review.

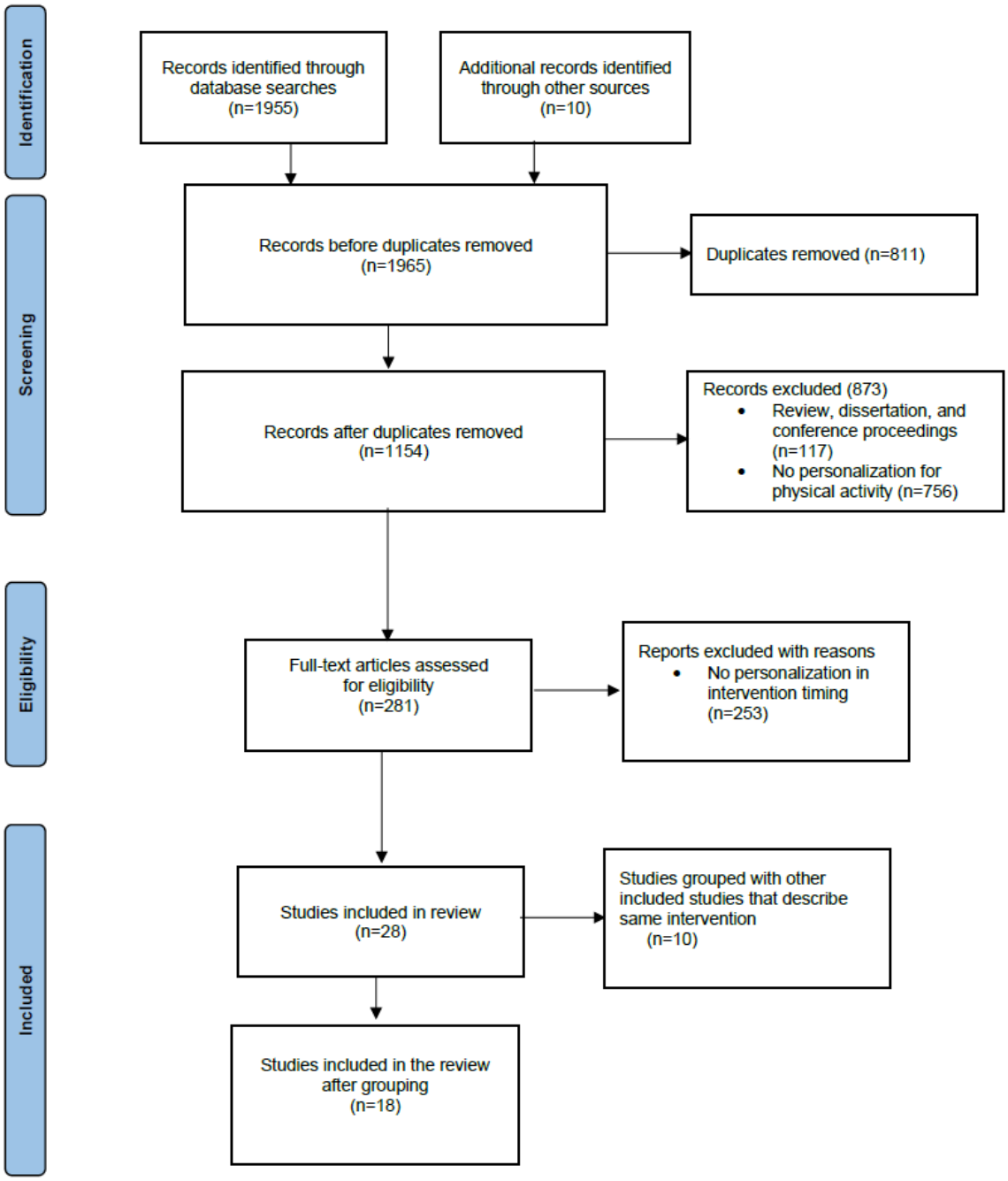
However, of the 28 articles included in the review, 6 (21%) presented different aspects of the same intervention or improvements of the same intervention system in multiple publications. All such related publications were grouped, and only a single publication with the most comprehensive intervention details was eventually selected to represent the studies using the same intervention system. This information about the articles referring to the same intervention and their representative publication selected for our review is presented in Table 1. After this grouping of related studies, out of 28 articles, 18 (64%) unique articles or intervention systems were included in this review. Figure 1 illustrates the flowchart representing the entire study selection procedure.

**Table 1.** Related studies regarding an intervention and the representative study chosen.

| Intervention | Related studies | Representative study |
|---|---|---|
| B-MOBILE JITAI[a] | Bond et al [34] and Thomas et al [35] | Thomas et al [35] |
| e-Moms Roc | Fernandez et al [36], Graham et al [37], and Olson et al [38] | Graham et al [37] |
| HeartSteps | Greenewald et al [39], Klasnja et al [40], Klasnja et al [41], and Liao et al [42] | Klasnja et al [41] |
| MINI Movers | Downing et al [43] and Downing et al [44] | Downing et al [44] |
| Text4Heart | Dale et al [45], Maddison et al [46], and Pfaeffli Dale et al [47] | Maddison et al [46] |
| txt4two | Willcox et al [48] and Willcox et al [49] | Willcox et al [49] |

[a]JITAI: just-in-time adaptive intervention.

**Figure 1.** Flowchart of the study selection process.




### Data Extraction

The data extraction or data charting from each article was performed following the approach in the study by Arksey et al [33]. We captured the following variables, which together form the basis of our analysis:

- Objective and research question
- Theory used (if any)
- Study method, which included the information regarding the study method, such as study design, duration, and setting of the study (daily living or laboratory based)
- Participant sample, which included but was not restricted to the participants' demographics (such as age and gender)
- Intervention, which included all the characteristics of the intervention system, such as mode of delivery, the content of the intervention, the technique used for providing PIT, user-specific inputs used by the system to provide PIT, and the method and devices (if used) used to extract user-specific inputs
- Results, which included the intervention evaluation results, if provided

## Results

### Overview of Studies

We placed no restrictions on the intervention's research objective or methodology to be included in this review other than following our selection criteria. As a result, the studies differ concerning their research objectives, data collection methods, target users, and intervention. We summarize each of

these aspects of the interventions before reviewing the PIT provided by the studies included in this review.

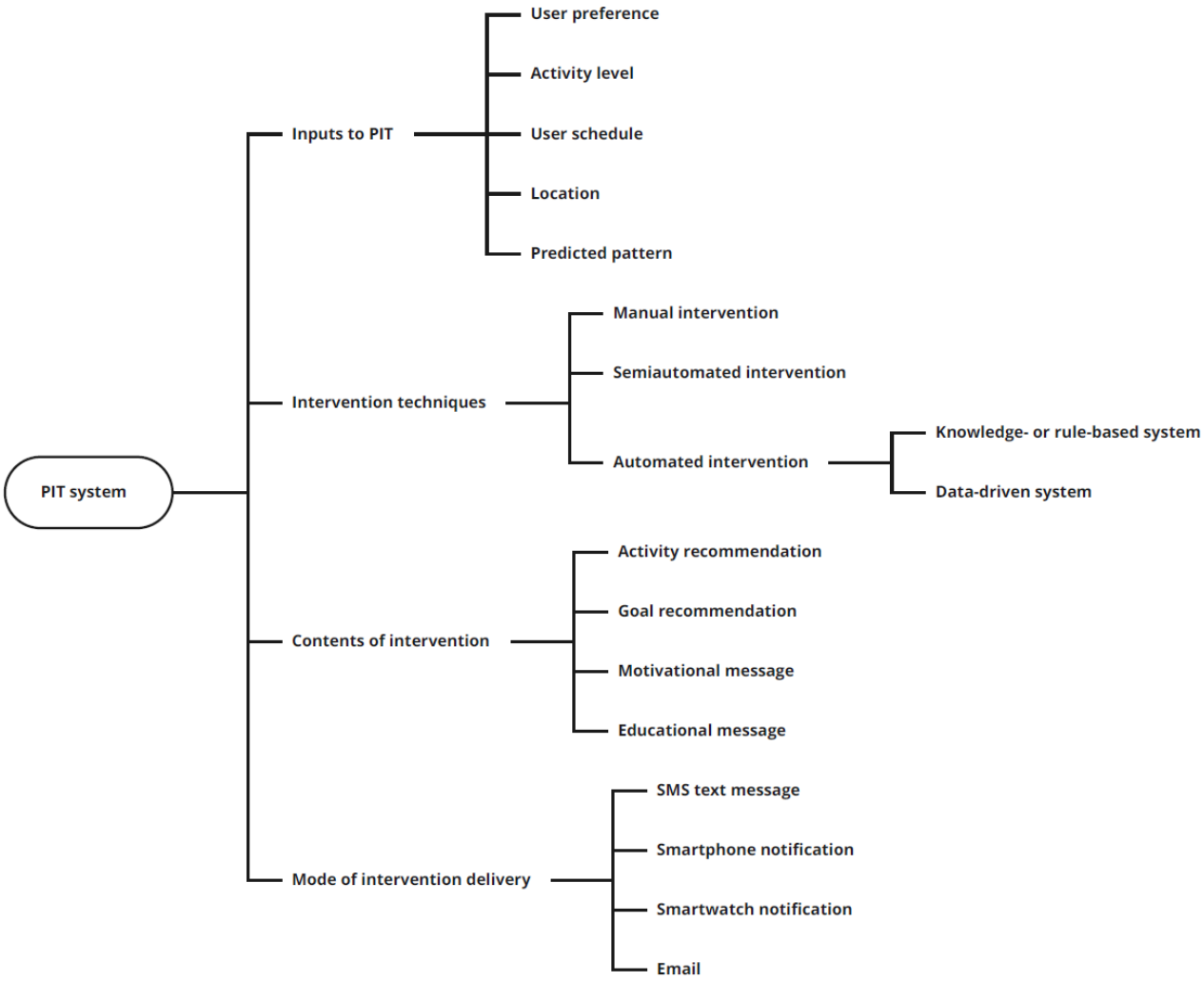
Increasing PA was the primary research objective in 72% (13/18) of the studies [35,41,44,50-59]. Of these 13 studies, 6 (46%) increased PA while reducing the SB of the participants [35,41,50,54,56,57]. Of the 18 studies, 3 (17%) had the primary goal of maintaining a healthy lifestyle, including diet management [49,60] and medication adherence [46], whereas weight loss and weight control were the research goals for 2 (11%) studies [37,61]. Increasing the PA levels of the participants was the secondary objective of these 5 studies.

The intervention systems collected user information in various ways; for example, using fitness trackers [41,44,50-52,59,61] (7/18, 39%), mobile phone sensors [50,54,56-59] (6/18, 33%), self-reported questionnaire [37,46,51,55,57] (5/18, 28%), smartwatches [35,53,54,57,59] (5/18, 28%), or through SMS text messages [44,49,60] (3/18, 17%). Of the 18 studies, 6 (33%) [44,50,51,54,57,59] had used more than one means to collect user information.

The target populations of the studies involved in this review were varied. They included older adults [53,55,57], children and parents [44,60], healthy adults who were sedentary [41], men and women with overweight [35], women with overweight who were sedentary [50], Hispanic individuals with overweight [61], African American individuals who were physically inactive [52], local residents [56,60], pregnant women [37,49], patients with cancer [51,54], and patients with chronic disease [46,58].

As discussed earlier, PIT can be divided into the following four components: (1) inputs to the intervention, (2) intervention techniques to process the inputs, (3) content of the intervention, and (4) mode of delivery for the intervention. Variations in each component were observed across the studies, as discussed in the following sections. These are synthesized in the taxonomy presented in Figure 2.

**Figure 2.** Taxonomy for various components of the personalized intervention timing (PIT) system.




## Inputs to PIT

### *Overview*

PIT was provided considering user-specific information such as user preference, activity level, location, and schedule. On the basis of the attributes used for providing PIT, we classified PIT inputs into five categories: (1) user preference (10/18, 56%), (2) activity level (6/18, 33%), (3) schedule (2/18, 11%), (4) location (1/18, 6%), and (5) predicted patterns (3/18, 17%). These categories are not mutually exclusive because 17% (3/18) of the studies used multiple input types. Table 2 shows the different input types used by the 18 studies in our review.

**Table 2.** Inputs to the personalized intervention timing used by the studies (N=18).

| Article reference | User preference | Activity level | User schedule | Location | Predicted pattern |
|---|---|---|---|---|---|
| Downing et al [44] | ✓ | | | | |
| Finkelstein et al [50] | | ✓ | | | |
| Godino et al [61] | ✓ | | | | |
| Gomersall et al [51] | ✓ | | | | |
| Graham et al [37] | ✓ | | | | |
| Kariuki et al [52] | ✓ | | | | |
| Klasnja et al [41] | ✓ | ✓ | | | |
| Li et al [53] | | | | | ✓ |
| Low et al [54] | | ✓ | | | |
| Maddison et al [46] | ✓ | | | | |
| Mehra et al [55] | ✓ | | | | |
| Militello et al [60] | ✓ | | | | |
| Sporrel et al [56] | | | ✓ | | ✓ |
| Taraldsen et al [57] | | ✓ | | | |
| Thomas et al [35] | | ✓ | | | |
| Vasankari et al [58] | | ✓ | | | |
| Willcox et al [49] | ✓ | | | | |
| Zhao et al [59] | | | ✓ | ✓ | ✓ |

### *User Preference*

This category achieves PIT by delivering the intervention at the user's preferred time. However, this is not a completely automated process. Human mediation from either the health care provider or the participant is needed to log the preferred timings manually.

Of the 18 studies included in this review, 10 (56%) used user preference as input to provide PIT [37,41,44,46,49,51,52,55,60,61]. Users could configure their preferred time of day to receive the intervention messages. Intervention systems in this category allowed the users to personalize the time of intervention delivery considering their schedules, leisure times, and so on. For example, the studies by Graham et al [37] and Mehra et al [55] allowed the user to configure the intervention timing as per their preferences through a website and smartphone app, respectively. It should be noted that such intervention systems require manual selection by the user for the preferred time of intervention message delivery. In contrast, intervention systems using user schedule (covered in the *User Schedule* section) as input infer this information implicitly through the user's calendar app and scheduled activities.

### *Activity Level*

Activity level–based PIT refers to the personalization in timing offered by considering the user's activity level in the recent past (typically 30, 60, or 120 minutes). Activity level–based PIT allows the intervention to be delivered every time the user has been inactive for a specific time, instructing them to either be involved in an MVPA or take a break from SB, typically by performing 2- to 3-minute exercises.

In this review, 33% (6/18) of the studies used user activity levels to offer PIT [35,41,50,54,57,58]. In the study by Finkelstein et al [50], if the user's step count was <15 in the past hour, a message was sent to encourage the user to engage in PA. In the study by Klasnja et al [41], the intervention was delivered at users' preferred time; however, the intervention was not delivered if the user was involved in PA at that moment or had just finished an activity bout in the past 90 seconds. The study by Low et al [54] alerted the user to engage in PA if 60 and 120 minutes of continuous SB occurred and the user reported no severe symptoms such as pain, fatigue, and shortness of breath in the most recent self-reported symptom ratings. In the study by Taraldsen et al [57], the user was prompted after every 30 and 60 minutes of continuous SB to engage in PA. The study by Thomas et al [35] prompted the user to take a 3-, 6-, or 12-minute walking break after 30, 60, or 120 minutes of continuous SB, respectively. Similarly, the study by Vasankari et al [58] notified the user to engage in PA if they had been sitting still for >60 minutes at a stretch. As the prespecified maximum limit of SB varied across studies, the frequency of the interventions also varied across studies.

### *User Schedule*

The studies in this category aimed to deliver an intervention according to the user's day-to-day schedule to ensure that the intervention was not delivered when they were busy. The premise is that even a tailored, actionable intervention is more likely to be ignored by the user if it is delivered when they are engaged in other activities. In this category, the intervention

systems attempt to discern the user's preference by taking their scheduled activities into account without requiring their direct input.

Of the 18 studies included in this review, 2 (11%) [56,59] used user schedules to provide PIT. To ensure that participants were not disturbed when they were otherwise engaged, both studies accessed the users' calendar app to avoid delivering the intervention when they were busy. In addition, the study by Sporrel et al [56] accessed users' calendar app to send them reminders and encouraging messages at the scheduled timings.

### *Location*

Location-based PIT considers user location data to provide an intervention tailored to the location and time. For instance, a recommendation to take a brisk walk can be delivered when the user is on the way to a frequently visited location.

In this review, of the 18 studies, only 1 (6%) [59] included user location to provide PIT. Zhao et al [59] considered the location-specific information captured through mobile phones and smartwatches to decide a suitable location and time for PA using a decision tree–based recommendation engine. Capturing the user's location information allowed the system to deliver a PA intervention with PIT. For example, the intervention system recommended a 15-minute walk to the user when leaving the workplace.

### *Predicted Pattern*

The studies in this category used the user's behavior pattern based on the recorded activity data to deliver an intervention at an appropriate time. The predicted pattern is not an output of an intervention technique; rather, it is the user's behavior pattern obtained from their PA log that is used as an input to provide PIT. The timing of the intervention could either be the predicted onset of SB or the user's frequent timings of PA. The primary difference between activity level–based PIT and predicted pattern–based PIT is when the intervention is provided. Activity level–based PIT provides the intervention on the occurrence of a specified period of user inactivity. In contrast, predicted pattern–based PIT tries to preemptively deliver the intervention to the user based on the behavior patterns extracted from the user's historical data.

In this review, of the 18 studies, 3 (17%) [53,56,59] used user behavior patterns to offer PIT. The study by Li et al [53] identified the patterns in the SB of the user. This study used the data collected by the fitness tracker at the baseline to determine the participant's most inactive period. Subsequently, the intervention for PA was scheduled during the participant's inactive period. In contrast, the study by Sporrel et al [56] determined the time to deliver the intervention based on the participant's PA metrics (such as frequency, duration, speed, and distance in the exercise) on receiving the intervention during a similar situation in the past. The situation was assessed based on weather type, calendar availability, time and date for the intervention delivery, and the PA performed. Finally, Zhao et al [59] used the information from the daily trained user activity model to predict the possible time for PA. For example, the user was recommended a walk to the bus stop on the days they commuted to work.

## Intervention Techniques

### *Overview*

In our reviewed studies, different approaches were used to create the PA intervention with PIT using the aforementioned inputs. On the basis of how the intervention system processed user-specific information, intervention studies could be classified into three categories: manual, semiautomated, and automated. The study by Li et al [53] did not specify the technique used for identifying participants' most inactive period and hence is not categorized. The remaining studies (17/18, 94%) are categorized and discussed in this section.

### *Manual Intervention*

Of the 17 studies, 8 (47%) [37,46,49,51,52,55,60,61] used manual techniques that relied on human mediation from the health care provider or the participant to generate the PIT. It should be noted that these systems relied on human mediation only to create the PIT, not to deliver it. All studies included in this review involved some IT element as the mode of delivery for the intervention.

In this category, the systems recorded user preferences of intervention timing to provide PIT. Of the 8 studies, 6 (75%) [46,49,51,52,60,61] recorded user preferences for receiving the intervention message at registration time or during follow-up sessions, whereas for the remaining 2 (25%) studies, the user could configure the intervention timing by means of a smartphone app [55] or a website portal [37].

### *Semiautomated Intervention*

Semiautomated interventions are systems where a combination of manual and automated techniques is used to determine the PIT specific to the user. The reviewed studies in this category typically used a rule-based approach to provide PIT automatically, along with the user being allowed further flexibility to configure the PIT according to their preference.

Of the 17 studies, 2 (12%) [41,44] used a semiautomated approach in their intervention systems for processing PIT. In the study by Downing et al [44], a few SMS text messages were scheduled to be delivered at particular times of the day to coincide with the activity recommended in the intervention. In addition, participants were asked to nominate a preferred time of the day to receive the SMS text messages. The study by Klasnja et al [41] marked 5 timings in a day, referred to as decision points in their study. At each of the 5 decision points, the system would automatically determine whether the intervention should be delivered to the user based on their availability. The participants were considered unavailable if they were involved in PA at the time or had just finished an activity bout in the past 90 seconds. In addition, the system also allowed users to configure the timings of the 5 decision points based on their schedules.

### *Automated Intervention*

#### Overview

Automated interventions were present in 41% (7/17) of the studies [35,50,54,56-59] and used either knowledge-based or data-driven approaches to automate the PIT. All the

knowledge-based systems were based on decision rules formulated from PA and clinical guidelines. All the data-driven systems used machine learning techniques to learn user models from their historical data.

### Knowledge- or Rule-Based Systems

Of the 7 studies in which automated interventions were present, 5 (71%) [35,50,54,57,58] used knowledge-based approaches. These systems were rule-based and provided feedback and recommendations based on the rules applied to user activity or other user-specific information. An intervention was delivered when the user's continuous inactivity period reached the prespecified limits of SB set in the intervention system [35,50,54,57,58]; for example, users were prompted with an intervention message encouraging them to engage in PA if they had been sitting continuously for >60 minutes.

### Data-Driven Systems

Data-driven intervention systems used machine learning approaches to achieve personalization. Of the 7 studies in which automated interventions were present, 2 (29%) [56,59] incorporated machine learning methods to determine the timing of the PA intervention delivery. The study by Sporrel et al [56] used a reinforcement learning module that optimized the personalized timing based on the user's behavior while using the app over time. To provide personalization in the initial stage (when user behavior data were lacking), training data from a separate study [62] involving 440,000 runs performed by >10,000 users with information about running performance, timing, and weather were used. The study by Zhao et al [59] used a decision tree–based recommendation engine that involved training a daily activity model for each user using activity-related data such as daily calories burned and steps, along with location information captured by mobile phone and smartwatch sensors.

## Contents of Intervention

Across the studies in this review, the types of content of the PA interventions with PIT included activity recommendations, goal recommendations, motivational messages, and educational messages. Table 3 shows the types of intervention contents for each study.

**Table 3.** Intervention contents in the included studies (N=18).

| Article reference | Activity recommendation | Goal recommendation | Motivational message | Educational message |
|---|---|---|---|---|
| Downing et al [44] | ✓ | | | ✓ |
| Finkelstein et al [50] | ✓ | | | |
| Godino et al [61] | ✓ | | | |
| Gomersall et al [51] | ✓ | ✓ | | ✓ |
| Graham et al [37] | | ✓ | | |
| Kariuki et al [52] | | | ✓ | |
| Klasnja et al [41] | ✓ | | | |
| Li et al [53] | ✓ | ✓ | ✓ | |
| Low et al [54] | ✓ | | | |
| Maddison et al [46] | ✓ | ✓ | | |
| Mehra et al [55] | | ✓ | | |
| Militello et al [60] | | | ✓ | |
| Sporrel et al [56] | ✓ | ✓ | | |
| Taraldsen et al [57] | | | ✓ | |
| Thomas et al [35] | ✓ | | | |
| Vasankari et al [58] | | ✓ | ✓ | |
| Willcox et al [49] | | ✓ | | |
| Zhao et al [59] | ✓ | | | |

Specifically, of the 18 studies, 11 (61%) [35,41,44,46,50,51,53,54,56,59,61] that provided activity recommendations prescribed one or more activities to the user. For example, in the study by Klasnja et al [41], participants were suggested to park their vehicle farther from the office and encouraged to walk during the morning commute to work. Of the 18 studies, 8 (44%) [37,46,49,51,53,55,56,58] that offered goal recommendations delivered personalized goals or a reminder to users to achieve their goals. For example, in the study by Sporrel et al [56], users would be reminded of their daily goal and given feedback on their current activity level when the intervention was delivered, encouraging them to achieve their goal. Of the 18 studies, 5 (28%) [52,53,57,58,60] aimed to encourage users to engage in PA by delivering motivational messages at an appropriate time. For example, in the study by Militello et al [60], participants could craft motivational messages that would be delivered to them at tailored timings during the following weeks. Finally, of the 18 studies, 2 (11%) [44,51] used educational messages that aimed to increase users' knowledge regarding the importance of PA.

For example, the study by Gomersall et al [51] informed users about the health benefits for the heart as a consequence of reducing SB and increasing PA.

### Mode of Intervention Delivery

As per our selection criteria, all the studies in our review used some form of IT to deliver the PA intervention. This is different from the manual intervention technique defined earlier, which implies that the PIT could be determined manually, albeit delivered through an IT-based system.

For the mode of intervention delivery, various forms of IT, such as SMS text messages, smartphone app notification, smartwatch notification, and emails, were used as a communication medium. In our review, 39% (7/18) of the studies [44,46,49-51,60,61] used only SMS text messages as the mode of intervention delivery. Furthermore, 28% (5/18) used smartphone notification alone [35,41,55,56,58]. In comparison, 22% (4/18) used both smartwatch and smartphone notifications [53,54,57,59], whereas 11% (2/18) used email and SMS text messages [37,52].

### Theories Used

Of the 18 studies included in this review, 11 (61%) used a theoretical framework for providing their intervention. The theories were used to make design decisions regarding the intervention delivery, selection of study variables for the intervention, or personalizing the intervention content. The theories used were the Beck cognitive theory [63] (1/18, 6%); behavioral intervention technology (BIT) model [64] (1/18, 6%); capability, opportunity, motivation, and behavior (COM-B) model [65] (1/18, 6%); integrative model of behavioral prediction [66] and behavior model for persuasive design [67] (1/18, 6%); control theory [68] (1/18, 6%); the Fogg behavior model (FBM) [69] (2/18, 11%); social cognitive theory (SCT) [70] (5/18, 28%); and self-efficacy theory (SET) [71] (2/18, 11%).

The Beck cognitive theory [63] was used in the study by Militello et al [60] to tailor the intervention content and identify study variables such as knowledge, perceived difficulty, beliefs, and behaviors. In contrast, the BIT model [64] and COM-B model [65] were used in the study by Sporrel et al [56] to guide the implementation and design of the persuasive strategies used in their intervention system. The COM-B model proposes the interrelationship among users' capability, opportunity for action, and the motivation required to change user behavior. The intervention system included goal setting, feedback, and reminders guided by the COM-B model [65]. The BIT model guided the implementation design decisions, such as the form and timing of the intervention, complexity, and esthetics of the app developed.

In the study by Graham et al [37], the theoretical framework provided by the integrative model of behavioral prediction [66] combined with the behavior model for persuasive design [67] was used in the formative research to determine the main features of the intervention. The behavior model for persuasive design [67] explored the role of computing systems as persuasive social actors and various persuasive strategies used to elicit a response from the user. The model provides insights on different persuasive techniques that can be used to increase human interaction with the systems, including health intervention systems. The integrative model of behavioral prediction [66] demonstrates the simultaneous use of two theories, behavioral prediction and media priming theory, to develop effective health interventions.

In the study by Mehra et al [55], elements of control theory [68] such as goal setting and self-monitoring were used in the design considerations to formulate the app's functional requirements designed for the intervention. Furthermore, various FBM [69] elements were used by 11% (2/18) [56,60] of the studies to construct the conceptual model of their intervention system. The FBM [69] states that the user must simultaneously have sufficient motivation, sufficient ability, and an effective trigger for the behavior to occur. For example, Militello et al [60] used SMS text messages as a medium to provide a trigger, which is among the three principal elements (motivation, ability, and trigger) defined in the FBM, to the user to promote healthy behavior. Similarly, in the study by Sporrel et al [56], a timely trigger was provided to the user through an app notification.

SCT [70], used by 28% (5/18) of the studies in this review [44,49,52,57,61], postulates the reciprocal relationship between an individual and the environment and personal factors such as self-efficacy, self-control, and behavioral capability to theorize how an individual acquires and maintains a particular behavior. SCT aims to focus on initiating behavior and explain how to achieve a behavior change that is maintained over time. The study by Downing et al [44] used the SCT taxonomy to tailor the content of the intervention. The study by Godino et al [61] used strategies for weight management that included evidence and SCT constructs to tailor the content of the intervention. In the study by Kariuki et al [52], SCT was used to select the workout videos recommended to the users to match their preferences. Elements of SCT were adopted in the study by Taraldsen et al [57] to make design decisions regarding the intervention system. In the study by Willcox et al [49], the design of the intervention system was based on SCT.

SET [71], used by 11% (2/18) of the studies in this review [46,53], defines self-efficacy as a personal judgment of "how well one can execute courses of action required to deal with prospective situations" [71]. The SET states that there are four approaches to increase a person's self-efficacy: enactive mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective feedback. In the study by Li et al [53], the intervention aimed to enhance the self-efficacy of the user by providing mastery experiences and verbal persuasion by recommending challenging yet attainable goals and providing interaction-enabled prompts and feedback to the user, whereas in the study by Maddison et al [46], the content of the intervention was based on the SET.

### Results of Individual Studies

In our review, only 33% (6/18) of the studies [41,43,49-51,53] presented evaluations of their interventions to increase PA. Of the remaining 12 studies, 4 (33%) [52,56-58] had not yet completed the intervention and thus did not present results, whereas 8 (67%) [35,37,46,54,55,59-61] did not evaluate the effects of the PA intervention. Of the 6 studies carrying out PA evaluations, 5 (83%) [41,43,49-51] conducted randomized

controlled trials, whereas 1 (17%) [53] conducted a pilot test. Table 4 shows the evaluation variables and results for these studies (n=6).

Specifically, Downing et al [44] evaluated the sitting time and MVPA in minutes for the participant children in the control and intervention groups at baseline and after the intervention. Parent-reported sitting time and objective sitting time, as measured by the activPAL device, recorded a decrease in children's sitting time. The reduction in objective sitting time in the intervention group was more than that in the control group: 25.8 minutes per day in the intervention group compared with 3.7 minutes per day in the control group.

**Table 4.** Results of the studies that evaluated their physical activity (PA) intervention (N=6).

| Article reference | Method of study | Participants, n | Variables evaluated | Results |
|---|---|---|---|---|
| Downing et al [44] | RCT[a] | 57 | Screen time and sitting time | Participants in the intervention group reduced their total screen time by 30.6 minutes per day, whereas the screen time increased by 7.5 minutes per day for participants in the control group. Sitting time was reduced in the intervention group by 25.8 minutes per day and in the control group by 3.7 minutes per day |
| Finkelstein et al [50] | RCT | 30 | Inactivity and number of steps | Inactivity was significantly lower ($P$<.02) during *message on* periods compared with *message off* periods. Increased mean (+584.34 steps) in the number of steps was recorded by group A, whereas group B recorded a reduced mean (–71.94 steps) during *message on* periods compared with *message off* periods |
| Gomersall et al [51] | RCT | 38 | MVPA[b] in minutes and sitting time | At the 12-week follow-up, the intervention group participants reduced their overall sitting and prolonged sitting time by 40-50 minutes per 16 hours awake and reported an increase in standing and light-intensity stepping. No significant changes were recorded in the objectively measured activity level of the control group. No group reported any significant change in MVPA assessed by the activPAL device |
| Klasnja et al [41] | RCT | 44 | Number of steps | Delivering a suggestion vs no suggestion increased the 30-minute step count by 14% ($P$=.06), an increase of 35 steps over the 253-step average |
| Li et al [53] | Pilot | 8 | Number of steps, sleep index, PA, and sedentary time | The participants' sedentary time decreased, and they spent less of their waking time on sedentary activities during the intervention ($P$=.03) and after the intervention ($P$<.01). On average, the participants' PA increased significantly after the intervention ($P$=.02) |
| Willcox et al [49] | RCT | 100 | Activity time in minutes and participants' weight | From the baseline to the conclusion of the intervention period, the women in the intervention group reported significantly smaller reductions in total, light-, and moderate-intensity PA ($P$=.001) than the women in the control group |

[a]RCT: randomized controlled trial.

[b]MVPA: moderate to vigorous physical activity.

In the study by Finkelstein et al [50], a randomized crossover design was used, with group A participants receiving tailored intervention messages for the first 4 weeks, followed by 4 weeks of no interventions. In contrast, group B participants received no intervention messages in the initial 4 weeks and were switched to tailored intervention messages in the 4 weeks that followed. *Message on* was used to indicate the duration of the study when interventions were delivered to the participants, whereas *message off* indicated the period when no interventions were delivered to the participants. Interestingly, although the overall inactivity period was significantly reduced, the mean number of steps recorded by group B was lower in the period when the intervention was delivered than when no intervention was delivered. In contrast, in the study by Gomersall et al [51], no significant differences in objectively measured MVPA were recorded between the intervention group and the control group, but a significant difference in self-reported MVPA between the groups was observed at the 4-week and 12-week follow-ups.

In the study by Klasnja et al [41], 30 minutes after the intervention was delivered, an increase in the average step counts was recorded. The group receiving contextually tailored activity suggestions also recorded an increase in the number of steps compared with the group with no interventions. However,

the group receiving the contextually tailored activity suggestions experienced a significant attrition rate.

The pilot test by Li et al [53] evaluated variables such as the number of steps, sleep index, PA, and sedentary time. Reduced SB and increased PA levels during the intervention and after the intervention were recorded. Finally, in the study by Willcox et al [49], participants in the intervention group were reported to be less likely to reduce PA levels throughout the intervention.

Thus, the studies in our review that carried out evaluations (6/18, 33%) have shown positive results regarding increasing PA and reducing SB.

## Discussion

### Principal Findings

In this study, we conducted a review of IT-based PA intervention studies that provided PIT and synthesized them to offer an overview of PIT research for PA improvement. We identified and categorized user-specific inputs, intervention techniques, intervention content, delivery modes, and theories used by intervention studies with PIT to increase PA.

Thus, this study contributes to the literature on personalization for PA interventions and specifically to the research on PIT. Although earlier reviews on personalized interventions for increasing PA focused on their classification [25] or evaluated the effectiveness of the interventions [25], we examined the intervention timing in depth, which is valuable for increasing intervention adherence and thereby improving PA [26,27]. Prior reviews, that is, the study by Tong et al [29], aimed more broadly at evaluating the effectiveness of a personalized mobile intervention in promoting lifestyle behavior change or focused more narrowly on studies of personalization for PA coaching systems, that is, the review by op den Akker et al [30]. Thus, our review is able to make a contribution by explicating the dimension of timing in the personalization of PA interventions.

In the next sections, we further discuss the implications of our review with respect to the inputs to PIT, intervention techniques, intervention content, mode of intervention delivery, theories used, and results.

### Inputs to PIT

The reviewed studies used input factors of user preferences, activity levels, user schedules, locations, and predicted patterns to provide PIT. Despite the evident influence of temporal and environmental factors such as day of the week, time of day, and weather on the choice made by the user regarding PA [17], only the study by Sporrel et al [56] used such temporal and contextual factors to provide PIT. Thus, we found that temporal and contextual factors have rarely been considered for the purpose of providing PIT aimed at increasing PA. Future studies could include and assess the effectiveness of such temporal and environmental factors as input attributes in the intervention system to provide PIT.

Of the 18 studies included in the review, 3 (17%) used more than one input attribute to provide PIT [41,56,59]. For instance, Zhao et al [59] considered user schedule, location, and user behavior pattern to provide PIT. However, none of the studies in this review evaluated the effectiveness of combining input types. The inclusion of multiple types of user-specific inputs in the evaluation could allow for a more holistic understanding of contextual information related to the user and therefore offer a better-informed decision for PIT. In addition, selecting a combination of types of inputs is also dependent on the intervention technique used by the system. For example, the selection of predicted behavior could be infeasible in a study that uses a manual approach as the intervention technique. Hence, a combination of input types appropriate to the system's intervention technique should be selected. Future research could consider evaluating the effectiveness of the combinations of input types applicable to the respective interventions' technique.

Of the 18 studies in this review, 6 (33%) [35,41,50,54,57,58] used user activity levels to provide PIT based on the prespecified maximum limit of SB. However, most of them did not clarify how they set the maximum SB limit. Although these studies had a prespecified maximum limit of SB of 30, 60, or 120 minutes, these limits do not adhere to the clinical or health institute and organization guidelines. Future studies could use the maximum limit of SB as stipulated in standard PA guidelines by clinical or health institutes and organizations such as the Health Promotion Board, Singapore [72], to be more scientific and rigorous.

Interestingly, the user's physical geolocation was used only in the study by Zhao et al [59] to provide location-triggered interventions. Although the study reported positive results, the method and design of the intervention study imply that the effectiveness cannot be attributed to that specific factor. Further research is needed to assess the effectiveness of physical geolocation regarding PIT for increasing PA. It should be noted that the lack of research studies including location information could be due to the privacy concerns posed by the location tracking of users.

### Intervention Techniques

The intervention studies included in this review used either manual, semiautomated, or automated methods to determine PIT. As mentioned earlier, semiautomated intervention systems used an automated component, typically knowledge- or rule-based, coupled with flexibility for human mediation. Both studies [41,44] using semiautomated intervention systems reported positive results; nevertheless, neither study evaluated the effectiveness of such intervention systems against the individual components, that is, manual approach and automated approach. Hence, it is unclear if the added complexity of combining automated and manual approaches in the semiautomated approach results in improved intervention effectiveness compared with using manual and automated approaches individually. Therefore, future studies could undertake a comparative study by evaluating the effectiveness of each intervention technique.

Similarly, among the studies providing automated interventions [35,50,54,56-59], none evaluated intervention effectiveness using a combination of rule-based and data-driven methodology to achieve PIT compared with individual components, that is, rule-based and data-driven approaches. An intervention design combining both methods would allow the integration of rules

based on guidelines with sophisticated machine learning algorithms. Multidimensional recommendation systems that consider context information, including user-specific information and environmental factors, novel approaches for temporal profiling, and similar advanced techniques, could be used to enhance the effectiveness of the intervention. Future studies could consider exploring and evaluating the effectiveness of intervention systems combining both approaches.

### Contents of Intervention

Across the studies included in this review, the intervention content types observed were activity recommendation, goal recommendation, motivational message, and educational message. Although a few studies offered multiple types of content for the intervention, none of the studies included in this review rigorously evaluated the effectiveness of each type of content or the combination of the types of content. Therefore, specifics of which type or combination of types of content should be used for the intervention delivered at personalized timings to maximize the effectiveness is unclear. For example, the likelihood of the user performing PA when provided with an activity recommendation can vary compared with the likelihood of performing PA if a motivational message was delivered. Future studies could consider designing the intervention system to evaluate the effectiveness of each intervention content type and combinations of types.

### Mode of Intervention Delivery

No conclusive evidence justifying the selection of the mode of communication was provided in the reviewed studies. This indicates a lack of research regarding the effectiveness of each mode of intervention delivery regarding PIT aimed at increasing PA. Furthermore, user preference for a particular mode of intervention delivery could also vary across user groups. For example, older adults might prefer SMS text messages over notifications on smart devices because of their simplicity. Future research could consider evaluating the effectiveness of each intervention delivery mode or a combination of intervention delivery modes and across different user groups.

### Theories Used

Theories can help to explain the mechanism and techniques that change user behavior and provide insights into various aspects, including system design decisions and personalization strategies. However, in this review, only the study by Mehra et al [55] used theory guidelines to personalize the intervention timing, whereas the remaining studies (10/11, 91%) did it only for the PA intervention itself. Future studies could consider further design and evaluation of theory-based interventions to personalize the intervention timing. In addition, theories on temporal aspects such as circadian rhythm theory [73] could be adopted to design intervention systems.

### Results of Individual Studies

The reviewed studies evaluated different PA-related metrics, with the number of steps being assessed by 50% (3/6) of the studies that evaluated PA-related metrics. Other interventions evaluated other metrics such as activity time in minutes, MVPA in minutes, and sitting time. Although the studies (6/18, 33%) reported positive results regarding the users' PA levels and SB, the lack of standard evaluation metrics for PA-related intervention studies hinders the objective comparison of the results across studies. Future studies could use standardized metrics for PA measurement and establish the correlations among existing metrics to facilitate evaluation and comparison of studies.

Although this review focused on PIT research for PA improvement, none of the studies in our review rigorously evaluated the effect of PIT on PA levels of the user. PIT was offered along with other forms of personalization, making it infeasible to evaluate the effectiveness of PIT alone to increase PA. Future research could rigorously assess PIT impacts; for example, by using randomized controlled trials to evaluate the effectiveness of PIT in improving PA levels compared with an intervention delivered with nonpersonalized timings.

### Limitations

This review includes a few limitations. First, it was restricted to select databases for searching relevant articles and the search query was limited to a time frame that was considered relevant for this review. This could have led to a few relevant studies being left out of this review because of their journal or indexing bias. Second, the reviewers were not blinded to each other's decisions during the study screening procedure, which may have led to a study selection bias. Third, because this is a scoping review, we have included studies without quality analysis and without any evaluation. Although this helps to identify the breadth of research, because the quality of studies is not assessed, the gaps identified may not be completely accurate. Fourth, the lack of strict restriction on the intervention method led to diverse outcomes across studies; therefore, a meta-analysis was not possible in this review. Hence, the results for the studies that evaluated the intervention's effectiveness for increasing PA could not be pooled together for statistical analysis. Finally, we could not assess those studies in which increasing PA was a secondary objective because they did not report the results of their study.

## Conclusions

This review assessed aspects of the intervention system providing PIT to increase PA. The studies evaluated PIT in conjunction with other personalization approaches such as activity recommendation, with no study evaluating the effectiveness of PIT alone. On the basis of the findings from this review, the following research directions for increasing the effectiveness of personalized interventions are proposed. First, the effectiveness of PIT in PA interventions is yet to be rigorously evaluated, although preliminary studies in this direction are promising. Second, the effectiveness of temporal and environmental factors as inputs and a combination of input types should be evaluated. Third, combinations of intervention content and mode of intervention delivery need to be evaluated. Fourth, standardized metrics for PA measurement and correlations among existing metrics should be established. Fifth, automated intervention systems need to be adapted to integrate clinical guidelines with sophisticated machine learning algorithms. Several important directions for future research are also highlighted in this review.

## Conflicts of Interest

None declared.

## References


1. Durstine JL, Gordon B, Wang Z, Luo X. Chronic disease and the link to physical activity. J Sport Health Sci 2013 Mar;2(1):3-11. [doi: 10.1016/j.jshs.2012.07.009]
2. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
3. Participation in leisure-time aerobic and muscle-strengthening activities that meet the federal. Centers for Disease Control and Prevention. 2008. URL: https://www.cdc.gov/nchs/data/hus/2019/025-508.pdf [accessed 2021-06-09]
4. Deb S. Information technology, its impact on society and its future. Adv Comput 2014;4(1):25-29 [FREE Full text]
5. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
6. Goasduff L. Gartner says global end-user spending on wearable devices to total $52 billion in 2020. Gartner. URL: https://www.gartner.com/en/newsroom/press-releases/2019-10-30-gartner-says-global-end-user-spending-on-wearable-dev [accessed 2021-05-30]
7. Fitbit. URL: https://www.fitbit.com/global/us/products/services/premium [accessed 2021-05-25]
8. Physical activity guidelines for Americans. U.S. Department of Health and Human Services. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2021-05-25]
9. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/336656/9789240015128-eng.pdf [accessed 2021-05-20]
10. Kreuter M, Farrell D, Olevitch L, Brennan L. Tailoring Health Messages: Customizing Communication With Computer Technology. New York: Lawrence Erlbaum Associates; 2000.
11. Marsaux CF, Celis-Morales C, Fallaize R, Macready AL, Kolossa S, Woolhead C, et al. Effects of a web-based personalized intervention on physical activity in European adults: a randomized controlled trial. J Med Internet Res 2015 Oct 14;17(10):e231 [FREE Full text] [doi: 10.2196/jmir.4660] [Medline: 26467573]
12. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
13. Henriksen A, Haugen MM, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
14. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
15. Marquet O, Alberico C, Adlakha D, Hipp JA. Examining motivations to play Pokémon GO and their influence on perceived outcomes and physical activity. JMIR Serious Games 2017 Oct 24;5(4):e21 [FREE Full text] [doi: 10.2196/games.8048] [Medline: 29066423]
16. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
17. Wang S, Scheider S, Sporrel K, Deutekom M, Timmer J, Kröse B. What are good situations for running? A machine learning study using mobile and geographical data. Front Public Health 2020;8:536370 [FREE Full text] [doi: 10.3389/fpubh.2020.536370] [Medline: 33490006]
18. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
19. 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]
20. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151. [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
21. Zimmer P, Baumann FT, Oberste M, Schmitt J, Joisten N, Hartig P, et al. Influence of personalized exercise recommendations during rehabilitation on the sustainability of objectively measured physical activity levels, fatigue, and fatigue-related biomarkers in patients with breast cancer. Integr Cancer Ther 2018 Jun;17(2):306-311 [FREE Full text] [doi: 10.1177/1534735417713301] [Medline: 28617135]

22. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]
23. Smith KL, Kerr DA, Fenner AA, Straker LM. Adolescents just do not know what they want: a qualitative study to describe obese adolescents' experiences of text messaging to support behavior change maintenance post intervention. J Med Internet Res 2014;16(4):e103 [FREE Full text] [doi: 10.2196/jmir.3113] [Medline: 24713407]
24. Haile C, Kirk A, Cogan N, Janssen X, Gibson A, MacDonald B. Pilot testing of a nudge-based digital intervention (Welbot) to improve sedentary behaviour and wellbeing in the workplace. Int J Environ Res Public Health 2020 Aug 10;17(16) [FREE Full text] [doi: 10.3390/ijerph17165763] [Medline: 32784956]
25. Ghanvatkar S, Kankanhalli A, Rajan V. User models for personalized physical activity interventions: scoping review. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11098 [FREE Full text] [doi: 10.2196/11098] [Medline: 30664474]
26. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
27. Sellak H, Grobler M. MHealth4U: designing for health and wellbeing self-management. In: Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering Workshops.: Association for Computing Machinery; 2020 Presented at: ASE '20: 35th IEEE/ACM International Conference on Automated Software Engineering Workshops; 2020; New York, NY, USA URL: https://doi.org/10.1145/3417113.3422179 [doi: 10.1145/3417113.3422179]
28. Aldenaini N, Orji R, Sampalli S. How effective is personalization in persuasive interventions for reducing sedentary behavior and promoting physical activity: a systematic review. In: Proceedings of the Persuasive Technology Conference. 2020 Presented at: Persuasive Technology Conference; 2020; Aalborg, Denmark URL: https://tinyurl.com/bderbpt3
29. Tong HL, Quiroz JC, Kocaballi AB, Fat SC, Dao KP, Gehringer H, et al. Personalized mobile technologies for lifestyle behavior change: a systematic review, meta-analysis, and meta-regression. Prev Med 2021 Jul;148:106532. [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]
30. op den Akker H, Jones VM, Hermens HJ. Tailoring real-time physical activity coaching systems: a literature survey and model. User Model User-Adap Inter 2014 Jun 25;24(5):351-392. [doi: 10.1007/s11257-014-9146-y]
31. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
32. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
34. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
35. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
36. Fernandez ID, Groth SW, Reschke JE, Graham ML, Strawderman M, Olson CM. eMoms: electronically-mediated weight interventions for pregnant and postpartum women. Study design and baseline characteristics. Contemp Clin Trials 2015 Jul;43:63-74 [FREE Full text] [doi: 10.1016/j.cct.2015.04.013] [Medline: 25957183]
37. Graham ML, Uesugi KH, Niederdeppe J, Gay GK, Olson CM. The theory, development, and implementation of an e-intervention to prevent excessive gestational weight gain: e-Moms Roc. Telemed J E Health 2014 Dec;20(12):1135-1142 [FREE Full text] [doi: 10.1089/tmj.2013.0354] [Medline: 25354350]
38. Olson CM, Groth SW, Graham ML, Reschke JE, Strawderman MS, Fernandez ID. The effectiveness of an online intervention in preventing excessive gestational weight gain: the e-moms roc randomized controlled trial. BMC Pregnancy Childbirth 2018 May 09;18(1):148 [FREE Full text] [doi: 10.1186/s12884-018-1767-4] [Medline: 29743026]
39. Greenewald K, Tewari A, Klasnja P, Murphy S. Action centered contextual bandits. Adv Neural Inf Process Syst 2017 Dec;30:5973-5981 [FREE Full text] [Medline: 29225449]
40. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
41. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
42. Liao P, Greenewald K, Klasnja P, Murphy S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-22 [FREE Full text] [doi: 10.1145/3381007] [Medline: 34527853]

43. Downing KL, Salmon J, Hinkley T, Hnatiuk JA, Hesketh KD. A mobile technology intervention to reduce sedentary behaviour in 2- to 4-year-old children (Mini Movers): study protocol for a randomised controlled trial. Trials 2017 Mar 03;18(1):97 [FREE Full text] [doi: 10.1186/s13063-017-1841-7] [Medline: 28253904]
44. Downing KL, Salmon J, Hinkley T, Hnatiuk JA, Hesketh KD. Feasibility and efficacy of a parent-focused, text message-delivered intervention to reduce sedentary behavior in 2- to 4-year-old children (Mini Movers): pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e39 [FREE Full text] [doi: 10.2196/mhealth.8573] [Medline: 29426816]
45. Dale LP, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Improving coronary heart disease self-management using mobile technologies (Text4Heart): a randomised controlled trial protocol. Trials 2014 Mar 04;15:71 [FREE Full text] [doi: 10.1186/1745-6215-15-71] [Medline: 24588893]
46. Maddison R, Stewart R, Doughty R, Scott T, Kerr A, Benatar J, et al. Text4Heart II – improving medication adherence in people with heart disease: a study protocol for a randomized controlled trial. Trials 2018 Jan 25;19(1):70 [FREE Full text] [doi: 10.1186/s13063-018-2468-z] [Medline: 29370829]
47. Dale LP, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the text4heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
48. Willcox JC, Campbell KJ, McCarthy EA, Wilkinson SA, Lappas M, Ball K, et al. Testing the feasibility of a mobile technology intervention promoting healthy gestational weight gain in pregnant women (txt4two) - study protocol for a randomised controlled trial. Trials 2015 May 07;16:209 [FREE Full text] [doi: 10.1186/s13063-015-0730-1] [Medline: 25947578]
49. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
50. Finkelstein J, Bedra M, Li X, Wood J, Ouyang P. Mobile app to reduce inactivity in sedentary overweight women. Stud Health Technol Inform 2015;216:89-92. [Medline: 26262016]
51. Gomersall SR, Skinner TL, Winkler E, Healy GN, Eakin E, Fjeldsoe B. Feasibility, acceptability and efficacy of a text message-enhanced clinical exercise rehabilitation intervention for increasing 'whole-of-day' activity in people living with and beyond cancer. BMC Public Health 2019 Jun 03;19(Suppl 2):542 [FREE Full text] [doi: 10.1186/s12889-019-6767-4] [Medline: 31159752]
52. Kariuki JK, Gibbs BB, Erickson KI, Kriska A, Sereika S, Ogutu D, et al. The feasibility and acceptability of a web-based physical activity for the heart (PATH) intervention designed to reduce the risk of heart disease among inactive African Americans: protocol for a pilot randomized controlled trial. Contemp Clin Trials 2021 May;104:106380. [doi: 10.1016/j.cct.2021.106380] [Medline: 33798731]
53. Li J, Hodgson N, Lyons MM, Chen K, Yu F, Gooneratne NS. A personalized behavioral intervention implementing mHealth technologies for older adults: a pilot feasibility study. Geriatr Nurs 2020 May;41(3):313-319 [FREE Full text] [doi: 10.1016/j.gerinurse.2019.11.011] [Medline: 31810730]
54. Low CA, Danko M, Durica KC, Kunta AR, Mulukutla R, Ren Y, et al. A real-time mobile intervention to reduce sedentary behavior before and after cancer surgery: usability and feasibility study. JMIR Perioper Med 2020 Mar 23;3(1):e17292 [FREE Full text] [doi: 10.2196/17292] [Medline: 33393915]
55. Mehra S, Visser B, Dadema T, van den Helder J, Engelbert RH, Weijs PJ, et al. Translating behavior change principles into a blended exercise intervention for older adults: design study. JMIR Res Protoc 2018 May 02;7(5):e117 [FREE Full text] [doi: 10.2196/resprot.9244] [Medline: 29720358]
56. Sporrel K, De Boer RD, Wang S, Nibbeling N, Simons M, Deutekom M, et al. The design and development of a personalized leisure time physical activity application based on behavior change theories, end-user perceptions, and principles from empirical data mining. Front Public Health 2020;8:528472 [FREE Full text] [doi: 10.3389/fpubh.2020.528472] [Medline: 33604321]
57. Taraldsen K, Mikolaizak AS, Maier AB, Boulton E, Aminian K, van Ancum J, et al. Protocol for the PreventIT feasibility randomised controlled trial of a lifestyle-integrated exercise intervention in young older adults. BMJ Open 2019 Mar 20;9(3):e023526 [FREE Full text] [doi: 10.1136/bmjopen-2018-023526] [Medline: 30898801]
58. Vasankari V, Halonen J, Husu P, Vähä-Ypyä H, Tokola K, Suni J, et al. Personalised eHealth intervention to increase physical activity and reduce sedentary behaviour in rehabilitation after cardiac operations: study protocol for the PACO randomised controlled trial (NCT03470246). BMJ Open Sport Exerc Med 2019;5(1):e000539 [FREE Full text] [doi: 10.1136/bmjsem-2019-000539] [Medline: 31354960]
59. Zhao Z, Arya A, Orji R, Chan G. Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study. JMIR Serious Games 2020 Nov 17;8(4):e19968 [FREE Full text] [doi: 10.2196/19968] [Medline: 33200994]
60. Militello L, Melnyk BM, Hekler EB, Small L, Jacobson D. Automated behavioral text messaging and face-to-face intervention for parents of overweight or obese preschool children: results from a pilot study. JMIR Mhealth Uhealth 2016 Mar 14;4(1):e21 [FREE Full text] [doi: 10.2196/mhealth.4398] [Medline: 26976387]

61. Godino JG, Golaszewski NM, Norman GJ, Rock CL, Griswold WG, Arredondo E, et al. Text messaging and brief phone calls for weight loss in overweight and obese English- and Spanish-speaking adults: a 1-year, parallel-group, randomized controlled trial. PLoS Med 2019 Sep;16(9):e1002917 [FREE Full text] [doi: 10.1371/journal.pmed.1002917] [Medline: 31553725]
62. Wang S, Timmer J, Scheider S, Sporrel K, Akata Z, Kröse B. A data-driven study on preferred situations for running. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. New York, NY, USA: Association for Computing Machinery; 2018 Presented at: UbiComp '18: The 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing; October 8 - 12, 2018; Singapore p. 283-286. [doi: 10.1145/3267305.3267552]
63. Beck A, Rush A, Shaw B, Emery G. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
64. 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]
65. 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]
66. Fishbein M, Yzer MC. Using theory to design effective health behavior interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]
67. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Burlington, Massachusetts, United States: Morgan Kaufmann; 2003.
68. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality–social, clinical, and health psychology. Psychol Bull 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]
69. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive 2009; 4th International Conference on Persuasive Technology; April 26 - 29, 2009; Claremont, California, USA p. 1-7. [doi: 10.1145/1541948.1541999]
70. Bandura A. Social foundations of thought and action: a social cognitive theory. In: Prentice-Hall Series in Social Learning Theory. Hoboken, New Jersey, United States: Prentice-Hall Inc; 1985.
71. Bandura A, Adams N. Analysis of self-efficacy theory of behavioral change. Cogn Ther Res 1977 Dec;1(4):287-310 [FREE Full text] [doi: 10.1007/bf01663995]
72. National Physical Activity Guidelines: summary guide for professionals. Health Promotion Board, Singapore. URL: https://www.healthhub.sg/sites/assets/Assets/PDFs/HPB/PhysicalActivityPDFs/NPAG_Summary_Guide.pdf [accessed 2021-10-21]
73. Drust B, Waterhouse J, Atkinson G, Edwards B, Reilly T. Circadian rhythms in sports performance--an update. Chronobiol Int 2005;22(1):21-44. [doi: 10.1081/cbi-200041039] [Medline: 15865319]


## Abbreviations

**BIT:** behavioral intervention technology
**COM-B:** capability, opportunity, motivation, and behavior
**FBM:** Fogg behavior model
**IT:** information technology
**MVPA:** moderate to vigorous physical activity
**PA:** physical activity
**PIT:** personalized intervention timing
**SB:** sedentary behavior
**SCT:** social cognitive theory
**SET:** self-efficacy theory


*Edited by L Buis; submitted 17.06.21; peer-reviewed by B Chaudhry, S Bhattacharjya, O Perski; comments to author 27.08.21; revised version received 21.10.21; accepted 07.01.22; published 28.02.22.*

*Please cite as:*
*Chaudhari S, Ghanvatkar S, Kankanhalli A*
*Personalization of Intervention Timing for Physical Activity: Scoping Review*
*JMIR Mhealth Uhealth 2022;10(2):e31327*
*URL: https://mhealth.jmir.org/2022/2/e31327*
*doi:10.2196/31327*
*PMID:35225811*

Original Paper

# Using Smartphones to Reduce Research Burden in a Neurodegenerative Population and Assessing Participant Adherence: A Randomized Clinical Trial and Two Observational Studies

Anna L Beukenhorst[1,2], MSc, PhD; Katherine M Burke[3], MSc, DPT; Zoe Scheier[3], BSc; Timothy M Miller[4], MD, PhD; Sabrina Paganoni[3,5], MD, PhD; Mackenzie Keegan[3], BSc; Ella Collins[3], BSc; Kathryn P Connaghan[6], PhD; Anna Tay[4], BSc; James Chan[7], MA, MEd; James D Berry[3], MPH, MD; Jukka-Pekka Onnela[1], DSC

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[2]Centre for Epidemiology Versus Arthritis, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom
[3]Neurological Clinical Research Institute and Sean M. Healey & AMG Center for ALS, Massachusetts General Hospital, Boston, MA, United States
[4]Department of Neurology, Washington University, Saint Louis, MO, United States
[5]Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA, United States
[6]MGH Institute of Health Professions, Charlestown, MA, United States
[7]Biostatistics Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

**Corresponding Author:**
Anna L Beukenhorst, MSc, PhD
Department of Biostatistics
Harvard T.H. Chan School of Public Health
4th Floor
677 Huntington Avenue
Boston, MA, MA 02115
United States
Phone: 1 (617) 4951000
Email: beuk@hsph.harvard.edu

## Abstract

**Background:** Smartphone studies provide an opportunity to collect frequent data at a low burden on participants. Therefore, smartphones may enable data collection from people with progressive neurodegenerative diseases such as amyotrophic lateral sclerosis at high frequencies for a long duration. However, the progressive decline in patients' cognitive and functional abilities could also hamper the feasibility of collecting patient-reported outcomes, audio recordings, and location data in the long term.

**Objective:** The aim of this study is to investigate the completeness of survey data, audio recordings, and passively collected location data from 3 smartphone-based studies of people with amyotrophic lateral sclerosis.

**Methods:** We analyzed data completeness in three studies: 2 observational cohort studies (*study 1*: N=22; duration=12 weeks and *study 2*: N=49; duration=52 weeks) and 1 clinical trial (*study 3*: N=49; duration=20 weeks). In these studies, participants were asked to complete weekly surveys; weekly audio recordings; and in the background, the app collected sensor data, including location data. For each of the three studies and each of the three data streams, we estimated time-to-discontinuation using the Kaplan–Meier method. We identified predictors of app discontinuation using Cox proportional hazards regression analysis. We quantified data completeness for both early dropouts and participants who remained engaged for longer.

**Results:** Time-to-discontinuation was shortest in the year-long observational study and longest in the clinical trial. After 3 months in the study, most participants still completed surveys and audio recordings: 77% (17/22) in study 1, 59% (29/49) in study 2, and 96% (22/23) in study 3. After 3 months, passively collected location data were collected for 95% (21/22), 86% (42/49), and 100% (23/23) of the participants. The Cox regression did not provide evidence that demographic characteristics or disease severity at baseline were associated with attrition, although it was somewhat underpowered. The mean data completeness was the highest for passively collected location data. For most participants, data completeness declined over time; mean data completeness was typically lower in the month before participants dropped out. Moreover, data completeness was lower for

people who dropped out in the first study month (very few data points) compared with participants who adhered long term (data completeness fluctuating around 75%).

**Conclusions:** These three studies successfully collected smartphone data longitudinally from a neurodegenerative population. Despite patients' progressive physical and cognitive decline, time-to-discontinuation was higher than in typical smartphone studies. Our study provides an important benchmark for participant engagement in a neurodegenerative population. To increase data completeness, collecting passive data (such as location data) and identifying participants who are likely to adhere during the initial phase of a study can be useful.

**Trial Registration:** ClinicalTrials.gov NCT03168711; https://clinicaltrials.gov/ct2/show/NCT03168711

***(JMIR Mhealth Uhealth 2022;10(2):e31877)*** doi:10.2196/31877

**KEYWORDS**

digital phenotyping; mobile health; trial; smartphones; attrition; mobile phone

## Introduction

### Background

Participation in clinical research requires an effort. More research visits create a greater burden and, thus, a larger barrier to long-term participation. Clinical trialists often design studies that collect data relatively infrequently to reduce the burden of clinical research for participants. Although this reduces the research burden, it also reduces statistical power [1]. Data collection from participants' smartphones may allow high-frequency data at a low burden of participation. As smartphones are now increasingly common and typically carried by their users throughout the day, every day, they can be used for nearly continuous, unobtrusive data collection in everyday settings [2,3].

This opportunity to collect research data frequently at a low participant burden is appealing for research on neurodegenerative diseases. For participants, clinic visits are especially onerous, owing to the progressive decline in their cognitive and physical function. Study teams also feel the burden on staff time, as assessment visits often are 1-3 hours in duration. Research sponsors see ballooning costs from staffing requirements [4].

Digital data collection from smartphones could reduce all of these burdens while providing relevant, quantitative, and frequent study data directly from participants. Smartphones can be used to collect a rich variety of data for clinical research. These data include *active data*, which requires data entry by the participant (eg, surveys), and *passive data* (eg, sensor and log data) that do not require activity by the participant beyond installing a research app [5]. These voluminous passive data can be converted into meaningful and interpretable variables that describe individual-level traits, habits, and behavior. If a study makes use of a participant's own phones (therefore enabling collection of naturalistic data without requiring additional instrumentation) and collects raw high-throughput data from the phones (therefore enabling generation of study specific metrics over prepackaged metrics with enhanced reproducibility), the approach is referred to as *digital phenotyping* [6].

### Smartphone Studies in Neurodegenerative Diseases

In many cases, people with neurodegenerative diseases remain able to use their smartphones to participate in studies, despite the progressive nature of their disease. This is certainly true for people with amyotrophic lateral sclerosis (ALS), a neurodegenerative disease that causes a progressive decline in speech, respiratory function, and motor skills [7]. Recent studies have demonstrated that people with ALS use smartphones and can complete frequent surveys for research, even in the later stages of the disease [8]. Thus, smartphone-based digital phenotyping for neurodegenerative diseases is feasible.

At the same time, digital data collection has potential shortcomings that must be understood. Despite the low burden of data collection from participants' own devices, smartphone studies may have high attrition, even when focusing on passive data collection [9-12]. When participants discontinue app use, they introduce missing active data. In addition, sensor noncollection due to technological factors or participant behavior introduces missing passive data [11,12]. Missing data, whether active or passive, reduces statistical power, threatens the generalizability of results, and can introduce attrition bias [9,12,13]. For example, if participants with more severe disease at baseline dropout more frequently, the study findings may not generalize to these participants [13].

To assess the risk of attrition bias in smartphone-based medical research, we must understand the relationship between participant characteristics, disease severity, and rate of progression on the one hand, as well as attrition risk on the other hand [9]. Patterns of attrition may differ between observational studies and clinical trials [14]. Attrition has been reported for smartphone studies in some areas, including mental health [9,15,16], cancer [17], chronic diseases [18], neurodegenerative diseases [19,20] and healthy controls [19]. However, predictors of attrition or risk of attrition bias have not been thoroughly investigated for neurodegenerative diseases. In people with diseases such as ALS, immobility, challenges with activities of daily living, and cognitive decline can threaten their ability to comply with smartphone data collection.

### Study Aims

We investigated data completeness in 3 studies using the same platform for data collection from personal smartphones of people with ALS. Two of these were observational studies, and one

was a clinical trial. In all 3 studies, the participants contributed traditional ALS clinical outcome data during in-clinic visits. In addition, participants installed the front-end app from the Beiwe platform on their smartphones and used it for active and passive data collection [8]. We estimated time-to-discontinuation in each of the three studies, identified predictors of app discontinuation, and quantified data completeness for early dropouts and participants who remained engaged longitudinally.

## Methods

### Overview

We analyzed data from 3 studies using the Beiwe platform for smartphone data collection. In each of the three studies, data were collected in three ways: (1) traditional clinician-administered survey data during clinic visits or by telephone; (2) active data, including patient-reported outcomes from smartphone surveys administered through the Beiwe app, audio recordings where participants coughed, and audio recordings where participants recited a text shown on their smartphone screen; and (3) passive data from sensors and logs, automatically collected by the Beiwe smartphone app.

All data were collected and stored in compliance with local, state, and national laws, regulations, and policies. For study 1, participants were enrolled at the Massachusetts General Hospital (MGH) in Boston. For study 2, participants were enrolled at both MGH in Boston, United States, and in Washington University in St Louis, Missouri, United States. For study 3, participants were enrolled at MGH, Twin Cities ALS Clinic in Minneapolis, Minnesota, United States, and Holy Cross ALS Clinic in Fort Lauderdale, Florida, United States. The studies differed in duration and expected frequency of clinical data collection (Table 1). None of the studies included routine contact with participants to encourage engagement; there was no reimbursement for engagement; and outside of reminders from the smartphone app itself (known as notifications), no reminders were sent to participants.

**Table 1.** Characteristics of the 3 included studies.

| Study | Number of participants, N | Study duration (weeks) | Frequency of data collection | | |
|---|---|---|---|---|---|
| | | | Clinic visit | Smartphone survey | Smartphone sensors |
| Study 1 | 22 | 12 | 3 times | Weekly | GPS on for 1 minute and off for 10 minutes |
| Study 2 | 49 | 52 | 2 times | Weekly | GPS on for 1 minute and off for 10 minutes |
| Study 3 | 23 | 20 | 3 times | Weekly | GPS on for 1 minute and off for 10 minutes |

### Study 1: 12-Week Pilot Study

Study 1 was a pilot observational cohort study 12 weeks in duration, running from July 2016 to June 2018. Clinician-administered survey data were collected at baseline and at weeks 6 and 12. Study design and participant recruitment for study 1 have been previously published [8].

### Study 2: 52-Week Cohort Study

Study 2 was an observational cohort study 52 weeks in duration, running from November 2018 to March 2021. Clinician-administered survey data were collected at baseline and week 52. This study used the same methods for recruitment and data collection used in study 1 [8].

### Study 3: 20-Week Clinical Trial

Study 3 was the safety of rate elevation in ALS (SURE-ALS2) randomized, placebo-controlled clinical trial of inosine to raise urate levels (NCT03168711). The trial ran from November 2017 to December 2019. The participants were divided into an intervention group, receiving inosine and a control group, receiving matching placebo until week 16. In short, after consent and successful screening for the trial, the Beiwe smartphone app was installed on the participants' personal smartphones at the baseline visit. The app was uninstalled at the 20-week visit. Participants were asked to complete in-person visits for clinical outcomes at baseline, week 12, and week 20. They also received phone calls every 3 weeks throughout the study. The clinician-administered revised ALS functional rating scale (ALSFRS-R) was completed during in-person visits.

Study 3 had more restrictive selection criteria than the observational studies. Studies 1 and 2 required participants to have a diagnosis of ALS according to the El Escorial Criteria [21], at least moderate smartphone use, and no neurological disorders other than ALS. Study 3 included additional selection criteria requiring vital capacity >60% of predicted, plasma urate <5.5 mg/dL, and no medical history of gout, coronary artery disease, stroke, poorly controlled hypertension, or renal insufficiency.

### Ethics

Each study was approved by the Mass General Brigham Institutional Review Board (IRB). Study 2 was also approved by the Washington University IRB. Study 3 used a central IRB (the Mass General Brigham IRB) for all sites.

### Data Collection

We collected smartphone data through Beiwe, an open-source, end-to-end encrypted high-throughput digital phenotyping platform [22]. It consists of Android and iOS smartphone apps for data collection and an Amazon Web Services cloud-based system back-end for data collection and processing [23]. It has been used in both observational studies and clinical trials to collect self-administered surveys and various types of passive data [8,24].

The primary clinical outcome measure in the 3 studies was functional ability, as measured by the ALSFRS-R. The ALSFRS-R is a 12-item survey for measuring functional ability, each with 5-answer options, scored from 4 (*normal ability*) to 0 (*lowest functionality*) [25]. The questions are divided into four subdomains: the bulbar domain (questions 1-3), fine motor domain (questions 4-6), gross motor domain (questions 7-9), and respiratory domain (questions 10-12). The maximum domain score is 12 (*domain not affected*), with a lower score denoting lower functional ability. The total survey score is the sum of all questions and has a 48-point scale (from 48 points, indicating *normal function* to 0 points [25]).

### Baseline Visit

At the baseline visit, clinical characteristics were obtained in person and stored in an electronic data capture system. The Beiwe app was downloaded onto the participant's smartphone and activated by the study coordinator. Upon activation, the app delivered a baseline survey to the participant to gather demographic and clinical information and thereafter collected active and passive data as planned.

### Smartphone Data

The Beiwe app was configured to collect weekly self-administered ALSFRS-R scores, weekly recordings of speech, and weekly recordings of cough (not analyzed here). The app also collected metadata on survey completion, including clock-times of survey presentation on the smartphone screen, submission time of each question answer by the participant, and submission time of the completed survey. In addition, the app collected data from multiple smartphone sensors and logs (Table 1) [8,26]. GPS data were collected for a 1-minute interval followed by a 10-minute interval of noncollection, that is, GPS data were collected for 1 minute every 11 minutes (hence, approximately 6 times per hour).

### Statistical Methods

#### *Data Volume*

For clinic-based and smartphone-based surveys, we reported the number of clinic-based and smartphone-based ALSFRS-R surveys per participant. For GPS location data, we reported the number of participant-days for which data were available and the total data volume. Data were considered available if the app had recoded any location data on that day.

#### *Kaplan–Meier Estimates of Time-to-Discontinuation*

For smartphone survey data, we defined the date of dropout as the date of the first missed survey, that is, the week after a participant had completed their last smartphone survey. For smartphone sensor data, we defined the date of dropout as the day after the last recording of smartphone sensor data.

We used the Kaplan–Meier method to estimate time-to-discontinuation for smartphone survey data (model 1) and for smartphone sensor data (model 2). Both models were stratified by study type. Time-to-discontinuation was censored at the end of each study's follow-up period (after week 12, 52, or 20; see Table 1) if a participant died and, for trial participants, if they discontinued the trial because of side effects.

#### *Proportional Hazard of Dropping Out*

We used Cox proportional hazard regression to identify the predictors of smartphone survey data and smartphone sensor data. We tested whether the likelihood of dropout was higher for participants with certain demographic characteristics or with a higher disease severity at baseline.

The covariates we included in the model were as follows:

- Participant characteristics, such as age (in years), sex (male or female), and smartphone operating system (Android or iOS)
- Disease severity at baseline, such as baseline functional ability as measured by the 4 domains of the ALSFRS-R score. These four domains are the fine motor domain score, gross motor domain score, bulbar domain score, and respiratory domain score.

#### *Data Completeness*

Data completeness was defined as the percentage of days for which participants provided GPS data and the percentage of weeks for which participants submitted surveys and audio recordings. For GPS data, 100% data completeness meant that GPS data were available for each day from the participants' enrollment until their last day in the study. For survey and audio recording data, 100% data completeness meant that a participant had submitted 1 survey or audio recording per week for each week from their enrollment until their last day in the study.

We visualized the data in a boxplot of participants' data completeness during the time they contributed to the data. In addition, we calculated the data completeness for each 28-day period in which a participant was in the study. We used a 28-day period rather than a calendar month, because the total duration of all 3 studies was a multitude of 28 days; the maximum time in study was 3×28 days for study 1, 12×28 days for study 2, and 5×28 days for study 3. We plotted data completeness for each 28-day period (hereafter, *month*), stratified by participants' total duration in the study.

## Results

The 3 studies are referred to as *study 1* (12-week observational pilot study), *study 2* (52-week observational study) and *study 3* (20-week clinical trial) in this section.

### Participants

Demographic data for the 3 studies are presented in Table 2. There were 22 participants in study 1, 49 in study 2, and 23 in study 3. There were more male participants in the 2 observational studies (15/22, 68% and 29/49, 59%; in line with a higher prevalence of ALS in men), but fewer in the clinical trial (9/23, 39%). The mean age and baseline ALSFRS-R scores were similar across studies. Owing to the differences in inclusion criteria, mean disease duration was 5-7 months shorter for participants in the clinical trial, and baseline mean vital capacity, a measurement of lung volume, was higher for those in the clinical trial. Most participants were iOS users in all studies (56/94, 60%). In each of the three studies, one person died before the end of the study.

In total, we collected 185 ALSFRS-R scores during clinic visits (43 for study 1, 77 for study 2, and 65 for study 3), 1465 ALSFRS-R scores from smartphones (375 for study 1, 759 for study 2, and 331 for study 3), 3748 audio recordings from smartphones (678 for study 1, 1315 for study 2, and 609 for study 3), and a total of 10.4 GB of GPS location data (3.4 GB for study 1, 5.5 for study 2, and 1.5 GB for the study 3).

**Table 2.** Demographic characteristics of participants per study.

| Characteristics | Study 1 | Study 2 | Study 3 |
|---|---|---|---|
| Number of participants, N | 22 | 49 | 23 |
| Sex (male), n (%) | 15 (68) | 30 (59) | 9 (39) |
| Race (White), n (%) | 20 (91) | 48 (98) | 23 (100) |
| Phone operating system (iOS users), n (%) | 17 (77) | 36 (73) | 12 (52) |
| **Location of symptom onset, n (%)** | 21 (100) | 49 (100) | 23 (100) |
| Bulbar | 5 (23) | 11 (22) | 7 (30) |
| Limb | 16 (73) | 38 (78) | 15 (65) |
| Trunk | 1 (5) | 0 | 1 (4) |
| Age (years), mean (SD) | 56 (6) | 57 (11) | 58 (10) |
| Disease duration at baseline visit (months), mean (SD) | 31 (21) | 35 (23; n=48)[a] | 26 (14; n=22)[a] |
| Diagnostic delay[b] (months), mean (SD) | 17 (13) | 17 (14) | 12 (7; n=22)[a] |
| **Baseline ALSFRS-R[c] total score, mean (SD)** | 34 (7) | 35 (9; n=46)[a] | 36 (8) |
| Bulbar subscore | 10 (2) | 10 (3) | 9 (3) |
| Fine motor subscore | 8 (2) | 8 (3) | 8 (3) |
| Gross motor subscore | 7 (3) | 7 (3) | 8 (3) |
| Respiratory subscore | 9 (3) | 10 (2) | 11 (2) |

[a]Data were missing; mean and SD calculated over smaller sample size (smaller sample size provided as n, wherever applicable).

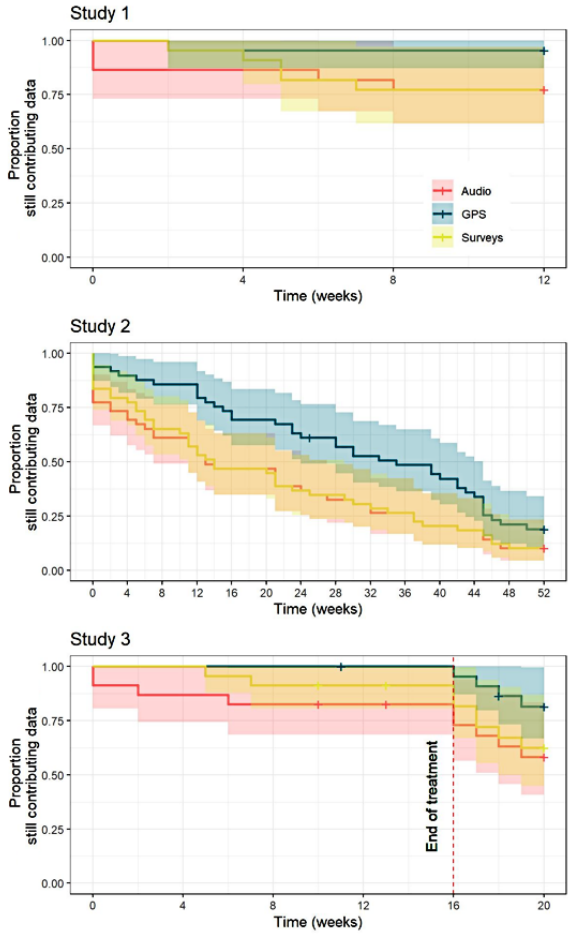
[b]Diagnostic delay: time between symptom onset and diagnosis.

[c]ALSFRS-R: revised amyotrophic lateral sclerosis functional rating scale.

## Time-to-Discontinuation

Kaplan–Meier estimates of the time-to-discontinuation for active data are presented in Figure 1 (red line for audio recordings, yellow line for surveys, and blue line for passive data). After 12 weeks, 77% (17/22) of the participants in study 1, 59% (29/49) of the participants in study 2, and 96% (22/23) of the participants in study 3 continued to contribute active data (surveys and audio recordings). For passive data, 95% (21/22) of the participants in study 1, 86% (42/49) of the participants in study 2, and 100% (23/23) of the participants in study 3 continued to contribute sensor data.

**Figure 1.** Kaplan–Meier plot estimates of time-to-discontinuation for 3 data types. Each color denotes a different data type: audio data in red, GPS data in blue, and survey data in yellow. Participants that were censored before the end of the study are denoted by + signs. Each panel shows time-to-discontinuation in a different study: study 1 (top, a 12-week pilot study), study 3 (middle, a 20-week clinical trial), and study 2 (bottom, a 1-year observational study).




## Predictors of Early Discontinuation

We used the Cox proportional hazards model to estimate whether study, participant demographics, and disease severity were associated with the risk of discontinuation, with a separate model for survey, audio recording, and GPS data. None of the variables were statistically significantly associated with the risk of discontinuation. The estimated associations between the study, participant demographics, and disease severity at baseline are presented in Multimedia Appendix 1.

## Data Completeness

The time-to-discontinuation model described above paints only part of the picture—how long the participants contributed to *any* data. We also explored data completeness, the proportion of days a participant provided GPS data, surveys, audio recordings of coughs, and audio recordings when participants recited a short text that was displayed on their screen. Data were 100% complete for GPS if any data were contributed for a given day, and data were 100% complete if surveys and audio recordings were completed each week when the task was presented.

Figure 2 shows boxplots for each study of the average data completeness for each data type before discontinuation (after discontinuation, data completeness is 0% by definition). In all studies, GPS data completeness was highest over the 3 studies (range 90%-100% of days that a participant stayed in the study), followed by survey data in study 1 (median 100%) and study 2 (median 90%) and by audio recordings (cough recordings; median 92%) in study 3. Of the 3 studies, study 3 (20-week clinical trial) had the highest data completeness, and study 2 (52-week observational study) had the lowest data completeness.

We then plotted data completeness for each 28-day period (hereafter, *month*), stratified by participants' total duration in the study (Figure 3). First, this analysis showed that participants who contributed data for longer (eg, for >2 months) had higher data completeness than participants who stopped contributing data in the first or second month. Participants who contributed data longer had a data completeness fluctuating around 75% for all the data types for their first months in the study, whereas early dropouts typically had low data completeness for the months that they were active. For those dropping out within the first month, mean data completeness across the studies ranged from 7.8% (audio recording) to 41% (surveys). For those who dropped out in the second study month, completeness ranged from 41% (audio cough recording) to 59% (GPS data). Second, for most participants, data completeness declined over time; mean data completeness was typically lower in the last month of the study (except for study 2, where participants who were active until the final month completed all surveys and audio

tasks). Third, data completeness was generally highest for GPS data (except for those who dropped out in the first month). Fourth, data completeness did not differ significantly between the 2 audio tasks.

**Figure 2.** Boxplot of participants' data completeness (in %) excluding the period after discontinuation. Data completeness was defined as percentage of days with any GPS data and percentage of weeks with a completed survey or audio recording.

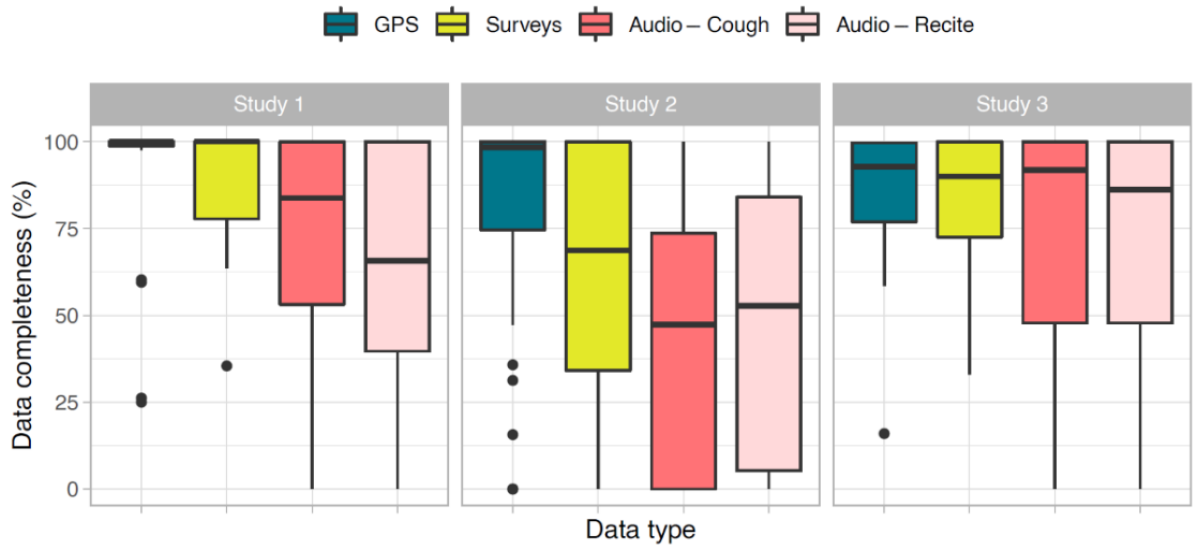
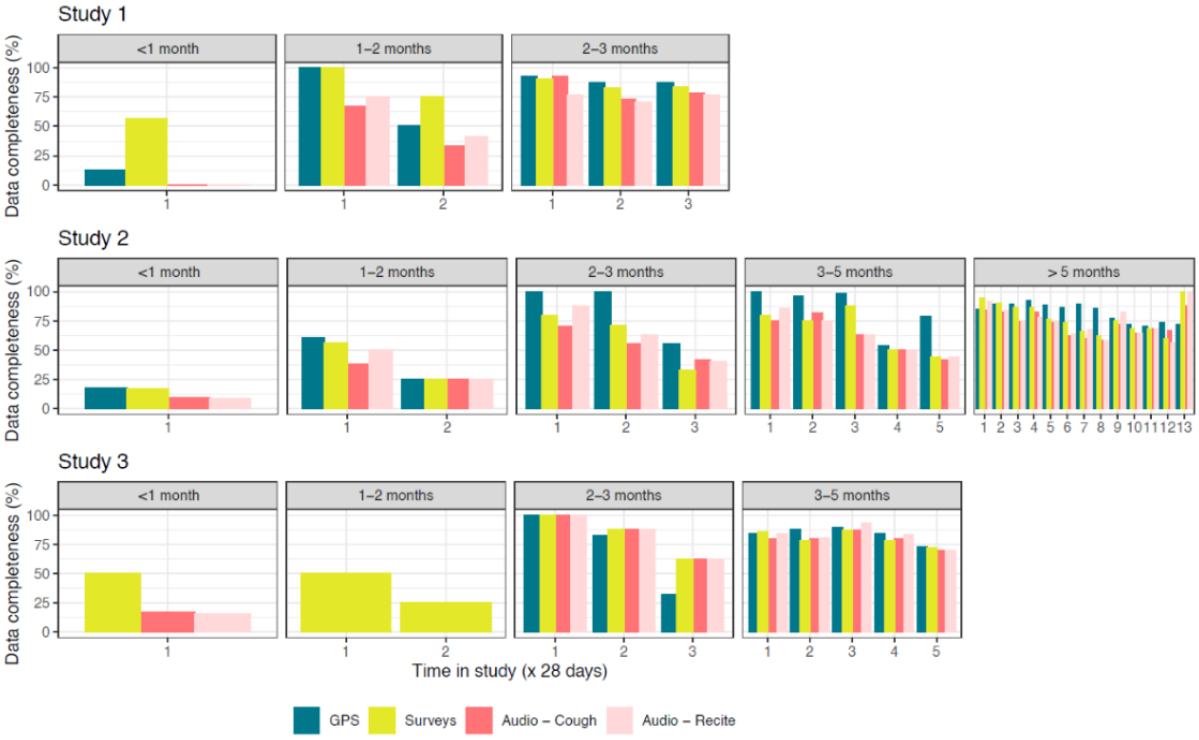



**Figure 3.** Bar graph of data completeness per month in study (excluding the period after discontinuation), stratified by time-to-discontinuation of the participant (gray bar indicates time-to-discontinuation). Number of participants for each panel from left to right are as follows: N=7, 4, and 18 for study 1; N=20, 6, 10, 8, and 33 for study 2; and N=5, 1, 2, and 22 for study 3. Data completeness was defined as percentage of days with any GPS data and percentage of weeks with a completed survey or audio recording.




## Discussion

### Principal Findings

In this study, we showed that smartphones can be used to collect frequent active and passive data from people with neurodegenerative diseases, specifically ALS, both in observational studies and in a clinical trial setting. Participant engagement, as measured by time-to-discontinuation, was higher than that in published data [9,10]. The two observational studies described in this paper, in which no adherence reinforcement or incentives were implemented, provide an important benchmark for participant engagement with a smartphone app in research.

Data completeness was lower for active data than for passive data. In other words, smartphones continued to collect passive

data even after participants had stopped completing surveys or recording audio.

### Understanding Participant Adherence

Time-to-discontinuation was higher in our studies compared with smartphone data collection studies in other domains, which often show an exponential dropout [9,10,27]. The lower dropout rate in our study may highlight the high commitment to research of people with ALS and neurodegenerative diseases despite the challenges of progressive functional and cognitive decline. In another smartphone study, participants with multiple sclerosis dropped out significantly later than healthy controls, who only remained active for 1 day [19]. In addition, time-to-discontinuation was shorter for clinic-referred participants than for self-referred participants (7 days vs 25.5 days) and shorter than in our study.

Our analysis did not provide evidence that demographic characteristics or disease severity at baseline were associated with attrition, although our analysis was underpowered to detect predictors of attrition.

## Strengths and Limitations

ALS is a rare disease, and our analysis of 3 studies, both in observational and interventional research contexts, is the first of its kind. Given that sample sizes were limited by the low prevalence of ALS, we were underpowered to detect associations between participant and disease characteristics and adherence to digital data collection. Furthermore, although neurodegenerative diseases share many characteristics, our results may not be generalizable to all neurodegenerative diseases.

## Improving Participant Adherence

Despite better than expected adherence compared with published studies, boosting adherence remains important, especially for clinical trials using smartphone-based outcomes. Participants of digital health studies are more likely to actively engage long term if they see the value of participation [14,27], which may have been the case, especially in the trial participants who received a novel therapeutic. Personal contact with study personnel helps participants feel valued and is a major driver of engagement [27]. Both perceived value and personal contact with study personnel may explain the better participant adherence in study 3, which had almost full adherence until the end of treatment with the study drug.

In future studies, we will test whether reminder phone calls, more frequent clinic visits, or financial incentives can improve adherence, particularly in longer studies. Another potential motivator for adherence could be allowing participants to view their data, including previous survey responses [28]. However, this may not always be scientifically advisable, as it may influence participants' responses through the Hawthorne effect and related forms of reporting bias [29].

Data completeness was higher and attrition was lower for passive data than for active data. Passive data incompleteness is due to both behavioral factors (eg, a participant disabling GPS) or technological issues (eg, smartphone blocking sensor data collection) [12,30,31]. Investigators familiar with passive smartphone data collection recognize that both commonly used smartphone operating systems (Android and iOS) implement power saving measures for apps running in the background to reduce consumption of the central processing unit resources, memory, and battery [12,31]. This means that no app can run in the background mode indefinitely, but instead the app needs to be brought to the foreground at least occasionally for the background data collection to persist [30]. Therefore, longitudinal passive data collection without active data collection is not possible. Factors such as device type, hardware, and operating system influence data completeness [30]. These technological factors can be difficult to modify, and they also change over time.

## Identifying High Adherence: Run-in and Withdrawal Design

Our analyses showed that participants in the clinical trial adhered best to the study regimen. When treatment ended, >80% were still answering surveys, and all eligible participants were still contributing sensor data. Nevertheless, for clinical trials, it could be useful to identify participants who are more likely to adhere. For studies requiring participants to use smartphones, especially trials, a *run-in and withdrawal design* has been suggested [9]. With this design, participants enter a *weed-out* period after enrollment. Only participants who still used the study app after the weed-out period were randomized. Our study showed that participants who stopped contributing surveys within 1 or 2 months of enrolling had lower data completeness than their engaged counterparts. This suggests that monitoring active data completeness during a screening period for a trial could help identify participants who are more likely to adhere.

## Conclusions

Our study demonstrates that it is possible to collect longitudinal research data from people with progressive neurodegenerative diseases using their personal smartphones. Our results are especially promising for clinical trials (longer time-to-discontinuation than in observational studies) and for studies collecting mainly passive data with a light active data component (higher data completeness and longer time-to-discontinuation than in studies prioritizing survey data). We identified putative predictors of dropout, which can be confirmed in future studies, and will allow researchers to target efforts to improve participant adherence to smartphone data collection.

## Acknowledgments

The authors thank the Salah Foundation for their philanthropic support. The authors thank all the study participants for their dedication in moving the amyotrophic lateral sclerosis research forward. This study was funded in part by federal funding (R15 DC018944).

## Conflicts of Interest

JPO is a cofounder and board member of a recently established commercial entity that operates in digital phenotyping. SP reports research grants from Amylyx Therapeutics, Revalesio Corporation, UCB/Ra Pharma, Biohaven, Clene, Prilenia, Seelos, The ALS Association, the American Academy of Neurology, ALS Finding a Cure, the Salah Foundation, the Spastic Paraplegia Foundation, the Muscular Dystrophy Association and reports personal consulting fees for advisory panels from Orion, Medscape and Cytokinetics. JDB reports equity in REACTNeuro. He reports consulting fees from Clene Nanomedicine, Biogen, Janssen, Sawai Pharmaceuticals, MT Pharma of America, MT Pharma Holdings of America. He reports research support from Amylyx, Alexion, Biogen, MT Pharma of America, Anelixis Therapeutics, Brainstorm Cell Therapeutics, nQ Medical, RAPA Therapeutics, NINDS, Muscular Dystrophy Association, ALS One, ALS Association, and ALS Finding A Cure. TMM reports licensing agreements with C2N and Ionis Pharmaceuticals, has served on advisory boards for Biogen and UCB Pharma, and is a consultant for Cytokinetics and Disarm Therapeutics.

**Multimedia Appendix 1**
Results of Cox models.
[DOCX File , 18 KB - mhealth_v10i2e31877_app1.docx ]

## References


1. Rothman K, Greenland S, Lash T. Modern Epidemiology, Third Edition. Pennsylvania, United States: Lippincott Williams and Wilkins; 2014.
2. Incel OD, Kose M, Ersoy C. A review and taxonomy of activity recognition on mobile phones. BioNanoSci 2013 May 11;3(2):145-171. [doi: 10.1007/s12668-013-0088-3]
3. Amor J, James C. Setting the scene: mobile and wearable technology for managing healthcare and wellbeing. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug. 25-29, 2015; Milan, Italy p. 7752-7755. [doi: 10.1109/embc.2015.7320189]
4. Pham L, Harris T, Varosanec M, Morgan V, Kosa P, Bielekova B. Smartphone-based symbol-digit modalities test reliably captures brain damage in multiple sclerosis. NPJ Digit Med 2021 Feb 24;4(1):36 [FREE Full text] [doi: 10.1038/s41746-021-00401-y] [Medline: 33627777]
5. Kourtis L, Regele O, Wright J, Jones G. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. NPJ Digit Med 2019;2:1-9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
6. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Dec;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
7. Brooks BR. Natural history of ALS: symptoms, strength, pulmonary function, and disability. Neurology 1996 Oct 01;47(4 Suppl 2):71-81. [doi: 10.1212/wnl.47.4_suppl_2.71s] [Medline: 8858055]
8. Berry JD, Paganoni S, Carlson K, Burke K, Weber H, Staples P, et al. Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. Ann Clin Transl Neurol 2019 May;6(5):873-881 [FREE Full text] [doi: 10.1002/acn3.770] [Medline: 31139685]
9. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
10. Dorsey ER, Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The use of smartphones for health research. Acad Med 2017 Dec;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
11. Kiang MV, Chen J, Krieger N, Buckee C, Alexander M, Baker J, et al. Sociodemographic characteristics of missing data in digital phenotyping. Sci Rep 2021 Jul 29;11(1):15408 [FREE Full text] [doi: 10.1038/s41598-021-94516-7] [Medline: 34326370]
12. Beukenhorst AL, Sergeant JC, Schultz DM, McBeth J, Yimer BB, Dixon WG. Understanding the predictors of missing location data to inform smartphone study design: observational study. JMIR Mhealth Uhealth 2021 Nov 16;9(11):e28857 [FREE Full text] [doi: 10.2196/28857] [Medline: 34783661]
13. Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. Br Med J 2012 May 18;344:e2809. [doi: 10.1136/bmj.e2809] [Medline: 22611167]
14. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
15. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]

16. 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]
17. Benze G, Nauck F, Alt-Epping B, Gianni G, Bauknecht T, Ettl J, et al. PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. Ann Palliat Med 2017 Oct 26:104-111 [FREE Full text] [doi: 10.21037/apm.2017.07.05] [Medline: 29156896]
18. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
19. Pratap A, Grant D, Vegesna A, Tummalacherla M, Cohan S, Deshpande C, et al. Evaluating the utility of smartphone-based sensor assessments in persons with multiple sclerosis in the real-world using an app (elevateMS): observational, prospective pilot digital health study. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e22108 [FREE Full text] [doi: 10.2196/22108] [Medline: 33107827]
20. Trister AD, Neto EC, Bot BM, Perumal T, Pratap A, Klein A, et al. mPower: a smartphone-based study of Parkinson's disease provides personalized measures of disease impact. In: Proceedings of the 2016 International Congress. 2016 Presented at: 2016 International Congress; June 19-23, 2016; Berlin, Germany URL: https://www.mdsabstracts.org/abstract/mpower-a-smartphone-based-study-of-parkinsons-disease-provides-personalized-measures-of-disease-impact/
21. Brooks BR, Miller RG, Swash M, Munsat TL, World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000 Dec 10;1(5):293-299. [doi: 10.1080/146608200300079536] [Medline: 11464847]
22. Onnela J, Dixon C, Griffin K, Jaenicke T, Minowada L, Esterkin S, et al. Beiwe: a data collection platform for high-throughput digital phenotyping. J Open Source Softw 2021 Dec;6(68):3417. [doi: 10.21105/joss.03417]
23. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
24. Beukenhorst AL, Collins E, Burke KM, Rahman SM, Clapp M, Konanki SC, et al. Smartphone data during the COVID-19 pandemic can quantify behavioral changes in people with ALS. Muscle Nerve 2021 Feb;63(2):258-262 [FREE Full text] [doi: 10.1002/mus.27110] [Medline: 33118628]
25. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. J Neurolog Sci 1999 Oct;169(1-2):13-21. [doi: 10.1016/s0022-510x(99)00210-5]
26. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
27. Druce K, Dixon W, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
28. Beukenhorst AL, Howells K, Cook L, McBeth J, O'Neill TW, Parkes MJ, et al. Engagement and participant experiences with consumer smartwatches for health research: longitudinal, observational feasibility study. JMIR Mhealth Uhealth 2020 Jan 29;8(1):e14368 [FREE Full text] [doi: 10.2196/14368] [Medline: 32012078]
29. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
30. Beukenhorst AL, Schultz DM, McBeth J, Lakshminarayana R, Sergeant JC, Dixon WG. Using smartphones for research outside clinical settings: how operating systems, app developers, and users determine geolocation data quality in mHealth studies. Stud Health Technol Inform 2017;245:10-14 [FREE Full text] [Medline: 29295042]
31. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan 17;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]

## Abbreviations

**ALS:** amyotrophic lateral sclerosis
**ALSFRS-R:** revised amyotrophic lateral sclerosis functional rating scale
**IRB:** Institutional Review Board
**MGH:** Massachusetts General Hospital

*Edited by L Buis; submitted 11.07.21; peer-reviewed by R Almeida, M Stach, R Marshall; comments to author 06.11.21; revised version received 10.11.21; accepted 14.12.21; published 04.02.22.*

*Please cite as:*
*Beukenhorst AL, Burke KM, Scheier Z, Miller TM, Paganoni S, Keegan M, Collins E, Connaghan KP, Tay A, Chan J, Berry JD, Onnela JP*
*Using Smartphones to Reduce Research Burden in a Neurodegenerative Population and Assessing Participant Adherence: A Randomized Clinical Trial and Two Observational Studies*
*JMIR Mhealth Uhealth 2022;10(2):e31877*
*URL: https://mhealth.jmir.org/2022/2/e31877*
*doi:10.2196/31877*
*PMID:35119373*

Original Paper

# Defining the Enablers and Barriers to the Implementation of Large-scale, Health Care–Related Mobile Technology: Qualitative Case Study in a Tertiary Hospital Setting

Ravi Aggarwal[1,2], MBBS; Sheena Visram[3], BSc, MClinRes; Guy Martin[1,2], MBBS, PhD; Viknesh Sounderajah[1,2], MBBS; Sanjay Gautama[4], MBBS; Kevin Jarrold[4], BA, MSc; Robert Klaber[4], MBBS, MD; Shona Maxwell[4]; John Neal[4]; Jack Pegg[4], BA; Julian Redhead[4], MBBS; Dominic King[1], MBBS, PhD; Hutan Ashrafian[1,2], MBBS, PhD; Ara Darzi[1,2], MD

[1]Institute of Global Health Innovation, Imperial College London, London, United Kingdom
[2]Department of Surgery & Cancer, Imperial College London, London, United Kingdom
[3]Department of Computer Science, University College London, London, United Kingdom
[4]Imperial College Healthcare NHS Trust, London, United Kingdom

**Corresponding Author:**
Hutan Ashrafian, MBBS, PhD
Institute of Global Health Innovation
Imperial College London
10th Floor, Queen Elizabeth Queen Mother Building
St Marys Hospital, Praed Street
London, W2 1NY
United Kingdom
Phone: 44 7799871597
Email: h.ashrafian@imperial.ac.uk

## Abstract

**Background:** The successful implementation of clinical smartphone apps in hospital settings requires close collaboration with industry partners. A large-scale, hospital-wide implementation of a clinical mobile app for health care professionals developed in partnership with Google Health and academic partners was deployed on a bring-your-own-device basis using mobile device management at our UK academic hospital. As this was the first large-scale implementation of this type of innovation in the UK health system, important insights and lessons learned from the deployment may be useful to other organizations considering implementing similar technology in partnership with commercial companies.

**Objective:** The aims of this study are to define the key enablers and barriers and to propose a *road map* for the implementation of a hospital-wide clinical mobile app developed in collaboration with an industry partner as a data processor and an academic partner for independent evaluation.

**Methods:** Semistructured interviews were conducted with high-level stakeholders from industry, academia, and health care providers who had instrumental roles in the implementation of the app at our hospital. The interviews explored the participants' views on the enablers and barriers to the implementation process. The interviews were analyzed using a broadly deductive approach to thematic analysis.

**Results:** In total, 14 participants were interviewed. Key enablers identified were the establishment of a steering committee with high-level clinical involvement, well-defined roles and responsibilities between partners, effective communication strategies with end users, safe information governance precautions, and increased patient engagement and transparency. Barriers identified were the lack of dedicated resources for mobile change at our hospital, risk aversion, unclear strategy and regulation, and the implications of bring-your-own-device and mobile device management policies. The key lessons learned from the deployment process were highlighted, and a road map for the implementation of large-scale clinical mobile apps in hospital settings was proposed.

**Conclusions:** Despite partnering with one of the world's biggest technology companies, the cultural and technological change required for mobile working and implementation in health care was found to be a significant challenge. With an increasing requirement for health care organizations to partner with industry for advanced mobile technologies, the lessons learned from

our implementation can influence how other health care organizations undertake a similar mobile change and improve the chances of successful widespread mobile transformation.

***(JMIR Mhealth Uhealth 2022;10(2):e31497)*** doi:10.2196/31497

**KEYWORDS**

mHealth; implementation science; mobile technology; mobile apps; clinical applications; smartphone apps; health care industry; stakeholders; mobile phone

## Introduction

### Background

The implementation of mobile technologies in hospital workflows has the potential to significantly improve patient safety, transform health care delivery, and positively affect patient outcomes [1]. Although there is widespread agreement about the importance and potential benefits of mobile technologies to tackle critical challenges in health care, successful implementation of this emerging technology in clinical settings has proven to be challenging [2-4]. In the National Health Service (NHS), the deployment of hospital-wide clinical mobile apps in secondary care is still uncommon despite the ubiquity of smartphone ownership among health care professionals (HCPs) [5,6]. HCPs continue to use their own smartphones in hospitals for daily clinical tasks, including communication among teams, accessing clinical apps such as decision support aids or medical calculators, and educational purposes [7-10]. Although medical apps are freely available for download from *App Stores*, there has been limited deployment of hospital-wide clinical mobile apps for HCPs in the NHS.

Streams (Google Health) is a multifunctional smartphone app displaying a range of patient clinical information that was implemented on a *bring-your-own-device* (BYOD) basis at our hospital (Imperial College Healthcare NHS Trust [ICHNT]). Streams was iteratively developed by a multidisciplinary team of researchers, clinicians, and developers in a tripartite partnership with DeepMind Health and Google Health [11] and Imperial College London (ICL). This partnership was initiated as part of ICHNT's goals to be one of the most digitally mature organizations in the NHS and one of 16 Global Digital Exemplar providers in the United Kingdom.

The ambition was to process and display routinely collected clinical results on the clinicians' own smartphone devices through integration with the hospital's existing information systems and electronic health records (EHRs; Cerner Corporation). Streams was developed for use on iOS devices only as it has been demonstrated that 75.6% of physicians and 58.4% of nurses at our institution use an iOS device [5]. Streams was registered with the Medicines and Healthcare Products Regulatory Agency as a Class I, nonmeasuring, nonsterile medical device under the EU Medical Device Directive (1993). The implementation of the Streams app at our hospital began in early 2019 with a small pilot group of clinicians. Further development of the app and instigation of mobile device management (MDM) software ensued before widespread deployment commenced in January 2020 with the app available to all HCPs across each hospital site within the organization.

Before Streams was implemented at our hospital, a limited version of the app had been deployed in a focused capacity at another London-based hospital network to aid HCPs in the detection and management of acute kidney injury [12]. This deployment attracted significant public and media interest because of an investigation by the Information Commissioner's Office (ICO) into the nature of data processing between the Trust and the industry partner [13-15]. In a separate initiative that learned from these well-publicized issues, a wide-scale implementation of the Streams app was undertaken at our hospital. To the best of our knowledge, this was the first large-scale, hospital-wide deployment of a BYOD clinical smartphone app using MDM technology in the NHS. As such, the implementation process provided important insights into the opportunities and challenges of delivering this type of innovation to the NHS and health systems more widely.

### Objective

In light of previous experiences, the difficulties encountered and the *lessons learned* from this deployment may be generalizable and applicable to other health care organizations that are considering working with industry partners as data processors to deploy similar mobile technology and with academic partners to independently evaluate these interventions. Therefore, the aim of this study is to characterize the key enablers and barriers and to propose a *road map* for the implementation of a hospital-wide clinical mobile app developed in collaboration with an industry partner as a data processor and an academic partner for independent evaluation.

## Methods

### Design

Semistructured interviews were conducted individually with the study participants at a single time point. An in-depth literature review was undertaken to identify implementation and change management frameworks applicable to digital health interventions. These findings led to the creation of a structured topic guide that drew heavily from the *Digital Change in Health and Social Care* document published by The King's Fund in 2018 [16]. This seminal report highlights five key areas to consider when undertaking digital change implementation in health care: leadership and management, user engagement, information governance, partnerships, and resourcing and skills. These areas formed the domains upon which the enablers and barriers were characterized.

### Participants and Sample Size

The participants were purposively recruited [17] following a key informant strategy [18] to ensure that a well-informed,

representative sample of staff members was obtained from all participating stakeholders who were involved in the implementation of the Streams app. These were predominantly members of the Streams Steering Group (SSG), which consists of academics, clinicians, and technicians from ICHNT, the Institute of Global Health Innovation at ICL, and Google Health.

The total number of individuals who were involved in the implementation process was small, which unavoidably restricted the number of interviewees. The sample size was guided by repeated assessments of the emerging data and in line with international consensus guidance and previously published work [19,20]. Although the sample was necessarily heterogeneous to ensure sampling of all the various roles in the steering committee, the wider research team agreed that the data set was adequate for the stated objectives to be met [21]. In total, 16 members of the steering committee were invited to participate in the study by email, with 14 (88%) consenting to be interviewed. All 3 participating organizations were adequately represented. Each participant was interviewed once. The mean duration of each interview was 35.53 (SD 12.36) minutes, and a total of 497 minutes of audio recordings were transcribed for analysis.

### Data Collection and Analysis

All interviews were conducted on the web with the participants over Microsoft Teams videoconferencing software. Audio recordings were made of the interviews, which were then transcribed verbatim. All interviews were conducted by a single male researcher (RA) who is a practicing physician and conducted the study as part of a wider research project. The interviewer took field notes during the interview, which were used to adapt the interview guide depending on the verbal responses given.

A broadly deductive approach to data analysis was used [22], with the topic guide adapted, as noted, from the King's Fund *Digital Change in Health and Social Care* [16] document that formed the basis of an initial predefined coding framework and, thus, a consistent focus for interpretation. The analysis was conducted by 2 independent researchers (RA and SV). After familiarization with the data, an iterative process of coding and indexing was adopted to ensure that important aspects of the data were not missed from the predefined coding framework. A working analytical framework was developed and applied to the coding of all the transcripts. The coded data were then charted to emerging themes, which were then summarized into the framework matrix. All data were coded, indexed, and charted using NVivo for Mac v12 (QSR International).

### Ethical Considerations

Ethical approval was granted for this study by the Joint Research Compliance Office at ICL under the Science Engineering Technology Research Ethics Committee process (reference 20IC5854). Informed consent was obtained from all the participants. All data were deidentified for the purposes of analysis, with each individual interview identified by an alphanumeric code. The participants were acknowledged only by their organization to avoid the identification of specific participants.

## Results

### Overview

The reported enablers and barriers to the implementation of Streams at ICHNT are described across the 5 key themes in successful digital change management [16]. The development of overarching themes and subthemes is summarized in Table 1 and Figure 1.

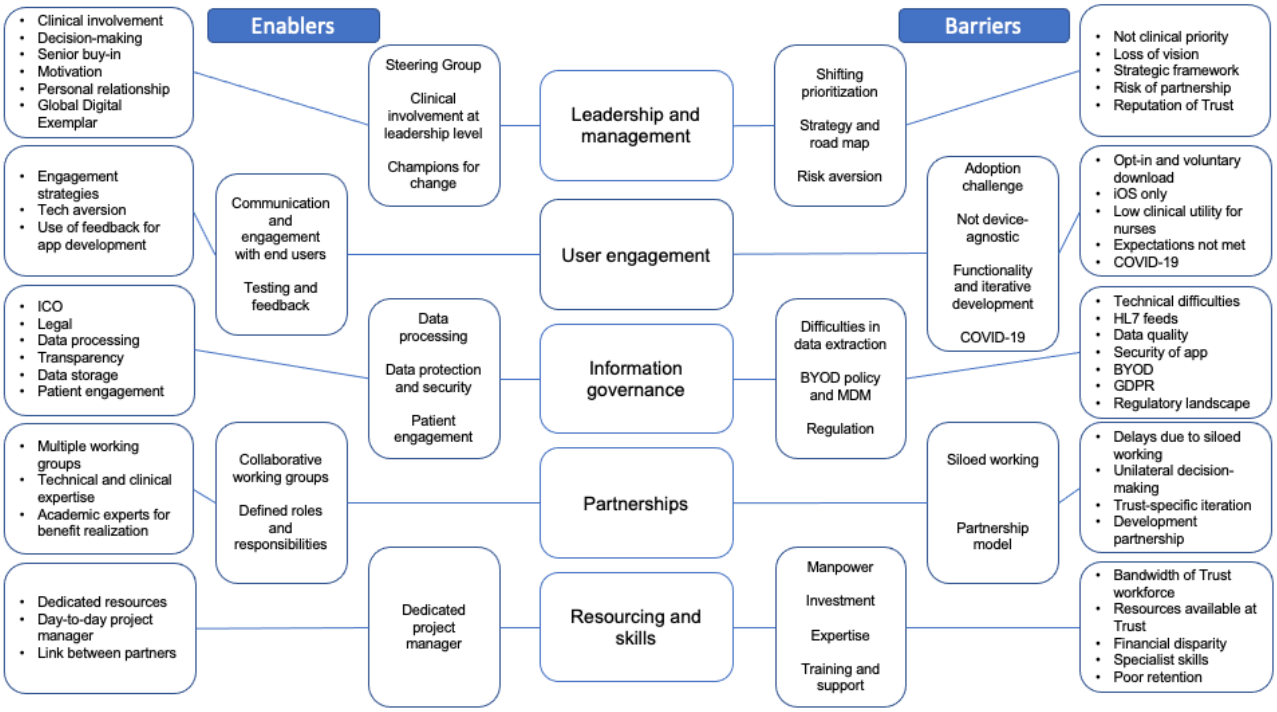
**Table 1.** Reported enablers and barriers to the implementation of the Streams app.

| Theme | Enablers | Barriers |
|---|---|---|
| Leadership and management | • Steering Group<br>• Clinical involvement at leadership level<br>• Motivation and champions for change | • Shifting prioritization<br>• Unclear strategy and road map<br>• Risk aversion |
| User engagement | • Communication and engagement with end users<br>• Testing and feedback | • Adoption challenge<br>• Not device-agnostic<br>• Functionality and iterative development<br>• COVID-19 |
| Information governance | • Data processing and information sharing<br>• Data protection and security<br>• Patient engagement and transparency | • Difficulties in data extraction<br>• BYOD[a] policy and MDM[b]<br>• Regulation |
| Partnerships | • Collaborative working groups<br>• Defined roles and responsibilities | • Siloed working<br>• Partnership model |
| Resourcing and skills | • Dedicated project manager | • Personnel<br>• Investment<br>• Expertise<br>• Training and support |

[a]BYOD: bring-your-own-device.

[b]MDM: mobile device management.

**Figure 1.** Thematic map for semistructured interviews demonstrating developed subthemes and overarching core themes. BYOD: bring-your-own-device; GDPR: General Data Protection Regulations; HL7: Health Level 7; ICO: Information Commissioner's Office; MDM: mobile device management.




## Leadership and Management

The involvement of senior leadership in the implementation process helped with decision-making, highlighted the importance of the project within the organization, and helped motivate other key stakeholders in the process. Among the enablers identified was the establishment of the SSG, consisting of key stakeholders and leaders from all 3 organizations, including the Chief Clinical Information Officer (CCIO), Caldicott Guardian, and Medical Director from the Trust; UK Lead and Project Managers from Google Health; and the Chief Scientific Advisor at the university partner. The SSG met every 6 weeks with an overarching remit to manage the project with decision-making by consensus. Clinical leaders were also recruited to the SSG and were able to act as champions for change among clinical teams:

> *I think the steering committee overall worked well. It had sort of senior people from both parties that met on a regular basis. I think it was important that there was that senior buy in [...] the senior people from both parties regularly engaged, despite busy schedules.* [Participant #13, Google Health]

> *There's a lot of clinical involvement and I think that ends with a product that is, at its core, clinically safe and has clinical utility.* [Participant #7, Google Health]

Barriers identified at a leadership level included shifting prioritization of the project, competing with a myriad of other information technology (IT) projects ongoing at the hospital, which affected the amount of dedicated resources allocated:

> *I think because Streams was not a key clinical system [...] you would not expect streams to be prioritized over other key clinical systems, obviously.* [Participant #13, Google Health]

An unclear strategic framework for deployment, risk aversion, and extensive due diligence caused by the alliance with a high-profile industry partner were also identified as barriers:

> *What there wasn't was almost that strategic framework within which to sit it, and I think we were both to blame for that, and actually possibly had either side pushed the other one a bit harder on that, that would have helped, but I think we were both a bit amateur on that front.* [Participant #1, ICHNT]

## User Engagement

Multiple strategies were trialed in an attempt to engage end users to participate in the implementation process. The enablers identified were the broad range of communication and promotional activities used to drive uptake, including attendance to routine clinical meetings and inductions, regular emails from various sources, and visual media placed around the hospital. Involving clinicians in testing and feedback sessions at an early stage of the change process also helped:

> *So on the whole that was good, the amount of feedback that we got back. And it helped iterate the product.* [Participant #14, Google Health]

The barriers identified were the adoption challenge where the system was opt-in and not integral to any clinical workflows. This was partially due to the limited functionality of the app when first deployed and the slow iterative improvements and updates during the development cycle. Moreover, the app was solely available on iOS devices; therefore, the potential user base was restricted:

> *I think ultimately the user base is driven by the utility of the product. If the product is super useful and provides value, then people will use it. Any limitation*

> *in the number of downloads, in my mind, always reflects back to the core value offering of the product.* [Participant #7, ICHNT]

> *I think also, in many cases, I think a lot of clinical users can be quite tech averse. And maybe that's from previous experience through your existing systems they're currently having to use. So there can be that barrier around, this is just going to have to be another thing that I'm going to have to use and it's almost coming at it from a "you need to show me the value before I actively engage in helping use this, I don't want it to be another burden on my clinical time."* [Participant #12, Google Health]

## Information Governance

Collaborating with industry partners as data processors raised information governance issues during the implementation process. Extra governance and precautions were required because of the public interest and scrutiny in the partnership, with proactive engagement with external bodies such as the ICO, National Data Guardian, and organizational legal teams:

> *We had to jump through more hoops, and we had to be a little bit more careful, because normally we'd have a process where we can sign off on systems and suppliers in a fairly straightforward manner. We tend to only engage with the ICO and National Data Guardian if we think there's a major problem.* [Participant #8, ICHNT]

Respondents commented that data protection and security aspects were handled well, especially during the migration of the data center to a cloud platform despite the significant delays it caused to the implementation process:

> *I think it's the rigour of the processes that we put in place around our information governance and I think we do have a very strong information governance capability within the organisation. I think it's been working collectively through that, but inevitably these things take time, don't they?* [Participant #6, ICHNT]

Comprehensive patient engagement was also identified as an enabler:

> *We need just to make sure that we were on board with the right messaging and we were engaging with all of the right partners, in addition to the public. We went out to the public in a number of different ways just to make sure we were transparent and in good faith, and really clear on the intent.* [Participant #9, ICHNT]

Difficulties in data extraction and assessing data quality from the EHR and the regulatory burden were noted to be barriers with the effect of delaying the implementation process. This involved extensive engagement over many months with a large number of clinical, technical, and legal stakeholders at the Trust to review data processing agreements and assess the quality, accuracy, and safety of the data being processed:

> *I think there was probably maybe slight frustrations on both sides [with regards to delays], but I think there was also recognition that we need to get this completely right, and it was much better to be delayed [...] than go fast and have another cycle.* [Participant #2, Google Health]

Furthermore, issues with BYOD policies and a change in plan midway through the development cycle to require an MDM solution for extra security were also noted to have delayed the implementation process:

> *I think had someone with a lot of experience in MDMs just been around, they could have just sorted it out in a week. [...] I think that's a lack of experience on both sides, probably.* [Participant #7, Google Health]

## Partnerships

Mutually reinforcing partnerships can help organizations with digital change. The working practices between the partners were frequently commented on by respondents. Enablers identified were the technical, implementation, user engagement, and clinical collaborative working groups that convened weekly and were established to oversee specific aspects of the project:

> *We've had a formal governance arrangement in place, which has built into it a series of meetings for different groups. We've got a Streams steering group that has met every six weeks. We've had a technical working group that has met fortnightly. We've had a programme management weekly meeting, and we've had very well-defined attendees and good regular attendance from the right people for those meetings. That has got us into quite a good cycle of good communication for particular areas.* [Participant #5, ICHNT]

Well-defined roles and responsibilities between partners were also established. The technical expertise of Google Health team members was used to develop and implement the app. Respondents felt that this was not something that the Trust would be capable of doing unilaterally:

> *I think most health systems and NHS organisations are, 'We should go out and partner with start-ups and established companies,' and I think that's the right approach.* [Participant #2, Google Health]

The established link and connections between the Trust and the University were used to engage academic experts in the implementation process. This enabled rigorous continuous evaluation and benefit realization of the app:

> *The university, particularly in this setting, were clinician scientists who have a feel for both the clinical practice and also the research.* [Participant #11, ICL]

Occasional siloed working practices among partners were identified as barriers, causing delays in the implementation and making the partnership model feel like a traditional supplier–client model rather than the development partnership that was envisaged:

> *If you want a development partner, you have to work in a much more integrated, collaborative manner, and they don't do that at the moment. They go away*

*with an idea. They say they've done a lot of thinking, and this is what they're planning to do, [...] We need to be a development partner.* [Participant #8, ICHNT]

### Resourcing and Skills

The ability of an organization to support mobile transformation is dependent on the resources and skills involved in the project. The presence of a dedicated, full-time project manager to oversee day-to-day running of the deployment was a key enabler in this process:

*I think what really helped was appointing a project manager at the Trust whose principal responsibility was to bring all the different competing teams at the Trust together, and act as a single point of contact and project manage it.* [Participant #3, ICL]

*I really feel like the project management at Imperial is very good and it's often not the organisation or the project management that's lacking, it's just it takes a long time because there isn't enough resource to actually do some of the work that the project manager is organising.* [Participant #7, Google Health]

However, respondents noted that the project manager was the only dedicated resource at the Trust and, therefore, a major barrier was the lack of personnel, investment, expertise, training, and support to help with the implementation process:

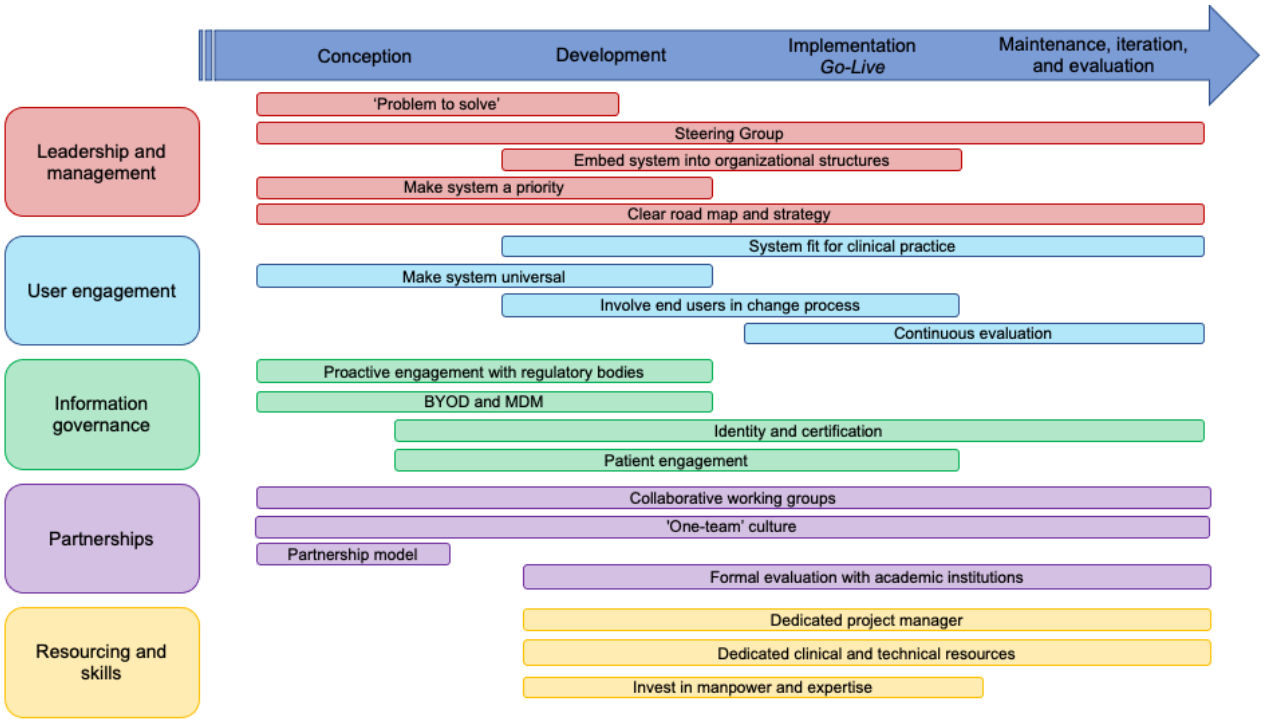
*I think if we'd have more resource we could have moved more quickly and we might have realised more of the original scope. I think it's been resource-constrained.* [Participant #6, ICHNT]

## Lessons Learned

### *Overview*

The implementation of mobile technology and working with commercial partners in NHS organizations were a significant challenge. The key lessons learned from this process are described in this section in addition to the proposed road map for the implementation of clinical mobile technology developed with commercial partners at scale. These are illustrated with temporal relationships in Figure 2.

**Figure 2.** Proposed road map for large-scale, hospital-wide mobile implementation. BYOD: bring-your-own-device; MDM: mobile device management.




### *Leadership and Management*

#### Clarify the Problem to Solve

Be specific about the clinical problem that needs solving and assess at an early stage whether this problem can be ameliorated by mobile technology. This will help create a shared vision across the organization.

#### Make the System a Priority

Create a sense of urgency in the organization by making the deployment a priority among the leadership and end users.

#### Steering Committee With Senior Key Stakeholders and Clinical Involvement

Establish a board-level steering committee comprising high-level key stakeholders to guide the implementation. They should act as a decision-making body and build consensus on the strategic vision. This should include clinical leaders such as the CCIO and deputy CCIOs, which will help engage other clinicians and aid with dissemination.

#### Clear Road Map and Strategy

Implementation strategies need to be tailored to organizational circumstances and should be well-planned, allowing room for flexibility in timelines.

#### Embed Program Into Existing Organizational Structures

Ensure that mobile change is implemented as a key part of the organization's digital transformation rather than as an isolated project and that it is interoperable with other systems already in place.

### *User Engagement*

#### Choose a System That Is Fit for Clinical Practice and Focused on End Users

The system being deployed needs to be usable and effective for clinicians. It should fit into existing clinical workflows to improve the quality of care or efficiency. The value of the system should be demonstrated to clinicians to warrant adoption.

#### Make the System Universal

Choose a system that can be universally adopted. It should be device- and operating software–agnostic and offer functionality that will be useful to multiple clinical user groups.

#### Involve End Users in the Change Process

Give end users a sense of ownership over the change process and involve them in the iterative development of the product. A clinical user group should be established as a forum to discuss new features and contribute to user acceptance testing, and local champions should be enlisted. A variety of communication strategies to engage with end users to promote the system and ensure adoption should also be used. The strategy should be linked with wider Trust communications to help with distribution of promotional material.

#### Continuous Feedback to Evaluate and Inform Iterative Development

It is essential to capture user feedback on the system and respond in a timely manner. An agile, rapid turnaround should be targeted for iterative development to demonstrate to end users that their feedback is being regarded.

### *Information Governance*

#### Proactive Engagement With Governance and Regulatory Organizations

If required, proactive engagement with the ICO and other regulatory bodies is recommended to ensure that all data processing, data security, and regulatory guidance are adhered to. Data processing, particularly if undertaken by industry partners, should be transparent and within legal boundaries.

#### Consider BYOD and MDM Policies at Early Stage

BYOD and MDM policies should be formulated to ensure security and privacy with widespread mobile implementation. If MDM is chosen, investment in the software will need to be considered. Tackling staff perceptions and attitudes toward MDM on personal devices will also need to be explored further before MDM is widely accepted.

#### Identity and Certification

The product should be embedded in the digital ecosystem of the NHS organization. The system should be made secure by using existing active directories for account creation and role-based access permissions.

#### Patient Engagement

Be open and transparent with patient and public groups about the nature of the partnership and data processing and security aspects.

### *Relationship Between Partners*

#### Collaborative Working Groups With Defined Escalation Pathways

Clinical and technical work streams with defined roles and responsibilities should be established. These working groups should have defined terms of reference and clear escalation paths.

#### Frequent Communication Between Partners at All Levels

Frequent and effective communication channels between partners should be established to instill a *one-team* culture.

#### Decide on Partnership or Procurement Model

Appraise the nature of the partnership with the commercial supplier and decide whether a customizable system or an *off-the-shelf* system is required.

#### Partner With Academic Institutions to Perform Formal Evaluation

Pre-existing links with academic institutions should be used for continuous independent evaluation and benefit realization.

### *Resourcing and Skills*

#### Dedicated Project Manager to Drive Through Vision and Act as Key Point of Call

Appoint a dedicated project manager who focuses on all activities related to clinical implementation. They should act as the key point of call to liaise with all stakeholders to help overcome any barriers.

#### Dedicated Clinical Implementation and Technical Teams

Ensure that resources are available for dedicated clinical implementation and technical teams within the NHS organization to support the implementation process.

#### Investment in Personnel and Expertise

Investments should be made in specialist expertise that may be required, personnel to aid implementation, training of staff, and ongoing support.

## Discussion

### Principal Findings

This study sought to (1) identify the enablers and barriers to the widespread implementation of mobile technology in an NHS Trust and (2) formulate an implementation road map from the experiences and perspectives of those leading and heavily involved with the change management process. In doing so, key enablers and barriers and the implementation road map were mapped onto 5 overarching themes that encompassed all the crucial aspects of the digital change management process: *leadership and management, user engagement, information governance, partnerships,* and *resources and skills* [16].

We identified that, despite the implementation occurring in a Global Digital Exemplar NHS Trust with world-class IT infrastructure [23] in partnership with one of the world's biggest technology companies and with strong support from an array of key stakeholders, the cultural and technological change required for mobile working and implementation in health care was a significant challenge. Widespread deployment of the mobile app was pursued at our organization; however, multiple barriers and hurdles were encountered along the process. These barriers were acknowledged to have either contributed to delays in the implementation or decreased the adoption of the app among end users. The key barriers identified were as follows: (1) delays to implementation (shifting prioritization, risk aversion, instigation of MDM policies and investment in software, problems with data quality and data extraction from the EHR, limited resources at the Trust, and the migration of data storage to a cloud-based platform) and (2) decreased adoption (limited functionality of the app that did not integrate into clinical workflows and was not clinically useful for large proportions of the workforce, the tardiness of iterative development and responding to feedback, and the fact that the app was not device-agnostic).

By considering the experiences and perspectives of key stakeholders in overcoming the aforementioned barriers, together with the enablers that were recognized to be present within the partnership, we proposed a novel implementation road map for mobile technology deployment at scale. This road map highlights the key lessons learned, which may act as a blueprint for multi-stakeholder scoping processes in health care organizations considering mobile transformation. This may help avoid some of the commonly encountered pitfalls and improve the likelihood of successful implementation of mobile technology.

This study also identified and exposed some of the difficulties NHS organizations may encounter if working with industry partners for digital change. With regard to Streams at ICHNT, added transparency about the partnership was required because of the media scrutiny and public interest [24]. This high-profile partnership led to some risk aversion on behalf of the NHS organization; however, the implementation of the General Data Protection Regulations and proactive engagement with regulators provided a secure backdrop for data processing. Furthermore, the relationship between the Trust and the industry partner was recognized as a development partnership rather than a supplier–customer partnership [25]. With this type of partnership being relatively novel and all partners having different ways of working, it is perhaps inevitable that frustrations were noted about the levels of collaborative working and the alignment of goals between the partners.

Although mobile change shares much similarity with digital change, it comes with its own unique challenges [26,27]. In this study, although many of the barriers can also be applied to digital change, additional mobile-specific barriers were identified, such as the adoption challenge and privacy and security concerns related to HCPs using their own smartphones for clinical purposes. Long-held beliefs about the appropriateness of mobile phones in hospital settings may be a further barrier [28,29]. This is particularly relevant with the prospect of HCPs using their personal smartphones for clinical purposes on a BYOD basis. This strategy can blur the lines between professional and personal use of smartphones, potentially creating conflicts that could arise related to their use in the wards or when off-site [30]. Furthermore, BYOD raises governance, information security, and patient confidentiality issues [27], which must be addressed. MDM software can be a solution to these concerns because of its ability to enforce security policies and secure devices [31]; however, implications such as the loss of control and privacy felt by staff and the financial investment required by the organization with this system may not be universally acceptable. Other options, including *hardening* the security of the app through features such as 2-factor authentication and geolocation, could be considered if MDM is not appropriate [32]. Overcoming these barriers is important to ensure widespread acceptability of mobile devices in clinical settings and requires increased awareness among both HCPs and patients as to the benefits of these technologies.

Although the cultural barriers may take more time and resources to overcome, there does now appear to be extensive and widespread progress in overcoming the technical and information governance barriers to mobile change identified in our study. The introduction of the General Data Protection Regulations in Europe has defined the legal framework for data processing, and guidance from NHSX [30] and the Medicines and Healthcare Products Regulatory Agency [33] has defined the current regulatory framework. The new EU Medical Device Regulation, implemented in 2021, will redefine the regulatory framework with more stringent and specific protocols for various types of medical device software and mobile apps [34]. A mobile technology investment toolkit was also recently published by NHS Digital that provided practical tools and resources for IT leads to deploy mobile technology [35]. Furthermore, the difficulties encountered in our study with data extraction, sharing, and interoperability will be alleviated with increased use of Fast Healthcare Interoperability Resources (a new international standard for health care data formats and elements) [36].

With increasing digital technology being introduced into the health care space and the potential introduction of advanced data-driven technologies, it is inevitable that NHS organizations will need to continue to work with commercial and industry partners to develop and implement interventions. This is now accepted, and the NHS relies on numerous strategic partnerships to improve outcomes and deliver its ambitions in all fields [37]. These partnerships need to be transparent and comply with legal, regulatory, and ethical boundaries [38] to ensure that the partnership is acceptable to patients and that they can trust it. However, the approach of technology companies in the health sector can be challenging as successful methods that have been used in other industries may not be appropriate in the regulated and necessarily risk-averse health care space. Maximizing new partnerships with technology companies requires the development of innovative interventions, the agile deployment of solutions in clinical environments, and ongoing evaluation and iterative development to improve the product [39]. Although more traditional medical device companies are limited in their

speed of introducing products to market because of the time needed for design, safety testing, manufacturing, and efficacy trials, technology companies may lack these restraints and lean toward rapid iteration and updates to evolve and improve products. This can create opposing views on the balance between careful evaluation and thorough evidence-based principles, and rapid technology development and fast product cycles that need to be addressed [40]. Tension with other partners such as academic institutions may also ensue as comprehensive evaluations of a novel intervention inevitably require time and added cost, which must be accounted for in the product road maps of commercial companies.

### Limitations

The study was limited to a single NHS hospital and the implementation of a single clinical mobile app. Although this inevitably influenced the perspectives of the interviewees and many of the findings were related to local contextual factors, we believe that the broad sample of key stakeholders interviewed and robust qualitative analysis identified issues that are generalizable to the implementation of other mobile technologies in hospital settings. The sample size and heterogeneity of the participants was unavoidable because of the limited number of key stakeholders and members of the SSG suitable for inclusion in the study. However, purposeful sampling was used to ensure that a representative cross-section of the SSG was included in the study. Furthermore, we accept that there will be a bias with members of the SSG reviewing their own role in the implementation process. Viewpoints of end users of the app were not explored to compare and provide a *top-to-bottom* view of the mobile change management process. Finally, differences across various health care settings, hospitals, and departments were also not explored in this study.

### Conclusions

The implementation of mobile technology in health care and working with commercial partners has been a significant and complex challenge for NHS Trusts. With the requirement of more industry partnerships for advanced digital technologies in the future, the findings of this study should influence how other organizations undertake similar mobile transformations and improve the probability of successful implementation and widespread adoption. By overcoming the cultural and technological barriers identified and observing the proposed road map, future deployments of mobile technology in health care settings could be facilitated and have a greater chance of success.

## Acknowledgments

The authors would like to thank all the participants in this study from Imperial College Healthcare National Health Service Trust, Imperial College London, and Google Health for their support and for agreeing to be interviewed for this study.

## Conflicts of Interest

AD is Chair of the Health Security Initiative at Flagship Pioneering UK Ltd. As a previous employee, DK holds stock in Google. HA is Chief Scientific Officer, Pre-emptive Medicine and Health Security at Flagship Pioneering

## References

1. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/M13-1531] [Medline: 24573664]
2. Wachter R. Making IT work: harnessing the power of health information technology to improve care in England. Report of the National Advisory Group on Health Information Technology in England. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2022-01-20]
3. Greenhalgh T, Swinglehurst D, Stones R. Rethinking resistance to'big IT': a sociological study of why and when healthcare staff do not use nationally mandated information and communication technologies. In: Health Services and Delivery Research. Southampton, UK: NIHR Journals Library; 2014.
4. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
5. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 07;1(4):174-181. [doi: 10.1136/bmjinnov-2015-000062]
6. Digital health in the UK: an industry study for the office of life sciences. Office for Life Sciences. 2015. URL: https://tinyurl.com/2fkjemmh [accessed 2022-01-20]
7. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
8. Payne KF, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]

9. Wu R, Rossos P, Quan S, Reeves S, Lo V, Wong B, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. J Med Internet Res 2011 Aug 29;13(3):e59 [FREE Full text] [doi: 10.2196/jmir.1655] [Medline: 21875849]
10. O'Connor P, Byrne D, Butt M, Offiah G, Lydon S, Mc Inerney K, et al. Interns and their smartphones: use for clinical practice. Postgrad Med J 2014 Feb;90(1060):75-79. [doi: 10.1136/postgradmedj-2013-131930] [Medline: 24243966]
11. Scaling streams with Google. DeepMind. URL: https://deepmind.com/blog/announcements/scaling-streams-google [accessed 2022-01-20]
12. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for acute kidney injury management in hospital emergency admissions. NPJ Digit Med 2019;2:67 [FREE Full text] [doi: 10.1038/s41746-019-0100-6] [Medline: 31396561]
13. Hern A. Google DeepMind 1.6 m patient record deal 'inappropriate'. The Guardian. 2017. URL: https://www.theguardian.com/technology/2017/may/16/google-deepmind-16m-patient-record-deal-inappropriate-data-guardian-royal-free [accessed 2022-01-19]
14. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
15. RFA0627721 provision of patient data to DeepMind. Information Commissioner's Office. 2017. URL: https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf [accessed 2022-01-20]
16. Maguire D, Evans H, Honeyman M, Omojomolo D. Digital Change in Health and Social Care. London: King's Fund; 2018.
17. Etikan I. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1. [doi: 10.11648/j.ajtas.20160501.11]
18. Marshall M. The key informant technique. Fam Pract 1996 Feb;13(1):92-97. [doi: 10.1093/fampra/13.1.92] [Medline: 8671109]
19. Marshall B, Cardon P, Poddar A, Fontenot R. Does sample size matter in qualitative research?: a review of qualitative interviews in is research. J Comput Inf Syst 2015 Dec 10;54(1):11-22. [doi: 10.1080/08874417.2013.11645667]
20. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Nov 21;18(1):148 [FREE Full text] [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]
21. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Global digital exemplars. NHS England. URL: https://www.england.nhs.uk/digitaltechnology/connecteddigitalsystems/exemplars/ [accessed 2022-01-20]
24. Google Health ties up data agreements with NHS trusts. Digital Health. URL: https://www.digitalhealth.net/2019/09/google-health-ties-up-data-agreements-with-nhs-trusts/ [accessed 2022-01-20]
25. Creating 'beyond the product' partnerships between providers and medtech players. McKinsey & Company. URL: https://tinyurl.com/4knrbkpf [accessed 2022-01-19]
26. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR Mhealth Uhealth 2019 May 03;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
27. Shah N, Martin G, Archer S, Arora S, King D, Darzi A. Exploring mobile working in healthcare: clinical perspectives on transitioning to a mobile first culture of work. Int J Med Inform 2019 May;125:96-101. [doi: 10.1016/j.ijmedinf.2019.03.003] [Medline: 30914187]
28. Thomairy NA, Mummaneni M, Alsalamah S, Moussa N, Coustasse A. Use of smartphones in hospitals. Health Care Manag (Frederick) 2015;34(4):297-307. [doi: 10.1097/HCM.0000000000000080] [Medline: 26506291]
29. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
30. Bring your own device (BYOD) IG guidance. NHSX. URL: https://www.nhsx.nhs.uk/information-governance/guidance/bring-your-own-devices-byod-ig-guidance/ [accessed 2022-01-20]
31. Sisala S, Othman SH. Developing a Mobile Device Management (MDM) security metamodel for Bring Your Own Devices (BYOD) in hospitals. Int J Innov Comp 2020 Nov 19;10(2). [doi: 10.11113/ijic.v10n2.273]
32. Mansfield-Devine S. Your life in your hands: the security issues with healthcare apps. Netw Security 2016 Apr;2016(4):14-18 [FREE Full text] [doi: 10.1016/s1353-4858(16)30038-1]
33. Medical devices: software applications (apps). GOV.UK. 2014. URL: https://www.gov.uk/government/publications/medical-devices-software-applications-apps [accessed 2022-01-20]
34. Keutzer L, Simonsson US. Medical device apps: an introduction to regulatory affairs for developers. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e17567 [FREE Full text] [doi: 10.2196/17567] [Medline: 32589154]

35. Mobile technology investment toolkit. NHS Digital. URL: https://digital.nhs.uk/services/mobile-technology-investment-toolkit [accessed 2022-01-20]
36. Mandel J, Kreda D, Mandl K, Kohane I, Ramoni R. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association 2016;23(5):a. [doi: 10.1093/jamia/ocv189]
37. Partnerships and relationships. NHS England. URL: https://www.england.nhs.uk/ourwork/part-rel/ [accessed 2022-01-20]
38. Making NHS data work for everyone. Reform. URL: https://reform.uk/research/making-nhs-data-work-everyone [accessed 2022-01-20]
39. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018;1:48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
40. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]

## Abbreviations

**BYOD:** bring-your-own-device
**CCIO:** Chief Clinical Information Officer
**EHR:** electronic health record
**HCP:** health care professional
**ICHNT:** Imperial College Healthcare National Health Service Trust
**ICL:** Imperial College London
**ICO:** Information Commissioner's Office
**IT:** information technology
**MDM:** mobile device management
**NHS:** National Health Service
**SSG:** Streams Steering Group

*Edited by L Buis; submitted 23.06.21; peer-reviewed by T Ong, S Liu; comments to author 11.09.21; revised version received 11.10.21; accepted 14.12.21; published 08.02.22.*

*Please cite as:*
*Aggarwal R, Visram S, Martin G, Sounderajah V, Gautama S, Jarrold K, Klaber R, Maxwell S, Neal J, Pegg J, Redhead J, King D, Ashrafian H, Darzi A*
*Defining the Enablers and Barriers to the Implementation of Large-scale, Health Care–Related Mobile Technology: Qualitative Case Study in a Tertiary Hospital Setting*
*JMIR Mhealth Uhealth 2022;10(2):e31497*
*URL: https://mhealth.jmir.org/2022/2/e31497*
*doi:10.2196/31497*
*PMID:35133287*

Original Paper

# Willingness of French General Practitioners to Prescribe mHealth Apps and Devices: Quantitative Study

Claire Della Vecchia[1,2], MPH, PhD; Tanguy Leroy[1,2], PhD; Charlotte Bauquier[1,2], PhD; Myriam Pannard[1,2], PhD; Aline Sarradon-Eck[3,4], MD, PhD; David Darmon[3,5], MD, PhD; Jean-Charles Dufour[3,6], PhD; Marie Preau[1,2], PhD

[1]Institut de Psychologie, Université Lyon 2, Bron, France

[2]Institut National de la Santé et de la Recherche Médicale, Unité 1296, "Radiations: Defense, Health and Environment" Centre Léon-Bérard, Lyon, France

[3]Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Institut de Recherche pour le Développement, Aix-Marseille Université, Marseille, France

[4]Institut Paoli-Calmettes, CanBios, Marseille, France

[5]Département d'Enseignement et de Recherche en Médecine Générale, Université Côte d'Azur, Nice, France

[6]Service Biostatistique et Technologies de l'Information et de la Communication, Assistance Publique Hôpitaux de Marseille, Hôpital de la Timone, Marseille, France

**Corresponding Author:**
Claire Della Vecchia, MPH, PhD
Institut de Psychologie
Université Lyon 2
5 Avenue Pierre Mendes France
Bron, 69676
France
Phone: 33 0672303682
Email: dellavecchia.claire@gmail.com

## Abstract

**Background:** The field of mobile health (mHealth) is constantly expanding. Integrating mHealth apps and devices in clinical practice is a major and complex challenge. General practitioners (GPs) are an essential link in a patient's care pathway. As they are patients' preferred health care intermediaries, GPs play an important role in supporting patients' transition to mHealth.

**Objective:** This study aims to identify the factors associated with the willingness of French GPs to prescribe mHealth apps and devices to their patients.

**Methods:** This study was part of the ApiAppS project whose overall objective was to help remove barriers GPs face when prescribing mHealth apps and devices by developing a custom-built platform to aid them. The study included GPs recruited from the general practice department of several medical faculties in France (Lyon, Nice, and Rouen) and mailing lists of academic GPs, health care professional associations, and social and professional networks. Participants were asked to complete a web-based questionnaire that collected data on various sociodemographic variables, indicators of their involvement in continued education programs and the amount of time they dedicated to promoting healthy behaviors during patient consultations, and indicators characterizing their patient population. Data on their perceptions of mHealth apps and devices were also collected. Finally, the questionnaire included items to measure GPs' acceptability of prescribing mHealth apps and devices for several health-related dimensions.

**Results:** Of the 174 GPs, 129 (74.1%) declared their willingness to prescribe mHealth apps and devices to their patients. In multivariate analysis, involvement in continued education programs (odds ratio [OR] 6.17, 95% CI 1.52-28.72), a better patient base command of the French language (OR 1.45, 95% CI 1.13-1.88), GP-perceived benefits of mHealth apps and devices for both patients and their medical practice and GP-perceived drivers for mHealth apps and device implementation in their medical practice (OR 1.04, 95% CI 1.01-1.07), and validation of mHealth apps and devices through randomized clinical trials (OR 1.02, 95% CI 1.00-1.04) were all associated with GPs' willingness to prescribe mHealth apps and devices. In contrast, older GPs (OR 0.95, 95% CI 0.91-0.98), female GPs (OR 0.26, 95% CI 0.09-0.69), and those who perceived risks for the patient or their medical practice (OR 0.96, 95% CI 0.94-0.99) were less inclined to prescribe mHealth apps and devices.

**Conclusions:** mHealth apps and devices were generally seen by GPs as useful in general medicine and were, for the most part, favorable to prescribing them. Their full integration in general medicine will be conditioned by the need for conclusive certification,

transparency (reliable and precise data concerning mHealth app and device methods of construction and clinical validation), software aids to assist GPs prescribe them, and dedicated training programs.

***(JMIR Mhealth Uhealth 2022;10(2):e28372)*** doi:10.2196/28372

**KEYWORDS**

mHealth; health applications; connected health and wellness devices; general practitioners; patients; prescription; quantitative study; mobile phone

## Introduction

### mHealth Apps and Devices Worldwide

The World Health Organization defines mobile health (mHealth) as a "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [1]. The area of mHealth continues to grow globally: in June 2021, there were over 350,000 health-related mobile apps worldwide, with more than 250 new apps being added to web-based stores every day [2]. In 2016, over 73 million connected health and wellness devices were sold worldwide. The report forecast a huge increase to 160 million devices sold in 2020 (the report on 2016-2020 data is not yet published) [3].

### The Prescribing of mHealth Apps and Devices

Although the feasibility of prescribing mHealth apps and devices in general practice, in Australia [4] and Spain [5] notably, has been demonstrated, their integration in clinical practice worldwide presents a complex challenge. In France, the various physician organizations agree on the role of the general practitioner (GP) in health care, especially to provide primary care based on a comprehensive approach that includes providing advice and support focusing on education, risk prevention, and health promotion. GPs also play a role in monitoring and coordinating patient care (guaranteeing communication between themselves and other professionals involved in their patients' care) [6]. GPs are essential links in the patient's care pathway [7,8]. According to a 2018 French general population study, 83% of French people consulted a GP at least once a year [9] and 90% of health problems are managed in primary care (especially in general practice) [10]. As GPs are patients' preferred health care intermediaries, they play a key role in patient support and patients' relationship with mHealth, especially by providing guidance and advice. To promote the full implementation and acceptance of mHealth in general medicine, it is necessary to consider upstream both human (attitudes, expectations toward mHealth, and the characteristics of GPs and their patient base) and technical implications (mHealth apps and device functionalities, ease of use, ease of data transfer, operability, compatibility with electronic medical records and computer software used by GPs, etc) [11]. In France, to facilitate the integration of mHealth, the ApiAppS project aims to propose a type of software (considering the perceptions GPs have toward mHealth) in primary care to help GPs prescribe mHealth apps and devices adapted to the patient's condition and provide reliable information regarding mHealth apps and devices [11].

### The Prescribing of mHealth Apps and Devices in the French Context

The national organization of French physicians published a report in 2015 that defined recommendations for good practices in mHealth app and device use [12]. The report indicated that mHealth apps and devices must support care to strengthen prevention behaviors, improve care monitoring and coordination, strengthen the patient–physician relationship, enable better access to care, and promote the empowerment of patients [12]. These same elements were highlighted in a French study investigating the drivers for the use of mHealth apps and devices in general medicine [13]. In this study, beyond the simple (informal) recommendation, it seemed interesting to investigate the potential prescription of mHealth apps and devices. In France, physicians are responsible for writing the prescription and making sure to give all the necessary information to the patient or their entourage to ensure proper compliance and the correct use of the prescribed elements [14]. The prescription is thus much more binding for physicians than a simple recommendation, constituting a material symbol of the patient–physician relationship [15] and the document required for reimbursement by the French social security health care insurance. In France, some connected health devices are reimbursed by social security health care insurance, especially devices for the management of diabetes, coagulation disorders, sleep apnea, and asthma. However, to our knowledge, only 2 mHealth apps—one for monitoring diabetes and the other for lung cancer—can currently be prescribed by physicians and then reimbursed by French social security health care insurance.

### Risks and Obstacles Linked to the Use of mHealth Apps and Devices

Although there are many potential advantages of mHealth apps and devices, their methods of construction, validation, and uses must be regulated. The international literature highlights the various primary types of risks and obstacles linked to the use of mHealth apps and devices in the following areas: data processing (data security and the use of personal data), reliability (lack of clinical validation, evaluation, precision of measurements, and reliable sources listing mHealth apps and devices) [12,13,16-19], the impact on patient care, and quality of the patient–physician relationship [13,20-22]. Physicians have also reported potential obstacles directly linked to their practice, in terms of the additional time spent during consultations processing digital information and providing patients support in the use of mHealth apps and devices [13,16,21,22], as well as the risk that the current divide between digitally literate and illiterate patients will become even greater, something that could increase inequalities in quality and access to care [13,16].

### GPs' Perceptions of mHealth Apps and Devices and Willingness to Prescribe

To date, few studies have investigated the perceptions of GPs about mHealth apps and devices and how these beliefs are associated with their willingness to prescribe mHealth apps and devices to their patients. This way, a qualitative study was conducted with French GPs with the aim of investigating their attitudes toward the prescription of mHealth apps and devices. They identified 3 groups of attitudes. The first group corresponds to GPs very willing to prescribe mHealth apps and devices, with positive perceptions of (1) the benefits for patients and for clinical practice and (2) ease of use; the second group represents GPs worried about the protection of patient data and the reliability of mHealth app and device content; and the third group corresponds to GPs concerned about the implications of mHealth apps and devices for their clinical practice (additional working time, modification of the patient–physician relationship, and the importance of mHealth app and device certification by independent entities) [23]. Consistently, an Australian study showed that the perceived barriers of GPs to prescribing mHealth apps and devices were generational digital divide, a lack of knowledge and reliable resources listing prescriptible mHealth apps and devices, additional working time it may represent for GPs, and concerns about data security [4]. To our knowledge, most studies have focused on factors associated with the intentions of physicians and other health care professionals to use mHealth apps and devices to support their own clinical practice (drug database, medical calculators, making appointments, etc) or studies based on informal recommendation (mostly oral) of mHealth interventions to their patients. Fewer studies focused on mHealth apps and devices intended for prescription or its equivalent in some contexts (formal and written recommendation), more binding for GPs and their patients. One study conducted among Turkish physicians showed that a manifest interest in mHealth apps and devices, very little fear about using them, perceiving mHealth apps and devices as useful for medical practice, and ease of access for physicians were associated with an increased willingness to use mHealth apps in medical practice [24]. These results were corroborated by 2 other studies conducted with Chinese health care professionals [25,26]. Unfortunately, none of these 3 studies provided much information on the determinants of prescribing mHealth apps and devices to patients in medical practice, as they focused on mHealth app and device acceptance by health care professionals (especially mHealth apps and devices for their own practice) and not on mHealth app and device prescription purposes. However, we found a study that investigated factors associated with the willingness of Malaysian GPs to recommend mHealth apps to their patients. They showed that in multivariate analysis, performance expectancy of the mHealth apps (improving patient health, improving chronic disease management, and encouraging patients to gain health knowledge) was associated with the willingness of GPs to recommend mHealth apps [27]. However, this study focused only on the mHealth app recommendation, which may have different implications for mHealth app and device prescription. A recent descriptive study was conducted to better understand German GPs' perceptions of mHealth apps. Of the 2138 GPs, although 60% recognized that mHealth apps could strengthen the involvement of people in the management of their health, only 36% reported global positive opinions of the health apps and only 18% frequently recommend mHealth apps to their patients, and the main criteria reported to recommend these apps were ease of use, guarantees for data privacy, and clinical validation [28]. However, this study focused only on mHealth app recommendation (not prescription), remained descriptive, thus results have to be corroborated by analytic studies.

### Objectives

On the basis of data from the literature, it seems essential to quantitatively describe GPs' perceptions of mHealth apps and devices and investigate the factors associated with the willingness of French GPs to prescribe mHealth apps and devices to their patients, constituting the objective of this study. This way, we hypothesize in a psychosocial perspective that characteristics related to GPs themselves, their practice, their patient base, and their perceptions of mHealth apps and devices may influence their willingness to prescribe mHealth apps and devices.

## Methods

### Study Design and Study Population

This study adheres to and has been reported following the Checklist for Reporting Results of Internet E-Surveys guidelines (Multimedia Appendix 1) [29].

This quantitative study was part of a larger project called ApiAppS (funded by the National Research Agency of France under grant ANR-17-CE19-0027 [11]), whose overall objective is to help remove barriers GPs face when prescribing mHealth apps and devices by developing a custom-built platform to aid them. This study aimed to confirm the results of a previous exploratory qualitative study [23], which investigated the attitudes of French GPs about prescribing patient-based mHealth apps and devices by analyzing their perceptions and expectations of mHealth apps and devices through semistructured interviews and focus groups [23].

For this quantitative study, we constructed a web-based, self-administered questionnaire on the basis of the results of the qualitative study concerning attitudes of GPs toward the prescription of mHealth apps and devices [23] and elements from the literature concerning mHealth apps and devices in current clinical practice. The various indicators measured by the questionnaire are described in detail in subsequent sections. The questionnaire was pretested with 8 GPs regarding the understanding of the different items in the questionnaire, and the researchers tested the technical functionalities of the questionnaire (any technical problems with posting the questionnaire on the web-based platform: no glitches in the layout of the questions and answers, the sequencing of the questions, and the recording of the questionnaire). These 2 test phases made it possible to correct, where necessary, the layout of the questionnaire (spelling, fonts, order of questions, etc) and the obligatory or nonobligatory nature of each question. From June 2019 to December 2019, the questionnaire was then distributed to GPs recruited through several academic

departments of general practice of several medical faculties in France (Lyon, Nice, and Rouen) and also from mailing lists of the academic GPs, health care professional associations, and social and professional networks. Participation in the study was voluntary and required a survey link, which headed toward the questionnaire (on the LimeSurvey platform). Information was provided about the time needed to fill in the questionnaire (about 15 minutes, 22 items), reminders about the rights of research participants under French law (anonymity, confidentiality, processing of data for research purposes, right of access, and data rectification), and email addresses of the researchers in charge of the study were provided. Once participants had validated their answers to the questionnaire, they could no longer review and change their answers. Only fully completed questionnaires were analyzed.

Ethical approval was obtained from the French Institute of Medical and Health Research Ethics Committee (IORG0003254 and FWA00005831) and the institutional review board (IRB00003888; opinion number 18-499).

### Questionnaire

The questionnaire, available in Multimedia Appendix 2, for this study included sociodemographic variables relating to GPs, indicators of their participation in continuing education programs, and the amount of time they dedicated to promoting healthy behaviors to their patients during consultations. It also included variables that allowed us to characterize the patient population. Finally, to meet the objective of this study, the questionnaire included variables aimed at gathering a greater understanding of GPs perceptions of mHealth apps and devices.

The objective of this study is to predict the willingness of GPs to prescribe mHealth apps and devices for twelve health-related dimensions: physical activity, dental health, nutrition, vaccination, sexual and reproductive health, well-being and mental health, addictions, asthma and allergies, dermatology, diabetes, first aid, and support for caregivers. For the analysis, to oppose 2 profiles of GPs, willingness to prescribe mHealth apps and devices was dichotomized into willingness and unwillingness to prescribe mHealth apps and devices for at least one health-related dimension.

Potential predictive variables were sociodemographic factors (age and gender) and factors related to GPs practices (indicators of GP involvement in continuous medical education, including subscription to professional journals, participation in peer groups, training, and presence in physician-based social networks), and the amount of time they dedicated during consultations in promoting healthy behaviors. We also considered psychosocial variables to help characterize the patient population as follows: the mean age of the practice population in the previous month; the place of residence; and the perceptions of GPs of the overall socioeconomic status, command of the French language, and self-management skills in terms of health of the patient base. On the basis of the previous qualitative study regarding attitudes GPs have toward mHealth app and device prescriptions [23], their perceptions of mHealth apps and devices were investigated as other potential predictors in our analysis:

- Facilitators regarding mHealth apps and devices implementation in general medicine, that is, both the potential perceived benefits of mHealth apps and devices and the levers to their implementation. This indicator included providing better access to care for patients, patient empowerment, better communication, quality of life, and work management for caregivers; obtaining additional information from patients (Patient-Reported Outcome Measures [30]); and facilitating links between the various professionals involved in patient care, an alternative to prescribing drugs, the strengthening of the patient–physician relationship, the perception of the importance of the role of the physician in the transition to mHealth, and the possibility of having a software aid that would automatically suggest mHealth apps and devices adapted to the needs of the patient.
- Obstacles to the implementation of mHealth apps and devices in general practice, that is, the risks and barriers associated with the use of mHealth apps and devices. This indicator included the dangers linked to misuse of mHealth apps and devices by patients, risks associated with self-medication, dehumanization of the patient–physician relationship, increase in patient anxiety because of the wealth of information available, use of personal data of patients, possibility of monitoring activities of GPs by health authorities, and devotion of additional time to mHealth apps and devices during consultations.
- Indicators relating to GPs perception of the importance of the development, clinical validation, and certification of mHealth apps and devices by GP-perceived trusted actors in health (eg, independent experts, patients' associations, academic researchers, physicians, health-related organizations, and stakeholders). Furthermore, GPs' perceptions of the importance of the involvement of health-related organizations and stakeholders in promoting the use of mHealth apps and devices in general medicine.

Scores on these indicators ranging from 0 to 100, with 100 representing the greatest perceived benefit or driver, risk or barrier, involvement, or utility, as applicable.

### Data Analyses

Several principal component analyses were performed to identify the underlying structure of data and highlight indicators (by grouping items belonging to the same 1D construct to generate a score). Specifically, for each component, the eigenvalues were extracted to capture the percentage of inertia explained by the component. Those greater than 1 (Kaiser criterion) were retained [31]. The choice of components was compared with the graph of the eigenvalues [32]. Finally, the results of these 2 methods were compared using parallel analysis to retain only those components that made the most sense at the theoretical level [33]. Internal consistency was assessed by calculating the Cronbach α coefficient [34]. All created indicators (patients' skills in self-management of their health, facilitators and obstacles to mHealth implementation, importance of involvement of trusted actors in health in the construction of mHealth apps and devices, usefulness of mHealth apps and devices certification, and importance of the involvement of health-related organizations and stakeholders in promoting the

use of mHealth in general medicine) were obtained by adding up the scores for each item in the indicator and converting them into a score from 0 to 100, with 100 representing the greatest perceived benefit, risk or barrier, involvement, or utility, as applicable.

Descriptive analyses of the variables in the sample were then performed, followed by a multivariate binomial logistic regression to investigate the willingness of GPs to prescribe mHealth apps and devices. To select the variables to be included in the multivariate model, we performed univariate logistic regressions, which made it possible to obtain the crude odds ratios (ORs) and their 95% CIs and the *P* value. Variables associated with a 20% *P* value threshold ($P<.20$) in the univariate analyses were retained in the final multivariate model [35]. Once the latter was established, we verified that there was no problem with multicollinearity, defined as a variance inflation factor greater than 2.5 [36]. To obtain the most efficient and parsimonious model reflecting our data, a stepwise selection combining forward and backward selection procedures was performed. The model with the lowest Akaike Information Criterion was retained. We compared this model with the starting model (variables significant at the 20% *P* value threshold in univariate analyses) using analysis of variance. The multiple logistic regression coefficients were presented as adjusted ORs with their 95% CIs. We tested interactions between GPs' gender and facilitators and obstacles in the model as it was shown in the general population that gender was a moderator between attitudes toward mHealth apps and the intention to use them [37]. To estimate the goodness of fit of the model, the McFadden pseudo-$R^2$ value was calculated. In addition, to assess the discrimination of the model, the area under the receiver operator characteristic curve was also determined. The level of significance for the multivariate model was set at the 5% *P* value threshold. Analyses were performed using RStudio (version 1.2.5033; RStudio Inc) [38].

## Results

### Sociodemographic Characteristics of the GPs and Their Patients

Among the 226 GPs who answered the first question of the survey, 174 (76.9%) fully completed the questionnaire. The study sample comprised thus 174 GPs. Almost two-thirds (112/174, 64.4%) were men, and the mean age was 45.1 (SD 13.0) years. Nearly half of the GPs (80/174, 45.9%) reported spending 40% or more of their consultation time promoting healthy behaviors. With regard to their patient base, 37.4% (65/174) of GPs declared having patients mainly aged between 45 and 69 years in the previous month. One-third (58/174, 32.8%) reported that their patient base was made up of people of different ages. Approximately, as many patients came from an urban setting as from a rural setting. Participating GPs estimated that, overall, their patient base had a middle socioeconomic status and quite a good command of the French language. However, in terms of self-management of their health, GPs perceived the skills of patients to be quite modest, with an average score of 49 (SD 15.3; Table 1).

**Table 1.** Sociodemographic characteristics of general practitioners (GPs) and their patients; characterization of GP practice and their perception of the self-management skills of patients in terms of health (N=174).

| Variables | Values |
|---|---|
| Age of GP (years), mean (SD) | 45.1 (13.0) |
| **Gender of GP, n (%)** | |
| Male | 112 (64.4) |
| Female | 62 (35.6) |
| **Time spent promoting good health behaviors during consultations (consultation %), n (%)** | |
| 0-20 | 45 (25.9) |
| 20-30 | 49 (28.2) |
| 30-50 | 49 (28.2) |
| 50-100 | 31 (17.8) |
| **Participation in a continued education program during the previous year, n (%)** | |
| No | 14 (8) |
| Yes | 160 (91.9) |
| **Participation in a peer group during the previous year, n (%)** | |
| No | 93 (53.4) |
| Yes | 81 (46.6) |
| **Subscription to a professional magazine, n (%)** | |
| No | 34 (19.5) |
| Yes | 140 (80.5) |
| **Part of a social network for physicians, n (%)** | |
| No | 105 (60.3) |
| Yes | 69 (39.7) |
| **Age of patients in the previous month (years), n (%)** | |
| 0-44 | 31 (17.8) |
| 45-69 | 65 (37.4) |
| 70 and older | 20 (11.5) |
| Other (not characterizable) | 58 (33.3) |
| Patient base place of residence (GPs perceived): urban setting (0)-rural setting (6), mean (SD) | 3.3 (2.1) |
| Patient base socioeconomic status (GPs perceived): low (0)-high (6), mean (SD) | 3.4 (1.5) |
| Patient base command of the French language (GPs perceived): poor (0)-excellent (6), mean (SD) | 4.6 (1.6) |
| Patient skills in self-management of their health[a] (GPs perceived; prevention behaviors, autonomous health management, and assessment of the reliability of information): low (0)-high (100), mean (SD) | 49 (15.3) |

[a]Internal consistency (Cronbach α)=.80.

## GPs and mHealth Apps and Devices

Participating GPs were more likely to have mHealth apps (mainly for mixed personal and professional use) than connected health and wellness devices (132/174, 75.9%, vs 84/174, 48.3%, respectively; Table 2).

**Table 2.** Participating general practitioners (GPs) and mobile health (mHealth) apps and devices (possession and perceptions; N=174).

| Variables | Values | Internal consistency (Cronbach α) |
|---|---|---|
| **Had a connected health or wellness device, n (%)** | | —[a] |
| No | 90 (51.7) | |
| Yes | 84 (48.3) | |
| **Had an mHealth app, n (%)** | | — |
| No | 42 (24.1) | |
| Yes | 132 (75.9) | |
| Facilitators: perceptions of GPs of the benefits of mHealth apps and devices for patients, caregivers, their own clinical practice, and GP-perceived drivers for mHealth apps and devices implementation in their medical practice[b], mean (SD) | 57.2 (16.6) | .91 |
| Obstacles: perceptions of GPs of risks for the patient and barriers for the GPs practice[b], mean (SD) | 54.1 (15.6) | .71 |
| Perceptions of GPs of the importance of the involvement of trusted actors in health in the construction of mHealth apps and devices[b], mean (SD) | 75.5 (19.8) | .76 |
| Perceptions of GPs of the usefulness of mHealth apps and devices certification[b], mean (SD) | 64.2 (15.3) | .71 |
| Perceptions of GPs of the importance of the involvement of health-related organizations and stakeholders in promoting the use of mHealth apps and devices in general medicine[b], mean (SD) | 64.6 (22.7) | .78 |
| Perceptions of GPs of the utility of validation of mHealth apps and devices using randomized studies (evidence-based medicine)[b], mean (SD) | 81.1 (21.4) | — |

[a]Cronbach α could not be estimated because of qualitative variables or a single quantitative item.

[b]Score ranging from 0 to 100, with 100 representing the greatest perceived benefit or driver, risk or barrier, involvement, or utility, as applicable.

## Perceptions of GPs of the Benefits and Drivers of mHealth Apps and Device Prescriptions and the Associated Risks and Barriers

GPs perceived as many benefits and potential drivers to mHealth apps and devices use by their patients (mean 57.2, SD 16.6) as they did risks and barriers (average score 54.1, SD 15.6, out of a possible score of 100; Table 2). More specifically, regarding benefits and potential drivers to mHealth devices implementation, the higher perception was that their patients would use mHealth apps and devices more if they recommended it (mean 5.3, SD 1.3, out of a possible score of 7), followed by the perception that mHealth apps and devices could strengthen the involvement of patients in the management of their health (average score 5.1, SD 1.1) and by an alternative to drug prescription (average score 4.9, SD 1.4). A wish for access to a software aid that could help them prescribe mHealth apps and devices (ie, software that would automatically suggest mHealth apps and devices adapted to the patient's needs) was expressed by GPs as facilitators (average score 4.9, SD 1.8). It is relevant to note that GPs reported a low level of knowledge regarding mHealth apps and devices (average score 3.0, SD 1.6; Figure S1 in Multimedia Appendix 3).

With regard to the perceived risks and barriers, their main concern (average score 5.5, SD 1.2, out of a possible score of 7) was that GPs must provide support in the use of mHealth apps and devices, meaning additional working time during and outside of consultations. This concern was followed by their fear that patient data would be used for commercial reasons (average score 5.4, SD 1.7). It should be noted that GPs shared low levels of concern regarding the risk of dehumanization of the patient–physician relationship (average score 3.3, SD 1.6; Figure S2 in Multimedia Appendix 3).

## GPs Perceptions of the Construction, Validation, and Certification of mHealth Apps and Devices

GPs reported a high importance of the implication of trusted actors in health in the construction of mHealth apps and devices (average score 75.5, SD 19.8, out of a possible score of 100). Precisely, GPs who participated in the study considered the involvement of physicians (average score 5.9, SD 1.3; out of a possible score of 7) and patients (average score 5.7, SD 1.4; out of a possible score of 7) in the construction and development of mHealth apps and devices content to be necessary. The average score was 5.0 (SD 1.5) when asked about the involvement of researchers, and the average score was 5.0 (SD 1; Figure S3 in Multimedia Appendix 3). GPs also underlined their strong belief that mHealth apps and devices should be clinically validated through randomized studies (average score 81.1, SD 21.4; out of 100) and obtain certification from trusted health actors (average score 64.2, SD 15.3; out of 100). More specifically, certification by independent experts, a college of physicians, or an ethics committee was necessary and even essential (average score 5.8 out of 7, SD 1.4 for the 3 items). University certification (average score 5.0, SD 1.5) or patients'

association certification (average score 4.8, SD 1.7) were also considered necessary. Conversely, GPs reported relatively unnecessary certification by private health companies (average score 1.9, SD 1.2; Figure S4 in Multimedia Appendix 3). The issues surrounding clinical validation and certification raise the question of financial implications. Implications of health-related organizations and stakeholders in promoting the use of mHealth apps and devices in general medicine were reported as an important issue (average score 64.6, SD 22.7; out of 100). Precisely, GPs considered it necessary to cover the costs of mHealth apps and devices by patient health care insurance or complementary health insurance firms (mean 5.0, SD 1.7; for both) and reported that it was necessary for health authorities to provide a financial incentive to GPs to prescribe mHealth apps and devices (average score 4.3, SD 2.0). The involvement of health authorities (French National Authority for Health) was clearly reported by GPs as a driver for the implementation of mHealth apps and devices in general practice, with an average score of 5.1 (SD 1.6) out of 7 (Figure S5 in Multimedia Appendix 3; Table 2).

## Willingness of GPs to Prescribe mHealth Apps and Devices

Of the 129 GPs, 97 (75.2%) were willing to prescribe mHealth apps and devices, that is, they were willing to prescribe mHealth apps and devices for at least one of the 12 health dimensions included in this study. More specifically, 60.5% (78/129) declared their willingness to prescribe mHealth apps and devices for physical activity, asthma and allergies, and vaccination; 79.8% (103/129) for diabetes; and 71.3% (92/129) for addictions (Multimedia Appendix 4).

## Univariate Binomial Logistic Regressions

Table 3 shows that at the 5% *P* value threshold, the following factors were associated with the willingness of GPs to prescribe mHealth apps and devices: having participated in a training program during the previous year (OR 6.20, 95% CI 2.01-21.28); having participated in a peer group during the previous year (OR 2.10, 95% CI 1.04-4.35); and GPs' perception that, overall, their patient base had a good command of the French language (OR 1.24 95% CI 1.01-1.51), facilitators of mHealth apps and devices implementation (perception of benefits and drivers; OR 1.04 95% CI 1.02-1.06). Conversely, GPs' age (OR 0.97, 95% CI 0.95-0.99) and their perception of risks for the patient and obstacles to their own practice (OR 0.97, 95% CI 0.95-0.99) were associated with unwillingness to prescribe mHealth apps and devices.

At the 20% *P* value threshold, subscribing to a professional journal (OR 1.77, 95% CI 0.77-3.91), owning a connected health or wellness device (OR 1.78, 95% CI 0.90-3.62), perceiving that the involvement of field-based actors in developing mHealth apps and devices is important (OR 1.02, 95% CI 1.00-1.03), and perceiving that validation of mHealth apps and devices by randomized studies is necessary (OR 1.01, 95% CI 1.00-1.03) were associated with greater willingness to prescribe mHealth apps and devices. In contrast, being a female GP (OR 0.60, 95% CI 0.30-1.21) was associated with the unwillingness to prescribe mHealth apps and devices (Table 3).

**Table 3.** Characteristics of general practitioners (GPs) and their patients, attitudes GPs have toward mobile health (mHealth) apps and devices, and their association with the willingness of GPs to prescribe mHealth apps and devices in univariate analyses (N=174).

| Variables | Unwilling to prescribe mHealth apps and devices (n=45) | Willing to prescribe mHealth apps and devices (n=129) | Crude OR[a] (95% CI) | *P* value |
|---|---|---|---|---|
| Age of GPs (years), mean (SD) | 48.7 (13.8) | 43.8 (12.6) | 0.97 (0.95-0.99) | .03 |
| **Gender of GPs, n (%)** | | | | .15 |
| Male | 25 (55.6) | 87 (67.4) | Ref[b] | |
| Female | 20 (44.4) | 42 (32.6) | 0.60 (0.30-1.21) | |
| **Time spent promoting healthy behaviors during consultations (% of consultation), n (%)** | | | | .24 |
| 0-20 | 9 (20) | 36 (27.9) | Ref | |
| 20-30 | 16 (35.6) | 33 (25.6) | 0.52 (0.19-1.30) | |
| 30-50 | 15 (33.3) | 34 (26.4) | 0.57 (0.21-1.45) | |
| 50-100 | 5 (11.1) | 26 (20.1) | 1.30 (0.40-4.65) | |
| **Participation in a continued education program during the previous year, n (%)** | | | | .002 |
| No | 9 (20) | 5 (3.9) | Ref | |
| Yes | 36 (80) | 124 (96.1) | 6.20 (2.01-21.28) | |
| **Participation in peer group during the previous year, n (%)** | | | | .04 |
| No | 30 (66.7) | 63 (48.8) | Ref | |
| Yes | 15 (33.3) | 66 (51.2) | 2.10 (1.04-4.35) | |
| **Subscription to a professional magazine. n (%)** | | | | .17 |
| No | 12 (26.7) | 22 (17.1) | Ref | |
| Yes | 33 (73.3) | 107 (82.9) | 1.77 (0.77-3.91) | |
| **Part of a social network for physicians, n (%)** | | | | .77 |
| No | 28 (62.2) | 77 (59.7) | Ref | |
| Yes | 17 (37.8) | 52 (40.3) | 1.11 (0.56-2.27) | |
| **Age of patient base in the previous month (years), n (%)** | | | | .71 |
| 0-45 | 7 (15.6) | 24 (18.6) | 1.12 (0.42-3.24) | |
| 45-70 | 16 (35.5) | 49 (38) | Ref | |
| ≥70 | 4 (8.9) | 16 (12.4) | 1.31 (0.41-5.06) | |
| Other (not characterizable) | 18 (40) | 40 (31) | 0.73 (0.33-1.60) | |
| Patient base place of residence (GP perceived): urban setting (0)-rural setting (6), mean (SD) | 3.6 (2.3) | 3.2 (2.1) | 0.92 (0.78-1.08) | .32 |
| Patient base socioeconomic status (GP perceived): low (0)-high (6), mean (SD) | 3.2 (1.6) | 3.5 (1.4) | 1.14 (0.91-1.44) | .25 |
| Patient base command of the French language (GP perceived): poor (0)-excellent (6), mean (SD) | 4.2 (1.8) | 4.8 (1.5) | 1.24 (1.01-1.51) | .04 |
| Patient skills in self-management of their health (GPs perceived): low (0)-high (100), mean (SD) | 48.8 (14.7) | 49 (15.6) | 1.00 (0.98-1.02) | .93 |
| **Had a connected health or wellness device, n (%)** | | | | .10 |
| No | 28 (62.2) | 62 (48.1) | Ref | |
| Yes | 17 (37.8) | 67 (51.9) | 1.78 (0.90-3.62) | |
| **Had an mHealth app, n (%)** | | | | .39 |
| No | 13 (28.9) | 29 (22.5) | Ref | |
| Yes | 32 (71.1) | 100 (77.5) | 1.40 (0.64-2.98) | |

| Variables | Unwilling to prescribe mHealth apps and devices (n=45) | Willing to prescribe mHealth apps and devices (n=129) | Crude OR[a] (95% CI) | *P* value |
|---|---|---|---|---|
| Facilitators: perceptions of GPs of the benefits of mHealth apps and devices for patients, caregivers, their own clinical practice, and GP-perceived drivers for mHealth apps and devices implementation in their medical practice[c], mean (SD) | 49.6 (18.1) | 59.9 (15.2) | 1.04 (1.02-1.06) | <.001 |
| Obstacles: perceptions of GPs of risks for the patient and barriers for the GPs practice[c], mean (SD) | 59 (15.8) | 52.4 (15.3) | 0.97 (0.95-0.99) | .02 |
| Perceptions of GPs of the usefulness of mHealth apps and devices certification[c], mean (SD) | 63 (15.4) | 64.6 (15.3) | 1.01 (0.98-1.03) | .54 |
| Perceptions of GPs of the importance of the involvement of health-related organizations and stakeholders in promoting the use of mHealth apps and devices in general medicine[c], mean (SD) | 61.5 (21.3) | 65.7 (23.2) | 1.01 (0.99-1.02) | .29 |
| Perceptions of GPs of the importance of the involvement of field-based actors in the construction of mHealth apps and devices[c], mean (SD) | 70.7 (19.8) | 77.2 (19.6) | 1.02 (1.00-1.03) | .06 |
| Perceptions of GPs of the utility of validating mHealth apps and devices using randomized studies (evidence-based medicine)[c], mean (SD) | 76.3 (24.7) | 82.8 (19.9) | 1.01 (1.00-1.03) | .08 |

[a]OR: odds ratio.

[b]Ref: reference.

[c]Scores ranging from 0 to 100, with 100 representing the greatest perceived benefit or driver, risk or barrier, involvement, or utility, as applicable.

## Multivariate Binomial Logistic Regression

Table 4 shows the results of the multivariate binomial logistic regression after stepwise selection. Factors associated with the willingness of GPs to prescribe mHealth apps and devices were as follows: having attended a continued education program during the previous year (OR 6.17, 95% CI 1.52-28.72); their perception that overall their patient base has a good command of the French language (OR 1.45, 95% CI 1.13-1.88); their perception that mHealth apps and devices could bring benefits to the patient and their own medical practice; and their perception of drivers for mHealth apps and devices implementation in their medical practice (OR 1.04, 95% CI 1.01-1.07); and a strong perception of the importance of validating mHealth apps and devices through randomized studies (evidence-based medicine; OR 1.02, 95% CI 1.00-1.04). In contrast, older age (OR 0.95, 95% CI 0.91-0.98), being a female GP (OR 0.26, 95% CI 0.09-0.69), and the perception of greater risks for the patient and barriers to their own medical practice (OR 0.96, 95% CI 0.94-0.99) were associated with the unwillingness of GPs to prescribe mHealth apps and devices (Table 4). This model had a very good fit and prediction properties, with an area under the receiver operator characteristic curve of 0.825 (excellent classification performance) and a McFadden pseudo-$R^2$ value of 0.28 (very good fit). Interactions between gender and facilitators and obstacles were not significant.

**Table 4.** Results of multivariate logistic regression (after the stepwise procedure) to explain the willingness of general practitioners (GPs) to prescribe mobile health (mHealth) apps and devices to their patients (N=174).

| Variables | aOR[a] (95% CI) | *P* value |
|---|---|---|
| Age of GPs (years) | 0.95 (0.91-0.98) | .003 |
| **Gender of GPs** | | |
| Male | Ref[b] | N/A[c] |
| Female | 0.26 (0.09-0.69) | .008 |
| **Participation in a continued training program during the previous year** | | |
| No | Ref | N/A |
| Yes | 6.17 (1.52-28.72) | .01 |
| **Participation in a peer group during the previous year** | | |
| No | Ref | N/A |
| Yes | 2.32 (0.94-6.01) | .07 |
| **Subscription to a professional magazine** | | |
| No | Ref | N/A |
| Yes | 2.41 (0.85-6.89) | .10 |
| Patient base command of the French language (GP perceived) | 1.45 (1.13-1.88) | .004 |
| Facilitators: perceptions of GPs of benefits of mHealth apps and devices for patients, caregivers, their own clinical practice, and GP-perceived drivers for mHealth apps and devices implementation in their medical practice | 1.04 (1.01-1.07) | .003 |
| Obstacles: perceptions of GPs of risks for the patient and barriers for the GP practice | 0.96 (0.94-0.99) | .01 |
| Perceptions of GPs of the importance of validating mHealth apps and devices using randomized studies (evidence-based medicine) | 1.02 (1.00-1.04) | .047 |

[a]aOR: adjusted odds ratio.

[b]Ref: reference.

[c]N/A: not applicable.

# Discussion

## Principal Findings

The objective of this study is to understand the factors influencing the willingness of French GPs to prescribe mHealth apps and devices. Our results highlighted that the several factors involved were as follows: sociodemographic characteristics of GPs, especially age and gender; factors linked to continued education, patient base–related characteristics, especially perceptions of GPs of their patient base's command of the French language; and factors linked to perceptions of GPs regarding mHealth apps and devices (benefits, drivers, risks, and barriers) and the perceived importance of clinical validation.

The health-related dimensions for which GPs were very willing to prescribe mHealth apps and devices were diabetes, asthma (ie, chronic diseases) and addictions, physical activity, and vaccination (ie, primary prevention) reflecting findings in the literature [16,18]. GPs were more likely to be willing to prescribe mHealth apps and devices in medical fields where they are already numerous and mHealth apps and devices that have already been clinically validated [13]. This result underlines the feasibility of the potential prescription of mHealth apps and devices. On the contrary, this finding may not be very revealing in terms of therapeutic areas where patients might need the most support.

This study reflects the previous findings about the importance of GP-perceived sociodemographic profiles of patients as a parameter in determining the integration of mHealth apps and devices into current clinical practice [22,39]. An Australian study showed that GPs who had been working longer were less willing to prescribe mHealth apps and devices [16], and a German study corroborated this fact as they showed that younger GPs saw mHealth apps more favorable [28], which reflects our findings here. One possible explanation for this finding is that younger GPs are more technologically savvy. In that sense, a French Barometer survey showed that physicians tended to prefer prescribing mHealth apps and devices to adolescent patients, professionally active patients, and to technologically savvy patients [39]. Furthermore, the fact that female GPs in this study were less willing to prescribe mHealth apps and devices than their male counterparts reflects the importance of the issue of gender in the appropriation of new technologies. Indeed, several studies on general populations have already shown that men are more inclined to use mHealth apps and devices than women [40,41], even if this association remains unclear, as other studies have shown that women are more inclined to use mHealth apps and devices [42-44]. However, we found no information that specifically concerned physicians

regarding this issue. Concerning the ease of communication GPs have with their patients, in this study, we found that GPs who perceived their patient base to have a good command of French were more willing to prescribe them mHealth apps and devices. Several studies have highlighted the importance of a patient's digital health literacy level and the difficulties faced in implementing mHealth apps and devices for both patients with low levels of eHealth literacy [16,17] and patients showing reluctance [13,17]. These findings highlight the possibility of a second-order digital divide, which is not a divide in terms of access to the internet and smartphones, rather a divide in the use of mHealth apps and devices, which in turn can widen the gap in health inequalities [45,46]. In contrast, patients' good command of the French language could be a predictor of their ability to use mHealth apps and devices and therefore may influence the decision of GPs to prescribe these types of interventions.

Our study highlights the important role that individual perceptions of mHealth apps and devices play in the willingness of GPs to use them in clinical practice. More specifically, our multivariate model highlighted that perceptions of the benefits, drivers, risks, and barriers of mHealth apps and devices were linked to the willingness of GPs to prescribe them, confirming the results of the qualitative study, which served as the basis for the construction of the questionnaire [23]. Consistent with our study, a Malaysian study found that GP-perceived benefits of mHealth apps (performance expectancy) were associated with the willingness of GPs to recommend them to their patients [27]. The aforementioned Australian study indicated different ways to encourage GPs to adopt mHealth apps and devices [16]. In particular, the need for training in mHealth apps and devices, the possibility of obtaining a list of safe and effective mHealth apps and devices that have been validated by a health authority, and access to detailed descriptions of mHealth apps and devices. Thus, only 22% of German GPs felt able to advise mHealth apps to their patients [28]. These results reflect the findings of this study, as GPs reported having a low level of knowledge regarding mHealth devices and a high perception level regarding the usefulness of a specific software aid that could help them when prescribing mHealth apps and devices [16].

### The Need for an mHealth App and Device Prescription Software Aid

Public and private initiatives have led to the creation of comprehensive lists of a number of mHealth apps and devices currently available. In the United Kingdom, 2 initiatives have been implemented to create a library of health apps to help patients navigate the various options available to them. First, PatientView was developed in 2013 by user groups and incorporated a visible app user rating system [47]. Second, the National Health Service Apps Library is a nationwide initiative developed by the National Health Service [48]. In the United States, an independent, private platform was created in 2009 for health care professionals based on the experience and opinions of their peers [49]. This platform led to the development of a specific app, called iPrescribeApps, which aids physicians in prescribing suitable mHealth apps and devices for their patient-specific medical conditions [50]. In Catalonia, we can also notice the platform AppSalut that references mHealth apps that have obtained accreditation in terms of technology, usability, security, and reliability (regarding medical content) [5].

An Australian before-and-after intervention pilot study aimed to investigate the feasibility of prescribing mHealth apps in general practice. The 36 GPs included were given a prescription guide for 6 apps (description of the app, download instructions, and cost). Video presentations for the 6 apps can also be found using the download instructions. After 2 months, the video presentation of one of the apps, randomly selected, was sent to each physician in the study just to remind the GP. The median number of apps prescribed before and after the intervention was almost quadrupled. However, the video presentations were not associated with this increase, highlighting the importance of having a prescription guide to prescribe mHealth apps and devices [51].

Given our study's findings and the initiatives and the literature on mHealth apps and devices described earlier, it would appear that there is great demand by GPs for an mHealth apps and devices prescription software aid. Such an aid would represent a real driver for the implementation of mHealth apps and devices in medical practice in France and would merit being developed, provided that clinical evaluation criteria of health-related organizations and stakeholders and the protection of personal data were considered.

### The Need for Training in mHealth

In this study, GPs reported not being sufficiently familiar with mHealth apps and devices. Training in mHealth would provide GPs with sufficient knowledge and confidence to prescribe mHealth apps and devices [52]. This issue was raised in a Dutch study [17], which included 621 GPs. Almost half of the participants declared their desire for remote learning (webinars, podcasts, etc) compared with only 12% who preferred face-to-face training [17]. It would be interesting to interview French GPs about this issue and ask their opinions about which training content would be most suitable to help them integrate mHealth into their medical practice. Governments, health systems, and authorities should provide digital health education to GPs [52] via continuing education programs and medical curricula. Training could be provided jointly by health authority mHealth referents, mHealth referent GPs, mHealth researchers, and developers of mHealth app and device national libraries.

### Issues Surrounding Certification, Data Privacy, and Development of mHealth Apps and Devices

As shown here and highlighted in several other studies, GPs are concerned about the protection of personal data and the reliability of mHealth apps and devices [13,17,21,22,28]. Indeed, the willingness and unwillingness of GPs to prescribe mHealth apps and devices reported the importance of certifying mHealth apps and devices by independent public bodies and the irrelevance, in their opinion, of certification by private health companies. This finding is corroborated by a study of different French physician organizations that found that three-quarters of those questioned reported that they trusted certification by a learned society or a health authority as opposed to only 2% who trusted certification by a private company [39]. In this study,

GPs strongly expressed the need for field-based actors (patients, physicians, and academic researchers) to be involved in the development of mHealth apps and devices. However, this factor was not significantly associated with the willingness of GPs to prescribe mHealth apps and devices in the multivariate analysis.

### The Need for Clinical Validation of mHealth Apps and Devices

For reimbursement by health insurance to become a possibility, it is essential that clinical validation—ideally by randomized studies (evidence-based medicine)—be performed. In this study, clinical validation appeared to be an essential element in the willingness of GPs to prescribe mHealth apps and devices. A 2018 overview of systematic reviews of randomized clinical trials focusing on mHealth apps showed that only 22 apps, most focusing on diabetes, obesity, and mental health, had been clinically validated. However, most of these 22 apps were clinically validated in pilot studies with small sample sizes, thereby limiting the validity of the results [18]. Clinical validation of mHealth apps and devices is a real challenge and deserves to be integrated in a more systematic fashion in health research projects.

### Issues Surrounding Care, Compensation, and Financial Incentives in Terms of mHealth Apps and Devices

The notion of covering the cost of mHealth apps and devices through health insurance of patients was an important point for the GPs in this study, as was the possibility of health authorities providing financial incentives for GPs to prescribe mHealth apps and devices. However, neither element was directly associated with the willingness of GPs to prescribe mHealth apps and devices in the multivariate analysis. In a descriptive way, our results showed that having to provide support to patients in the use of mHealth apps and devices—thereby leading to a longer working time—appeared to be the major perceived obstacle perceived by GPs. This reflects the literature that mentions the desire of GPs for financial compensation for the time spent (during and outside consultation) both processing information coming from mHealth apps and devices and training themselves and their patients in the use of mHealth apps and devices [13,16,17,22]. Studies have also reported the problem of the costs of mHealth apps and devices [13,16] and the lack of reimbursement [13] for these costs as obstacles to the prescription of mHealth apps and devices.

### Limitations

We decided to oppose, from the perspective of behavior change, in this study 2 profiles of GPs—those willing to mHealth apps and devices prescription and those not as in France prescription of mHealth apps and devices, especially in general medicine, which is not integrated in current practice. Then the variable willingness to prescribe was dichotomized; thus, we lost the information regarding the amount of mHealth apps and devices that GPs were willing to prescribe.

At the epistemological level, this study adopts a comprehensive approach that focuses on understanding the psychosocial processes involved in the initiation or noninitiation of a behavior and the meaning that individuals give to it. This approach is important for understanding behavior toward a phenomenon (in our case, the willingness or unwillingness to prescribe mHealth apps and devices). We did not base our study on registers or sampling techniques that ensure the representativeness of the French population of GPs. However, with the comprehensive approach, this study provides interesting elements to better understand the obstacles and facilitators of GPs' willingness to prescribe mHealth apps and devices to their patients. The study was not intended to be representative but sought to confirm the role of various factors associated with the willingness of GPs to prescribe mHealth apps and devices. Although the ratio of male to female GPs in this study reflects national numbers, GPs in this study were a little younger (mean age 45.1 (SD 13) vs 50.4 years at the national level) [53] and the patient base seemed to also be little younger compared with French national figures [54]. This may have resulted in a slight overestimation of GPs' willingness to prescribe mHealth apps and devices. In addition, our sample size was relatively small, and we were unable to obtain the response rate given our methodology for administering the questionnaires, which may question the representativeness of the responding GPs. In this study, we grouped willingness (or unwillingness) to prescribe mHealth apps or connected health and wellness devices in the same indicator, which could be interesting in further studies to investigate if there are differences in factors associated with mHealth apps prescription and connected health and wellness devices prescription. The willingness (or unwillingness) to prescribe mHealth apps and devices grouped several health categories, and further studies should be conducted to investigate whether the identified factors differ between these different health conditions.

Given our sample size and principal component analysis, we created indicators that aggregated several perceptions GPs have toward mHealth apps and devices, but we cannot identify the individual factors that have a significant impact. In France, the prescription of mHealth apps and devices is not integrated in clinical routine; we then investigated obstacles and facilitators perceived rather than experienced. Further studies need to be conducted after the implementation of mHealth apps and devices in general medicine to investigate obstacles and facilitators experienced. In this study, we focused on the perceptions of GPs, as they are the essential link in the patient's care pathway. Compared with GPs, it could also be interesting to investigate the perceptions of specialist physicians, as it can be assumed that they may have a different practice and a different relationship with their patients.

## Conclusions

To conclude, mHealth apps and devices represent an important dimension in general practice consultations that can complement other GP treatment methods. GPs in this study seemed inclined to fully integrate mHealth apps and devices into their practice, especially if they have access to tools to help them navigate their way in the field of digital health, similar to those that already exist for the prescription of drugs. Such tools should provide information on the benefits of mHealth apps and devices both for the GP practice and for the patient; the pros and cons of mHealth apps and devices; and data on how mHealth apps and devices are developed, validated, and certified.

Public authority–based initiatives for the certification of mHealth apps and devices are very important for mHealth apps and devices to become accepted in general medicine and must be extensively implemented. Clinical validation of mHealth apps and devices through scientific studies needs to be performed on a larger scale, not only with pilot studies. Indeed, validation should be integrated more systematically into health research projects. Training courses specifically designed to provide support GPs in fully integrating mHealth apps and devices into their practice are also indispensable. Such training, in turn, would ensure that GPs could provide the best support possible in terms of mHealth apps and devices use to their patients.

## Acknowledgments

The authors would like to thank the *Agence Nationale de la Recherche* for its support (grant ANR-17-CE19-0027; project ApiAppS). They would also like to thank all the general practitioners who agreed to participate in this study and Jude Sweeney (Milan, Italy) for translating and copyediting the English version of this manuscript.

## Authors' Contributions

All authors contributed to the study conception and design. The ApiAppS questionnaire was designed and developed by M Préau, M Pannard, CB, and TL, together with the ANR ApiAppS group. The questionnaire was administered on the web by Amandine Andrin. The analyses were performed by CDV who also wrote the first draft of the manuscript. All authors commented on the previous versions of the manuscript and read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys.
[DOCX File , 26 KB - mhealth_v10i2e28372_app1.docx ]

Multimedia Appendix 2
The ApiAppS questionnaire.
[DOCX File , 251 KB - mhealth_v10i2e28372_app2.docx ]

Multimedia Appendix 3
Perceptions of general practitioners regarding mobile health.
[DOCX File , 215 KB - mhealth_v10i2e28372_app3.docx ]

Multimedia Appendix 4
Proportion of general practitioners (GPs) willing to prescribe mHealth apps and devices according to the 12 different health dimensions included in this study (N=129; GPs who declared their willingness to prescribe mHealth apps and devices).
[DOCX File , 108 KB - mhealth_v10i2e28372_app4.docx ]

## References

1. mHealth-New horizons for health through mobile technologies. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2021-05-15]
2. Digital Health Trends 2021 : innovation, evidence, regulation, and adoption. IQVIA Institute. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2021-05-15]
3. Internet of things in healthcare market size, share and trends analysis. Grand View Research. 2019 Mar. URL: https://www.grandviewresearch.com/industry-analysis/internet-of-things-iot-healthcare-market [accessed 2021-12-31]
4. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020 Jul 30;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
5. Segui FL, Bufill CP, Gimenez NA, Roldan JM, Cuyas FG. The prescription of mobile apps by primary care teams: a pilot project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701 [FREE Full text] [doi: 10.2196/10701] [Medline: 29929948]
6. Référentiels métiers et compétences. Médecins généralistes, sages-femmes et gynécologues- obstétriciens. Berger-Levrault. 2010. URL: https://www.cnge.fr/les_productions_scientifiques/referentiels_metiers_et_competences/ [accessed 2021-12-31]
7. Hummers-Pradier E, Beyer M, Chevallier P, Eilat-Tsanani S, Lionis C, Peremans L, et al. Series: the research agenda for general practice/family medicine and primary health care in Europe. Part 2. Results: primary care management and community

orientation. Eur J Gen Pract 2010 Mar;16(1):42-50 [FREE Full text] [doi: 10.3109/13814780903563725] [Medline: 20100109]
8. Green LA, Fryer GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. N Engl J Med 2001 Jun 28;344(26):2021-2025. [doi: 10.1056/nejm200106283442611] [Medline: 11430334]
9. Les français et le médecin généraliste. Opinionway pour Medaviz. 2018. URL: https://www.medaviz.com/les-francais-et-le-medecin-generaliste-opinionway/ [accessed 2021-05-15]
10. Bourgueil Y, Marek A, Mousquès J. Trois modèles types d?organisation des soins primaires en Europe, au Canada, en Australie et en Nouvelle-Zélande. Irdes. 2009 Apr. URL: https://www.irdes.fr/Publications/Qes/Qes141.pdf [accessed 2021-05-15]
11. Dufour J, Grosjean J, Darmoni S, Yasini M, Marchand G, Simon C, et al. ApiAppS: A project to study and help practitioners in recommending mhealth apps and devices to their patients. Stud Health Technol Inform 2019 Aug 21;264:1919-1920. [doi: 10.3233/SHTI190713] [Medline: 31438407]
12. Santé connectée - De la e-santé à la santé connectée -Le livre blanc du Conseil national de l'Ordre des médecins. Conseil national de l'Ordre des médecins. 2015 Jan. URL: https://www.conseil-national.medecin.fr/sites/default/files/external-package/edition/lu5yh9/medecins-sante-connectee.pdf [accessed 2021-05-15]
13. El Amrani L, Oude Engberink A, Ninot G, Hayot M, Carbonnel F. Connected health devices for health care in french general medicine practice: cross-sectional study. JMIR Mhealth Uhealth 2017 Dec 21;5(12):e193 [FREE Full text] [doi: 10.2196/mhealth.7427] [Medline: 29269336]
14. Gérer mes ordonnances. Conseil national de l'Ordre des médecins. URL: https://www.conseil-national.medecin.fr/medecin/exercice/gerer-ordonnances [accessed 2021-05-15]
15. Fainzang S. Les patients et leurs ordonnances. In: Médicaments et Société. France: Presses Universitaires de France; 2001:23-40.
16. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
17. Peeters JM, Krijgsman JW, Brabers AE, Jong JD, Friele RD. Use and uptake of eHealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]
18. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018 May 9;1(1):12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
19. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Jan 24:439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
20. Senn N, Santiago-Delefosse M. Objets connectés de santé: patients et médecins déconnectés ? Rev med suisse. 2018 Dec 05. URL: https://www.ne.ch/autorites/DFS/SCSP/prevention/Documents/SantePerso_pdf_complet.pdf [accessed 2021-05-15]
21. Mutebi I, Devroey D. Perceptions on mobile health in the primary healthcare setting in Belgium. Mhealth 2018;4:44 [FREE Full text] [doi: 10.21037/mhealth.2018.09.01] [Medline: 30363764]
22. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
23. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
24. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: a mixed method study. Health Policy Technol 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
25. Ramdani B, Duan B, Berrou I. Exploring the determinants of mobile health adoption by hospitals in china: empirical study. JMIR Med Inform 2020 Jul 14;8(7):e14795 [FREE Full text] [doi: 10.2196/14795] [Medline: 32459630]
26. Wu J, Wang S, Lin L. Mobile computing acceptance factors in the healthcare industry: a structural equation model. Int J Med Inform 2007 Jan;76(1):66-77. [doi: 10.1016/j.ijmedinf.2006.06.006] [Medline: 16901749]
27. Lim H, Dunn A, Ooi SM, Teo C, Abdullah A, Woo W, et al. mHealth adoption among primary care physicians in Malaysia and its associated factors: a cross-sectional study. Fam Pract 2021 Jun 17;38(3):210-217 [FREE Full text] [doi: 10.1093/fampra/cmaa103] [Medline: 32968795]
28. Wangler J, Jansky M. The use of health apps in primary care-results from a survey amongst general practitioners in Germany. Wien Med Wochenschr 2021 Feb 11;171(7):148-156. [doi: 10.1007/s10354-021-00814-0] [Medline: 33570692]
29. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
30. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights 2013;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
31. Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960 Apr 01;20(1):141-151. [doi: 10.1177/001316446002000116]

32. Cattell RB. The scree test for the number of factors. Multiv Behav Res 1966 Apr 01;1(2):245-276. [doi: 10.1207/s15327906mbr0102_10] [Medline: 26828106]
33. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30:179-185. [doi: 10.1007/BF02289447] [Medline: 14306381]
34. Bland JM, Altman DG. Cronbach's alpha. Br Med J 1997 Feb 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
35. Aminot I, Damon MN. The use of logistic regression in the analysis of data concerning good medical practice. Revue Médicale de l'Assurance Maladie 2002 Apr;33:137-143 [FREE Full text] [doi: 10.1201/9781420057225-10]
36. Allison P. Multiple Regression: A Primer. Thousand Oaks, CA: SAGE Publications; Dec 1998.
37. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
38. RStudio Team. RStudio: Integrated Development for R. 2019. URL: http://www.rstudio.com/ [accessed 2021-12-31]
39. 4ème baromètre?: L'utilisation des smartphones par les médecins. VIDAL C. 2016 Mar. URL: https://www.vidalfrance.com/wp-content/download/info/Barometre_Mobile-VIDAL-CNOM-2016.pdf [accessed 2021-05-15]
40. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
41. Haluza D, Wernhart A. Does gender matter? Exploring perceptions regarding health technologies among employees and students at a medical university. Int J Med Inform 2019 Oct;130:103948. [doi: 10.1016/j.ijmedinf.2019.08.008] [Medline: 31442846]
42. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: user preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
43. Rafiullah M, David SK. Health apps usage and preferences among Saudi patients with diabetes: a survey. Int J Clin Pract 2019 May;73(5):e13345. [doi: 10.1111/ijcp.13345] [Medline: 30884059]
44. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of health app use and perceived effectiveness in people with cardiovascular diseases and diabetes: population-based survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179 [FREE Full text] [doi: 10.2196/12179] [Medline: 30920383]
45. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
46. Kumar D, Hemmige V, Kallen MA, Giordano TP, Arya M. Mobile phones may not bridge the digital divide: a look at mobile phone literacy in an underserved patient population. Cureus 2019 Feb 20;11(2):e4104. [doi: 10.7759/cureus.4104] [Medline: 31057998]
47. my health apps, tried and tested by people like you. PatientView. URL: http://myhealthapps.net/# [accessed 2021-05-15]
48. NHS Choices Health Apps Library. URL: https://www.nhs.uk/apps-library/ [accessed 2021-05-15]
49. iMedicalApps - Physician reviews of medical apps, health apps, fitness apps, and wearables. URL: https://www.imedicalapps.com [accessed 2021-05-15]
50. iPrescribeApps. URL: https://iprescribeapps.com [accessed 2021-05-15]
51. Byambasuren O, Beller E, Hoffmann T, Glasziou P. mHealth app prescription in Australian general practice: pre-post study. JMIR Mhealth Uhealth 2020 Jun 01;8(6):e16497 [FREE Full text] [doi: 10.2196/16497] [Medline: 32478660]
52. Moss RJ, Süle A, Kohl S. eHealth and mHealth. Eur J Hosp Pharm 2019 Jan;26(1):57-58 [FREE Full text] [doi: 10.1136/ejhpharm-2018-001819] [Medline: 31157099]
53. Atlas de la démographie médicale en France. Conseil national de l'Ordre des médecins. 2019. URL: https://www.conseil-national.medecin.fr/sites/default/files/external-package/analyse_etude/npcskp/demo_2019_atlas.pdf [accessed 2021-05-15]
54. Labarthe G. Direction de la Recherche des Etudes de l'Evaluation et des Statistiques (drees). Les consultations et visites des médecins généralistes - Un essai de typologie. 2004. URL: https://bdsp-ehesp.inist.fr/vibad/index.php?action=getRecordDetail&idt=304637 [accessed 2021-05-15]


## Abbreviations

**GP:** general practitioner
**mHealth:** mobile health
**OR:** odds ratio

*Edited by L Buis; submitted 03.03.21; peer-reviewed by I Berrou, A MacIsaac; comments to author 04.05.21; revised version received 25.06.21; accepted 27.08.21; published 11.02.22.*

*Please cite as:*
*Della Vecchia C, Leroy T, Bauquier C, Pannard M, Sarradon-Eck A, Darmon D, Dufour JC, Preau M*
*Willingness of French General Practitioners to Prescribe mHealth Apps and Devices: Quantitative Study*
*JMIR Mhealth Uhealth 2022;10(2):e28372*
*URL: https://mhealth.jmir.org/2022/2/e28372*
*doi:10.2196/28372*
*PMID:35147508*

Original Paper

# Evaluation of myCOPD Digital Self-management Technology in a Remote and Rural Population: Real-world Feasibility Study

Rowena Cooper[1*], MBBS; Adam Giangreco[2*], MSc, PhD; Michelle Duffy[3], RGN; Elaine Finlayson[3], BSc, RGN; Shellie Hamilton[3], RGN; Mahri Swanson[3], MSc, RGN; Judith Colligan[1], BSc; Joanna Gilliatt[4], BA, DPhil; Mairi McIvor[4], BSc; Elizabeth Kathryn Sage[1,2,5], MBBS, PhD

[1]Department of Respiratory Medicine, National Health Service Highland, Inverness, United Kingdom

[2]Division of Biomedical Sciences, University of the Highlands and Islands, Inverness, United Kingdom

[3]Specialist Community Respiratory Nursing Team, National Health Service Highland, Inverness, United Kingdom

[4]Technology Enabled Care Team, National Health Service Highland, Inverness, United Kingdom

[5]Institute for Applied Health Sciences, Centre for Rural Health, University of Aberdeen, Inverness, United Kingdom

[*]these authors contributed equally

**Corresponding Author:**
Elizabeth Kathryn Sage, MBBS, PhD
Department of Respiratory Medicine
National Health Service Highland
Raigmore Hospital
Old Perth Road
Inverness, IV2 3UJ
United Kingdom
Phone: 44 463 706294 ext 8823
Email: beth.sage@nhs.scot

## Abstract

**Background:** Chronic obstructive pulmonary disease (COPD) is a common, costly, and incurable respiratory disease affecting 1.2 million people in the United Kingdom alone. Acute COPD exacerbations requiring hospitalization place significant demands on health services, and the incidence of COPD in poor, remote, and rural populations is up to twice that of cities.

**Objective:** myCOPD is a commercial, digital health, self-management technology designed to improve COPD outcomes and mitigate demands on health services. In this pragmatic real-world feasibility study, we aimed to evaluate myCOPD use and its clinical effectiveness at reducing hospitalizations, inpatient bed days, and other National Health Service (NHS) resource use.

**Methods:** myCOPD engagement and NHS resource use was monitored for up to 1 year after myCOPD activation and was compared against health service use in the year prior to activation. A total of 113 participants from predominantly remote and rural communities were recruited via community-based care settings, including scheduled home visits, outpatient appointments, pulmonary rehabilitation, and phone or group appointments. There were no predetermined age, disease severity, geographical, or socioeconomic inclusion or exclusion criteria.

**Results:** Out of 113 participants, 89 activated myCOPD (78.8%), with 56% (50/89) of those participants doing so on the day of enrollment and 90% (80/89) doing so within 1 month. There was no correlation between participant enrollment, activation, or myCOPD engagement and either age, socioeconomics, rurality, or COPD severity. Most active participants used at least one myCOPD module and entered their symptom scores at least once (79/89, 89%). A subgroup (15/89, 17%) recorded their symptom scores very frequently (>1 time every 5 days), 14 of whom (93%) also used four or five myCOPD modules. Overall, there were no differences in hospital admissions, inpatient bed days, or other health service use before or after myCOPD activation, apart from a modest increase in home visits. Subgroup analysis did, however, identify a trend toward reduced inpatient bed days and hospital admissions for those participants with very high myCOPD usage.

**Conclusions:** Our results suggest that neither age, wealth, nor geographical location represent significant barriers to using myCOPD. This finding may help mitigate perceived risks of increased health inequalities associated with the use of digital health technologies as part of routine care provision. Despite high levels of activation, myCOPD did not reduce overall demands on health services, such as hospital admissions or inpatient bed days. Subgroup analysis did, however, suggest that very high myCOPD usage was associated with a moderate reduction in NHS resource use. Thus, although our study does not support implementation

of myCOPD to reduce health service demands on a population-wide basis, our results do indicate that highly engaged patients may derive benefits.

*(JMIR Mhealth Uhealth 2022;10(2):e30782)* doi:10.2196/30782

**KEYWORDS**

digital self-management; COPD; remote and rural; mobile health; application; chronic pulmonary obstructive disease; rural communities

## Introduction

Chronic obstructive pulmonary disease (COPD) is a common, costly, and incurable respiratory disease affecting 1.2 million people in the United Kingdom alone. Annually, it costs the National Health Service (NHS) £1.9 billion (US $2.43 billion), it requires over 1 million inpatient bed days due to acute exacerbations requiring hospitalization, and it places significant demands on health services [1-3]. The prevalence of COPD in poor, remote, and rural populations is twice that of cities [4].

Effective COPD self-management can reduce both exacerbation-induced hospital admissions and health service use when compared to standard care [5,6]. myCOPD is a digital health self-management technology designed to improve COPD outcomes and mitigate demands on services [7]. myCOPD modules include symptom scoring, inhaler technique, and a virtual pulmonary rehabilitation course. Previous studies indicate that myCOPD is associated with reduced inhaler technique errors, lower COPD Assessment Test (CAT) scores, and fewer hospital readmissions within 3 months of an exacerbation [8-11].

NHS Highland covers the largest geographical area of Scotland, contains regions of significant socioeconomic deprivation, and has a majority remote and rural population. Access to hospital services and delivering equity of care remains challenging, and digital health technologies represent one potential solution. In this pragmatic test-of-change study, we evaluated myCOPD and its effectiveness at reducing hospitalizations, inpatient bed days, and other health service use.

## Methods

### Study Design

This was a 1-year, longitudinal, test-of-change evaluation of the digital self-management technology myCOPD for patients with COPD. The study received Caldicott Guardian approval for anonymized health record data analysis; it received internal ethical approval by NHS Highland Research, Development & Innovation; and all participants provided written informed consent.

### Participants

Participants were recruited over a 6-month period, from May to October 2019, as part of routine community-based care, including scheduled home visits, outpatient appointments, pulmonary rehabilitation, and phone or group appointments. As this was a pragmatic real-world assessment of myCOPD, there were no predetermined age, disease severity, geographical, or socioeconomic inclusion or exclusion criteria. Participants lacking appropriate digital devices, technological skills, or connectivity were not enrolled. A total of 140 people throughout the Scottish Highlands with COPD were offered myCOPD, of whom 120 enrolled during routine health care encounters, including scheduled home visits (n=54, 45.0%), outpatient appointments (n=43, 35.8%), pulmonary rehabilitation (n=13, 10.8%), and phone or group appointments (n=10, 8.3%).

### Intervention

Once enrolled, participants activated the technology by registering and creating an account on the myCOPD platform, which was accessed via an email link sent to each participant. Up to four reminders were sent on a weekly basis to encourage myCOPD activation. Participants used the technology as they wished and did not receive further encouragement during the evaluation. Participants were provided with licenses at no cost to themselves.

### Data Analysis

All participant data were collected for the 12-month period prior to myCOPD enrollment and up to 12 months following technology activation. myCOPD engagement data were collected via the myCOPD clinician portal. Health service use data were obtained via NHS electronic care records, including the NHS Highland Clinical Portal system and out-of-hours contacts using Adastra. Participant rurality and Scottish Index of Multiple Deprivation (SIMD) data were calculated using participant postcodes and relevant lookup tables [12,13].

Health service data were evaluated on a longitudinal basis for all enrolled participants, comparing the incidence of daily hospital admissions, inpatient bed days, and other service use for the period before and after myCOPD activation to March 1, 2020. Enrollment was defined as a participant who consented to participate, received an invitation to enroll, and was allocated a myCOPD license. Activation was defined as a participant who accessed the myCOPD platform and completed account registration. Symptom scoring frequency was defined as follows: low (<1 time every 100 days), moderate (1-5 times every 100 days), high (6-20 times every 100 days), and very high (>20 times every 100 days). Health care usage data from participants who enrolled but did not activate myCOPD contributed to “before” activation results. All data were compliant with the General Data Protection Regulation.

### Statistical Analysis

Power calculations determined that a study size of 100 participants was sufficient to evaluate a primary endpoint of reduced inpatient bed days. Calculations were based on projected modest (10%) reductions in inpatient bed days and significant (25%) seasonal variability in COPD exacerbations, with significance ($\alpha$) at .05 and 80% power (1-$\beta$). Paired participant

health care usage data before and after myCOPD activation were analyzed using the Wilcoxon signed-rank test. Statistical analysis was not performed on user subgroups representing variable myCOPD symptom scoring frequency or module usage, as subgroups were not sufficiently powered.

## Results

Of the 140 people invited to enroll in myCOPD, 20 (14.3%) declined to participate, mostly for technology-related reasons. Of the remaining 120, 7 (5.8%) were excluded, as they died during the study period, leaving a total of 113 participants (Figure 1). The average participant age was 69.3 (SD 8.2) years, and 51.3% (58/113) were female. A total of 70.8% (80/113) of participants were from remote and rural areas, and 75.2% (85/113) represented the three most deprived SIMD quintiles. Most participants (69/113, 61.1%) had moderate or severe COPD, and 20.4% (23/113) had very severe disease according to their Global Initiative for Chronic Obstructive Lung Disease score (Table 1) [14].

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) flowchart of study participants who were offered myCOPD, showing the number of patients who declined and reasons why, the number enrolled, and the number included in final study. COPD: chronic obstructive pulmonary disease.

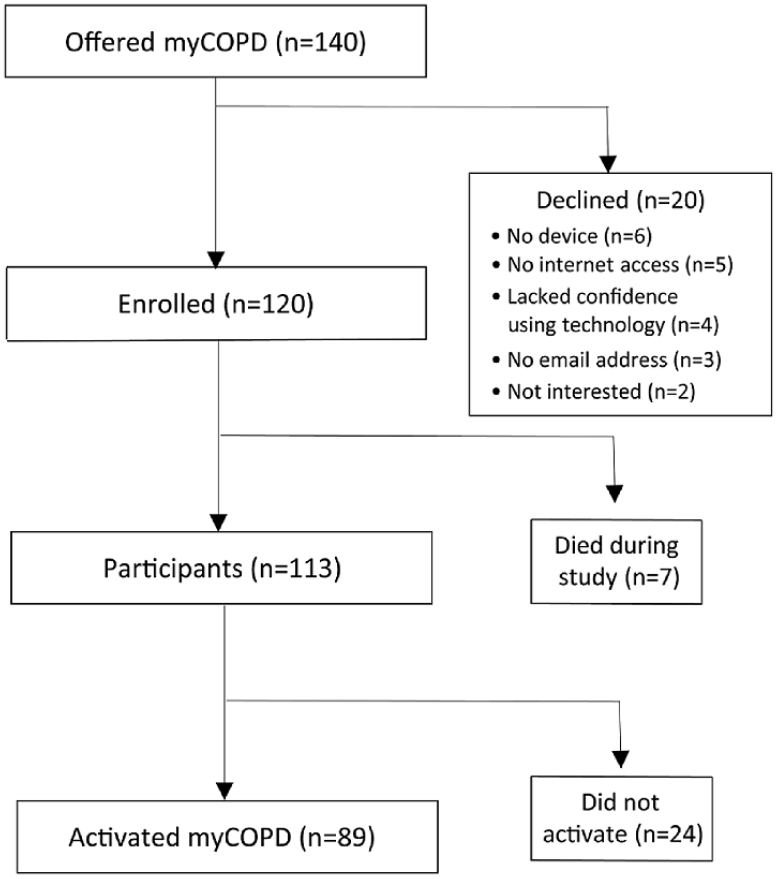


**Table 1.** myCOPD[a] participant characteristics.

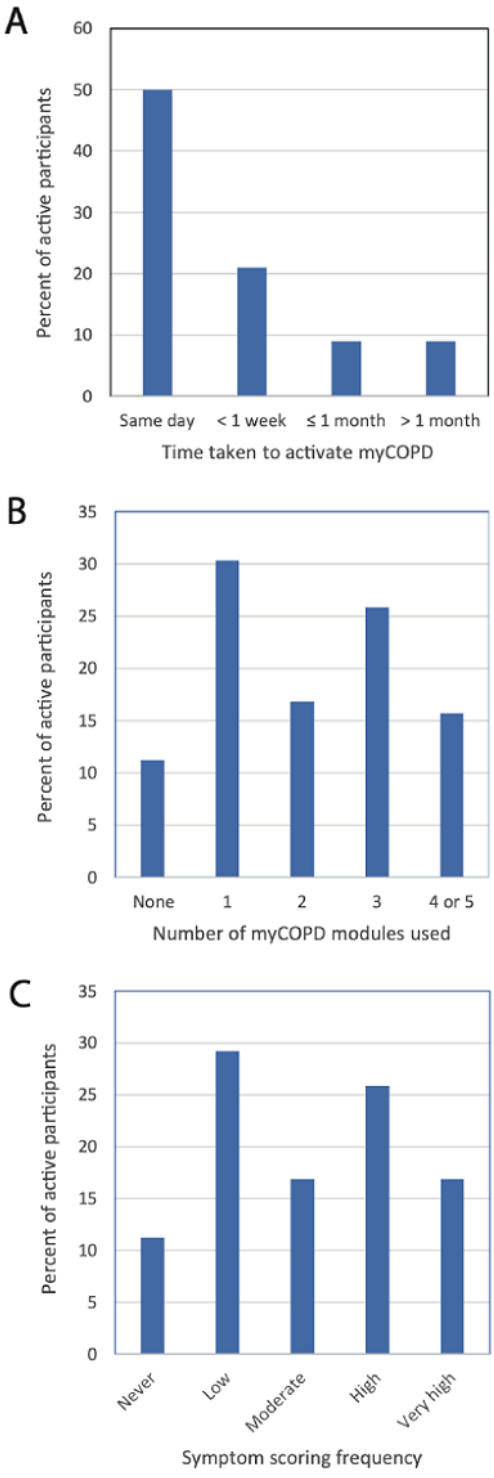
| Participant characteristics | Participants (N=113), n (%) |
|---|---|
| **Age in years** | |
| 31-40 | 0 (0) |
| 41-50 | 2 (1.8) |
| 51-60 | 21 (18.6) |
| 61-70 | 37 (32.7) |
| 71-80 | 35 (31.0) |
| ≥81 | 5 (4.4) |
| Not recorded | 13 (11.5) |
| **Sex** | |
| Female | 58 (51.3) |
| Male | 55 (48.7) |
| **Socioeconomics (SIMD[b] quintile)** | |
| 1 (most deprived) | 11 (9.7) |
| 2 | 27 (23.9) |
| 3 | 47 (41.6) |
| 4 | 25 (22.1) |
| 5 (least deprived) | 3 (2.7) |
| **Urban-rural classification** | |
| Large urban areas | 0 (0) |
| Other urban areas | 31 (27.4) |
| Accessible small towns | 2 (1.8) |
| Remote small towns | 26 (23.0) |
| Accessible rural areas | 7 (6.2) |
| Remote rural areas | 47 (41.6) |
| **COPD severity** | |
| Mild | 11 (9.7) |
| Moderate | 33 (29.2) |
| Severe | 36 (31.9) |
| Very severe | 23 (20.4) |
| Not recorded | 10 (8.8) |

[a]COPD: chronic obstructive pulmonary disease.

[b]SIMD: Scottish Index of Multiple Deprivation.

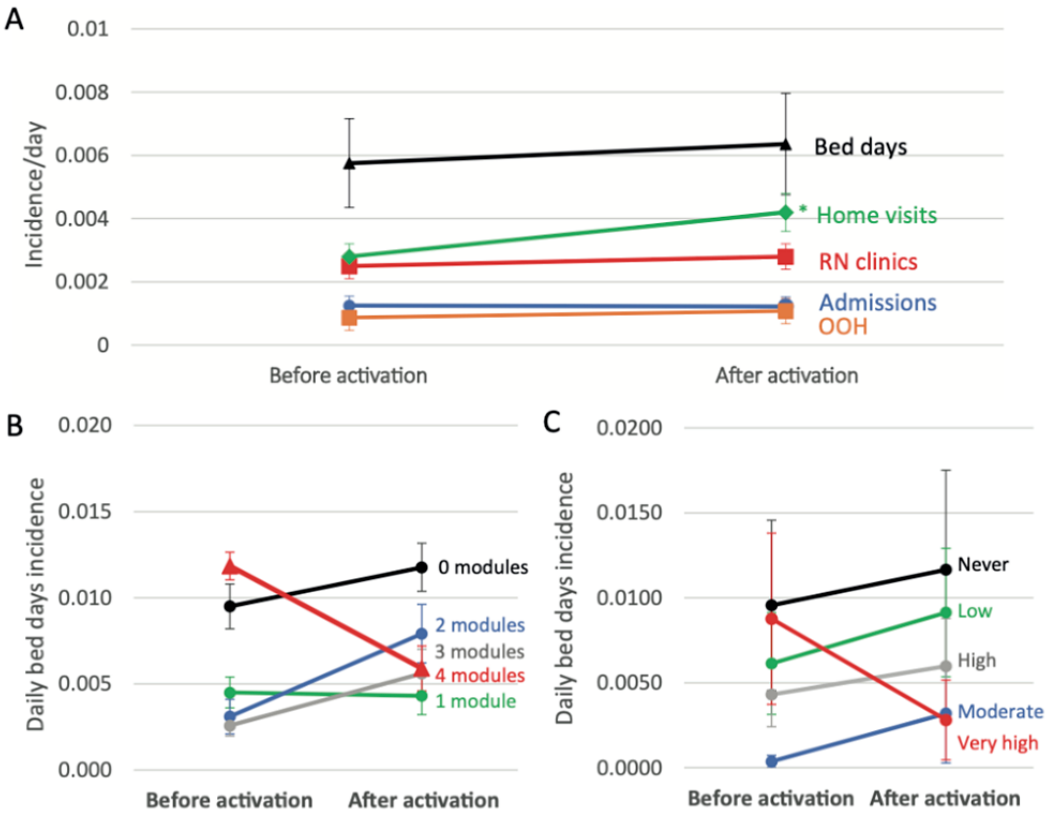
A total of 89 out of 113 (78.8%) participants activated myCOPD, with 56% (50/89) of them doing so on the day of enrollment and 90% (80/89) doing so within 1 month (Figure 2, A). Most active participants used at least one module and entered their symptom scores at least once (Figure 2, B and C; n=79, 89%). A total of 10 (11%) participants activated myCOPD but used no modules. Overall, 57% (n=51) of active participants recorded their CAT score one or more times, 39% (n=35) initiated pulmonary rehabilitation training, 24% (n=21) viewed educational course material, and 10% (n=9) watched at least one inhaler technique video. Out of 89 participants, 15 (17%) were very high users based on symptom scoring frequency (Figure 2, C), and 14 of these (93%) also used four or five myCOPD modules (Figure 2, B), suggesting a discrete subgroup of highly engaged users. There was no overall correlation between myCOPD engagement and participant age, SIMD status, or rurality (Multimedia Appendices 1 and 2).

**Figure 2.** myCOPD engagement. Bar charts showing (A) time taken to activate myCOPD account following enrollment, (B) participant module usage, and (C) frequency of entering symptom scores. COPD: chronic obstructive pulmonary disease.




To evaluate myCOPD effectiveness, we quantified the daily incidence of inpatient bed days, hospital admissions, home visits, clinic appointments, and out-of-hours care provision for an average of 375 (SD 32) days before and 239 (SD 46) days after myCOPD activation, for a total of 69,211 participant days (47,972 days before and 21,239 days after). There were no significant differences in hospital admissions, inpatient bed days, or other health service use before or after myCOPD activation, apart from a modest increase in home visits consistent with previous telemonitoring studies [15] (Figure 3, A). Even after excluding participants who did not activate their license, there remained no significant difference for any of the categories.

**Figure 3.** Daily incidence of health service usage before and after myCOPD activation. (A) Average daily incidence of health service use among all participants before and after myCOPD activation. Subgroup analysis of average daily inpatient bed day use before and after myCOPD activation according to (B) module usage and (C) frequency of symptom scoring. COPD: chronic obstructive pulmonary disease; OOH: out-of-hours; RN: registered nurse.




Subgroup analysis results can be seen in Multimedia Appendix 3, A-H. Although underpowered, subgroup analysis based on either module usage (Figure 3, B; Multimedia Appendix 3, E) or symptom scoring frequency (Figure 3, C; Multimedia Appendix 3, F) did identify trends toward reduced inpatient bed days and hospital admissions for highly engaged users. There were also increased home visits in all subgroups after myCOPD activation regardless of module usage or symptom scoring frequency (Multimedia Appendix 3, A and B). No other trends in health service use were observed based on subgroup analysis regarding clinic appointments (Multimedia Appendix 3, C and D) and out-of-hours contacts (Multimedia Appendix 3, G and H).

## Discussion

This study is the longest and largest evaluation of the digital health self-management technology myCOPD to date, the only one involving a predominantly remote and rural population, and the first to recruit patients from within community care settings using a pragmatic approach. Enrollment and engagement with myCOPD was popular, with 78.8% (89/113) of participants activating the technology and 89% (79/89) of these participants using at least one module or entering their symptom scores at least once. Only 14.3% (20/140) of people approached declined to participate in the study, and there was no correlation between participant enrollment, activation, or engagement and either age, socioeconomics, rurality, or disease severity, suggesting that these are not significant barriers to using myCOPD. This finding may help mitigate perceived risks of increased health inequalities associated with the use of digital health technologies as part of routine care provision.

Despite high levels of activation, myCOPD did not reduce overall demands on health services. These findings are consistent with the limited evidence supporting the use of COPD self-management technologies, but they contrast with previous studies involving myCOPD [8-11]. There are several possible explanations for this difference. First, our study involved community-based recruitment of stable patients with COPD irrespective of exacerbation frequency, whereas other trials recruited hospital-based patients immediately following an acute exacerbation where motivation to engage in self-management may be greater. Second, previous studies collected data for only 90 days and, therefore, evaluated acute rather than long-term myCOPD benefits [9]. Finally, it remains possible that cultural or socioeconomic differences between rural and urban participants might influence myCOPD engagement and impact. Our results highlight the need for further evaluation of myCOPD and other digital health technologies ahead of their widespread procurement and adoption as part of routine health services. It may be that myCOPD can function as an effective tool in reducing COPD exacerbations when offered to participants in hospital and at a time of crisis, whereas it may not function in

this manner when offered to patients in the community who are not actively in crisis or experiencing an exacerbation.

Despite no overall reduction in health service use, we did observe trends toward reduced hospital admissions and inpatient bed days in a subgroup of highly engaged users. This suggests the technology may be clinically beneficial if it is highly used and suggests that a greater emphasis is needed for understanding the motivation to use digital self-management tools and how to promote increased, meaningful user engagement. Paradoxically, previous studies indicate that those patients who may benefit most from engaging with digital self-management technologies are the least likely to do so [16]. Our observation that highly engaged myCOPD users were indistinguishable in terms of age, socioeconomics, rurality, or disease severity suggests that the factors driving meaningful user engagement are complex and require further attention. This will necessitate increased collaboration among a wide group of stakeholders, including patients, throughout all stages of digital health technology design, development, and testing.

One potential limitation of this study involves differences in the amount of data we collected before versus after myCOPD activation (47,972 participant days before and 21,239 after). Our original design involved collecting an equivalent quantity of data before and after myCOPD activation, but a decision was made to terminate the study on March 1, 2020, due to the emergent COVID-19 pandemic. We mitigated the impact of this change by evaluating data according to daily rather than annual individual health service usage. Interestingly, and despite cessation of formal data collection, we observed increased myCOPD engagement among many participants after March 2020, which may reflect changes in health behavior when access to face-to-face services was limited. A further impact of the COVID-19 pandemic was our limited ability to evaluate the effect of seasonality on exacerbation frequency, and it is possible that the inclusion of data beyond March 2020 may have resulted in different outcomes.

In conclusion, although our study does not support implementation of myCOPD to reduce health service demands on NHS Highland on a population-wide basis, our results do indicate that some highly engaged patients may derive benefits. Thus, individuals can be encouraged to individually adopt myCOPD as part of their self-management care should they find it useful. Further research is needed to understand what motivates some individuals to engage with digital health technologies, in order to facilitate the design and development of clinically and economically effective self-management tools.

## Acknowledgments

Jane Stokes and Liam Clutterbuck (my mHealth) provided training, advice, and support on the use of the myCOPD (chronic obstructive pulmonary disease) platform at an agreed cost. Technology license costs were £40 (US $51.30) per license, which was covered by external grant funding. This project was supported, in part, through Inverness and Highland City-Region Deal funding and the National Technology Enabled Care Test of Change funding scheme.

## Authors' Contributions

RC was involved in data analysis and wrote the manuscript. AG was involved in study design and data analysis and wrote the manuscript with RC and EKS. MD, EF, SH, MS, and JC were involved in study design, patient recruitment and enrollment, data collection, and manuscript preparation. JG and MM were involved in study design, data collection, patient enrollment, and manuscript preparation. EKS was involved in study design, data analysis, and manuscript preparation.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of myCOPD engagement and patient demographics. Average daily symptom scoring frequency according to participant age (top), SIMD quintile (middle), and rurality score (bottom). Trendline and $R^2$ values are representative of all participant data. COPD: chronic obstructive pulmonary disease; SIMD: Scottish Index of Multiple Deprivation.
[PNG File , 85 KB - mhealth_v10i2e30782_app1.png ]

Multimedia Appendix 2
Comparison of myCOPD engagement and patient demographics. myCOPD module usage according to participant age (top), SIMD quintile (middle), and rurality score (bottom). Trendline and $R^2$ values are representative of all participant data. COPD: chronic obstructive pulmonary disease; SIMD: Scottish Index of Multiple Deprivation.
[PNG File , 64 KB - mhealth_v10i2e30782_app2.png ]

Multimedia Appendix 3
Subgroup analysis of health service usage relative to myCOPD engagement. Average daily home visits (A, B), registered nurse (RN) clinics (C, D), hospital admissions (E, F), and out-of-hours (OOH) care (G, H) according to myCOPD module usage (A, C, E, G) or symptom scoring frequency (B, D, F, H) for all participant subgroups. COPD: chronic obstructive pulmonary disease.

[PNG File , 154 KB - mhealth_v10i2e30782_app3.png ]

## References


1. Lung disease in the UK - Big picture statistics. British Lung Foundation. 2021. URL: https://statistics.blf.org.uk/lung-disease-uk-big-picture [accessed 2021-05-21]
2. Chronic obstructive pulmonary disease (COPD) statistics. British Lung Foundation. URL: https://statistics.blf.org.uk/copd [accessed 2021-05-01]
3. Trueman D, Woodcock F, Hancock E. Estimating the Economic Burden of Respiratory Illness in the UK. London, UK: British Lung Foundation; 2017. URL: https://www.blf.org.uk/policy/economic-burden [accessed 2021-05-17]
4. Raju S, Keet CA, Paulin LM, Matsui EC, Peng RD, Hansel NN, et al. Rural residence and poverty are independent risk factors for chronic obstructive pulmonary disease in the United States. Am J Respir Crit Care Med 2019 Apr 15;199(8):961-969 [FREE Full text] [doi: 10.1164/rccm.201807-1374OC] [Medline: 30384774]
5. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 04;8:CD011682. [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
6. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: An overview of reviews. Chron Respir Dis 2017 Aug;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
7. myCOPD: The COPD app for controlling your symptoms. my mhealth. URL: https://mymhealth.com/mycopd [accessed 2021-05-21]
8. Bourne S, DeVos R, North M, Chauhan A, Green B, Brown T, et al. Online versus face-to-face pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: Randomised controlled trial. BMJ Open 2017 Jul 17;7(7):e014580 [FREE Full text] [doi: 10.1136/bmjopen-2016-014580] [Medline: 28716786]
9. North M, Bourne S, Green B, Chauhan AJ, Brown T, Winter J, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: The RESCUE trial. NPJ Digit Med 2020;3:145 [FREE Full text] [doi: 10.1038/s41746-020-00347-7] [Medline: 33145441]
10. Crooks MG, Elkes J, Storrar W, Roy K, North M, Blythin A, et al. Evidence generation for the clinical impact of myCOPD in patients with mild, moderate and newly diagnosed COPD: A randomised controlled trial. ERJ Open Res 2020 Oct;6(4):1-10 [FREE Full text] [doi: 10.1183/23120541.00460-2020] [Medline: 33263052]
11. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. NPJ Prim Care Respir Med 2020 Apr 01;30(1):11 [FREE Full text] [doi: 10.1038/s41533-020-0167-1] [Medline: 32238810]
12. Scottish Government Urban Rural Classification 2016. Edinburgh, Scotland: The Scottish Government; 2018 Mar. URL: https://www2.gov.scot/Resource/0054/00542959.pdf [accessed 2021-05-22]
13. Scottish Index of Multiple Deprivation. Edinburgh, Scotland: The Scottish Government; 2020. URL: https://simd.scot/#/simd2020/BFTTTTT/7.094052301976135/-5.5000/58.4374/ [accessed 2021-05-22]
14. Pocket Guide to COPD Diagnosis, Management, and Prevention: A Guide for Health Care Professionals. 2017 Report.: Global Initiative for Chronic Obstructive Lung Disease; 2017. URL: https://goldcopd.org/wp-content/uploads/2016/12/wms-GOLD-2017-Pocket-Guide.pdf [accessed 2021-05-17]
15. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: Researcher blind, multicentre, randomised controlled trial. BMJ 2013;347:f6070 [FREE Full text] [Medline: 24136634]
16. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: Analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]


## Abbreviations

**CAT:** COPD Assessment Test
**COPD:** chronic obstructive pulmonary disease
**NHS:** National Health Service
**SIMD:** Scottish Index of Multiple Deprivation

*Edited by L Buis; submitted 03.06.21; peer-reviewed by C Carlin, D Ryan; comments to author 14.07.21; revised version received 02.09.21; accepted 19.11.21; published 07.02.22.*

*Please cite as:*
*Cooper R, Giangreco A, Duffy M, Finlayson E, Hamilton S, Swanson M, Colligan J, Gilliatt J, McIvor M, Sage EK*
*Evaluation of myCOPD Digital Self-management Technology in a Remote and Rural Population: Real-world Feasibility Study*
*JMIR Mhealth Uhealth 2022;10(2):e30782*
*URL: https://mhealth.jmir.org/2022/2/e30782*
*doi:10.2196/30782*
*PMID:35129453*

Original Paper

# The Effect of a Mobile and Wearable Device Intervention on Increased Physical Activity to Prevent Metabolic Syndrome: Observational Study

Hee Jin Kim[1], MSc; Kang Hyun Lee[1], MD, PhD; Jung Hun Lee[1], MSc; Hyun Youk[1], MD; Hee Young Lee[1], PhD

Department of Emergency Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

**Corresponding Author:**
Kang Hyun Lee, MD, PhD
Department of Emergency Medicine
Yonsei University Wonju College of Medicine
20 Ilsan-ro
Wonju, 26426
Republic of Korea
Phone: 82 33 741 1612
Email: ed119@yonsei.ac.kr

## Abstract

**Background:** Research on whether wearable devices and app-based interventions can effectively prevent metabolic syndrome (MetS) by increasing physical activity (PA) among middle-aged people living in the rural areas of South Korea remains insufficient.

**Objective:** The aim of this study was to determine whether mobile and wearable device interventions can improve health indicators, including PA, in MetS risk groups in rural South Korea.

**Methods:** In this clinical trial, performed from December 2019 to June 2020, participants were asked to use a wearable device (GalaxyWatch Active1) alone (standard intervention) or the wearable device and mobile app (Yonsei Health Korea) (enhanced intervention). Clinical measures and International Physical Activity Questionnaire (IPAQ) scores were evaluated initially and after 6 months. The number of steps was monitored through the website. The primary outcome was the difference in PA and clinical measures between the enhanced intervention and standard intervention groups. The secondary outcome was the decrease in MetS factors related to the change in PA.

**Results:** A total of 267 participants were randomly selected, 221 of whom completed the 6-month study. Among the 221 participants, 113 were allocated to the enhanced intervention group and 108 were allocated to the standard intervention group. After 6 months, the body weight and BMI for the enhanced intervention group decreased by 0.6 (SD 1.87) and 0.21 (SD 0.76), respectively ($P$<.001). In both groups, systolic blood pressure, diastolic blood pressure, waist circumference, and glycated hemoglobin $A_{1c}$ ($HbA_{1c}$) decreased ($P$<.001). The total PA was approximately 2.8 times lower in the standard intervention group (mean 44.47, SD 224.85) than in the enhanced intervention group (mean 124.36, SD 570.0). Moreover, the enhanced intervention group achieved the recommended level of moderate to vigorous physical activity (MVPA), whereas the standard intervention group did not (188 minutes/week vs 118 minutes/week). Additionally, the number of participants in the enhanced intervention group (n=113) that reached 10,000 daily steps or more after the intervention increased from 9 (8.0%) to 26 (23.1%) ($P$=.002), whereas this number did not increase significantly in the standard intervention group (n=108), from 8 (7.4%) to 16 (14.8%) ($P$=.72). The number of participants without any MetS factors increased by 12 (11%) and 8 (7%) in the enhanced and standard intervention group, respectively.

**Conclusions:** PA monitoring and an intervention using wearable devices were effective in preventing MetS in a rural population in Korea. Blood pressure, waist circumference, and $HbA_{1c}$ were improved in both intervention groups, which were effective in reducing MetS factors. However, only the participants in the enhanced intervention group continuously increased their MVPA and step counts above the recommended level to prevent MetS. Body weight and BMI were further improved, and a higher number of participants with zero MetS factors was attained from the enhanced intervention.

**Trial Registration:** Clinical Research Information Service KCT0005783; https://cris.nih.go.kr/cris/search/detailSearch.do/16123

***(JMIR Mhealth Uhealth 2022;10(2):e34059)*** doi:10.2196/34059

**KEYWORDS**

mHealth; physical activity; wearable device; metabolic syndrome; health care; exercise; intervention; Asia; Korea; rural

# Introduction

## Background

Metabolic syndrome (MetS) is a disease characterized by three or more of the following five factors: abdominal obesity, high blood pressure, high blood sugar, high triglycerides, and high high-density lipoprotein cholesterol [1]. Approximately one-quarter of the world's adult population has MetS [2], which is a major cause of disability as well as a leading cause of death in 60% of the global population [3]. MetS is also a serious risk factor for heart disease, stroke, and type 2 diabetes [4]. Research shows that rural residents are less accessible and active than urban residents, who reported exercising in constructed environments such as neighborhood streets, parks, and shopping malls. There is also a difference in income level between urban and rural populations [5]. Moreover, studies have shown that rural residents lack education regarding proper eating habits compared to urban residents and are more likely to be obese; in addition, their socioeconomic status and access to medical services are lower than those of urban residents [6,7]. Rural residents appear to have a higher risk for MetS and a higher disease burden than urban residents due to these differences in infrastructure [8].

Global smartwatch sales continue to increase, which are expected to reach 109.2 million units in 2023 [9], and the wearable device market is expected to continue to expand [10]. Mobile health (mHealth) technologies using apps and wearable devices are becoming increasingly popular, as they allow patients to monitor their own health conditions [11,12]. Moreover, mobile apps suggest a variety of methods to prevent disease and maintain and improve patient health [13]. Wearable devices can improve the lifestyle of patients with chronic diseases [14]. Previous reviews on promoting physical activity (PA) in adults have shown evidence for the effectiveness of mHealth on increasing PA [15,16]. Most health-related behaviors such as eating well and exercising regularly can lead to significant improvements if sustained through motivation [4,17]. However, it is difficult to promote the self-management of chronic diseases among the elderly, which is also a valid concern among the elderly population in Korea [14]. According to a study by the Korea Institute for Health and Social Affairs, which measured the PA of Korean adults (N=697), the proportion of men aged 65 years or above who were engaged in vigorous PA decreased by approximately half from 9.3% in 2010 to 4.9% in 2018, and the proportion of those engaged in moderate PA decreased from 14.6% to 10%. During the same period, the proportion of women of the same age engaging in vigorous PA decreased from 3.3% to 2% and the proportion engaging in moderate PA decreased from 6.6% to 4% [18]. These results suggest that lack of PA might be a health risk factor for adults in Korea [19].

Previous studies have shown that PA offers a variety of health benefits, including reducing anxiety and depression; improving sleep and quality of life; and lowering the risk of developing diabetes, heart disease, and many cancers [20]. It has been reported that lack of PA increases health risks, including coronary heart disease, type 2 diabetes, and cancer, and shortens life expectancy from major noncommunicable diseases [19,20]. Regular PA can help to prevent aging-related declines in physical function, and reduces morbidity and mortality [21]. Since self-management of chronic diseases requires treatment or behavioral modification, support tools are needed to maintain practice in daily life [22]. Interventions through wearables and/or smartphone apps are effective in promoting PA in the adult population [23].

## Objectives

We hypothesized that intervention components, especially wearable devices and mobile apps, for the prevention of MetS will have a positive effect on PA in the middle-aged population in Korea. The purpose of the study was two-fold: (1) to compare the changes in clinical values and PA between an enhanced intervention group and standard intervention group that had not received the intervention for 6 months, and (2) to objectively reduce risk factors of MetS. We further explored whether the change in MetS risk factors is related to the measured PA change.

# Methods

## Ethical Considerations

The study was performed at Yonsei University Wonju Severance Christian Hospital (Wonju, Korea) between December 2019 and June 2020, and was approved after deliberation by the Research Ethics Review Committee (approval number CR319089; trial registration number KCT0005783). Participants of the clinical trial were informed of the purpose and procedure of the study through the consent form for participation in the study and were asked to fill out the consent form in writing. Instructions were provided in the questionnaire, including statements that no personal information will be exposed for purposes other than research, and that participation is voluntary and can be withdrawn at any time.

## Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) adults aged 40 to 80 years, (2) people with one or more MetS factors (Textbox 1), (3) participating in the Wonju-Pyeongchang cohort study via the Korea Centers for Disease Control and Prevention, (4) agree to participating in the clinical trial, and (5) able to participate after understanding the training and instructions. A diagnosis of MetS was based on the diagnostic criteria for MetS in Korea, modified from the National Cholesterol Education Program Adult Treatment Panel-III criteria. A total of 221 people (113 in the enhanced intervention group and 108 in the standard intervention group) were enrolled according to the standards for randomized controlled trials (Figure 1).

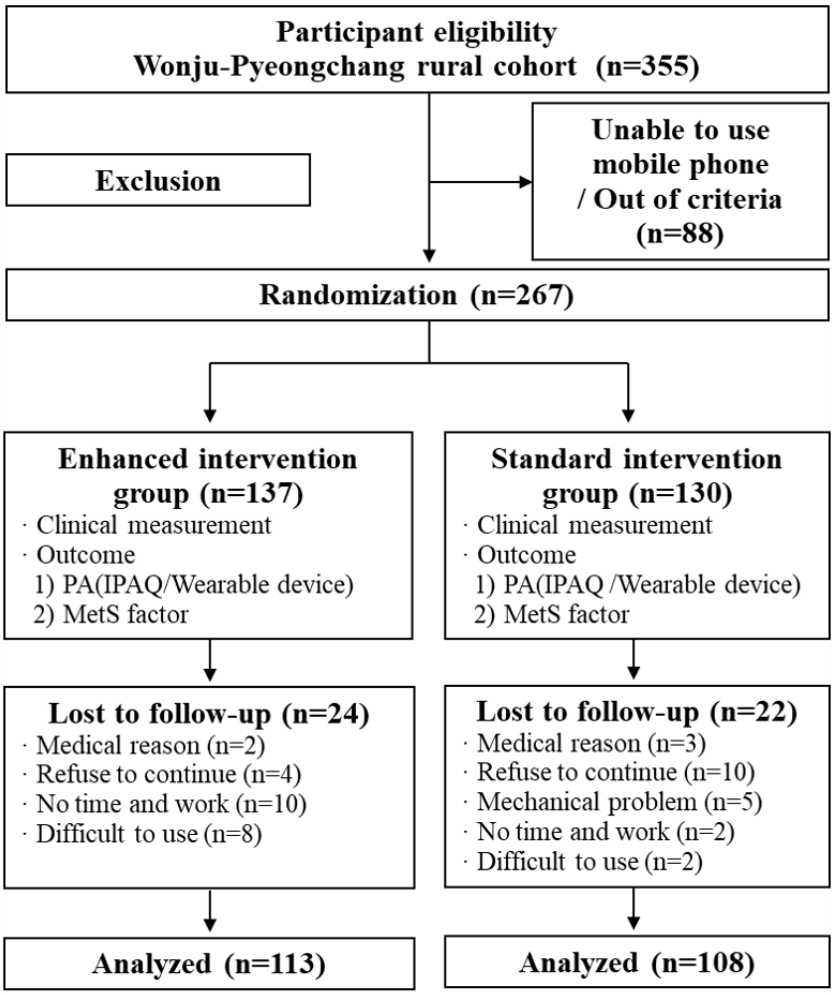
Besides having a MetS diagnosis or risk factor, participants also had to have a smartphone using the Android operating system. Participants had to be able to receive and read text messages,

without have any difficulty in using wearable devices and mobile apps that will send alerts. Some patients were excluded 1 month before participation due to the use of warfarin (eg, Coumadin); additionally, those with physical disabilities who could not use a wearable device, those with skin diseases and dysfunction, and those who had an aversion to the wearable device were excluded.

**Textbox 1.** Metabolic syndrome factors considered for study inclusion.

- Waist circumference ≥90 cm in men and ≥80 cm in women
- Systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥85 mmHg
- Triglyceride level ≥150 mg/dL
- High-density lipoprotein cholesterol level <40 mg/dL in men and <50 mg/dL in women
- Fasting plasma glucose level ≥100 mg/dL

**Figure 1.** Flowchart of study participants. PA: physical activity; IPAQ: International Physical Activity Questionnaire; MetS: metabolic syndrome.




## Measures

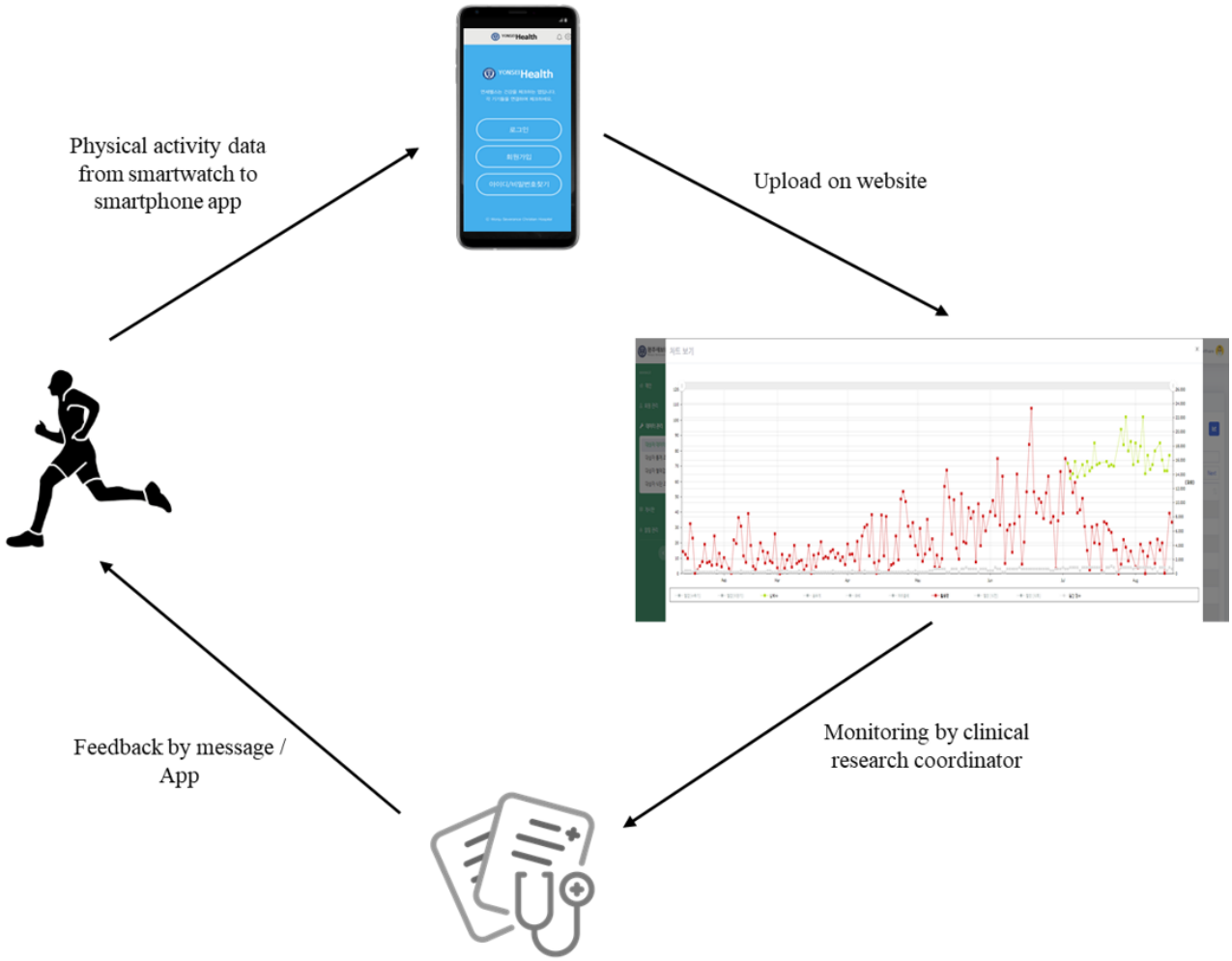
Participants were selected among the existing Wonju-Pyeongchang rural cohort list, and only those who were connected and willing to participate in the clinical study were recommended to visit the Smart Healthcare Support Center of Wonju Severance Christian Hospital. On the first visit, an approximate 1-hour session was performed including clinical measurements, MetS education, and device education. Clinical measurements comprised an 8-hour fasting blood test. Weight and height were measured using an automated digital scale (Tanita T6360) with shoes and jackets removed. Participants were measured in a fasting state without outerwear. BMI was calculated as the ratio of weight to height squared. Waist circumference was measured using a digital tape measure (PIE, Bagle Labs Co, Korea). An automated blood pressure meter (HEM-9000T, Omron Co, Japan) was used to measure diastolic/systolic blood pressure. All participants underwent clinical examination before and after the start of the study.

## Intervention

The structure and contents of the mobile app and text messages were developed by the researchers and clinical research coordinators. The National Health Insurance Service, the

guidelines for endocrinology, educational materials, and literature for existing MetS programs were referenced. The overall flow chart of the study is shown in Figure 2. A wearable device (Galaxy Watch Active1) was provided to both the enhanced and standard intervention groups. The participants in the enhanced intervention group were further provided with a mobile app (Yonsei Health) that allows users to check text messages, phone contacts, videos, and reference educational materials customized for the prevention and treatment of MetS. Participants were recommended to use the provided mobile app and wearable device at least 3 times a week (more than 8 hours/day). The data from the wearable device were transmitted to the web server through the website where the measurements were checked by a research nurse. On the website, changes in PA could be observed as the accumulated number of patients' steps, and an SMS text message could be sent to the patient directly. The standard intervention group received no intervention other than self-monitoring, whereas the enhanced intervention group was provided with feedback on their PA. The feedback comprised reward messages for increased activity or sufficient PA, and encouraging messages were provided every other week if the level of activity decreased or remained unchanged compared with that of the previous week. The feedback was centered on encouraging and maintaining PA. Health information provided through the mobile app contained guidance on PA and lifestyle to prevent or reduce risk factors for MetS, mainly comprising medical, lifestyle, nutrition, exercise, and cognitive categories.

**Figure 2.** Flow of intervention.




## Evaluation

### *Data Collection*

The evaluation was performed at baseline and 6 months later. Demographic data such as age, gender, education level, and average monthly income were collected face to face using questionnaires. Clinical measures were collected using the same parameters to ensure comparability before and after the study.

### *PA Assessment*

The short format of the International Physical Activity Questionnaire (IPAQ) is a form for population surveillance and the long format is designed to be used in research that requires information on various PA areas [24]. The long-form IPAQ includes a total of 31 items, providing a more detailed assessment of the level of PA than the short format [25]. The questions include home and garden work activities, occupational activities, transportation use, and leisure time in the last 7 days, which are designed to record the time of walking, vigorous and moderate PA, and time spent sitting or lying down. The

questionnaire was prepared by interviewing the research coordinator and constructing a self-written form. Metabolic equivalents of task (MET) was used to calculate official scores, which is an intensity unit of PA determined as the ratio between the metabolic rate during a given activity over the resting metabolic rate [26]. PA was represented using the number of days and duration (time) of activity, and was calculated based on the following formulas [27]:

Walking (MET hours/week) = 3.3 × walking hours × days **(1)**

Moderate PA (MET hours/week) = 4.0 × moderate PA hours × days **(2)**

Vigorous PA (MET hours/week) = 8.0 × vigorous PA hours×days **(3)**

Total MET minutes/week = sum of Walking + Moderate + Vigorous MET minutes/week scores **(4)**

### *Step Counts*

The number of steps was measured using a wearable device. After each measurement, data were automatically uploaded to the website through a Bluetooth-connected smartphone app. The wearable device used in this study was a smartwatch that enabled Bluetooth connection with the Android operating system. This device was adopted because the user interface and method of use were similar to those of smartphones; therefore, there was no difficulty for participants to use it.

### Statistical Analysis

The general characteristics of the subjects are summarized using descriptive statistics. A paired *t* test was used to compare clinical measures before and after the intervention, and a *t* test and a $\chi^2$ test were performed to compare differences between groups. SPSS 26.0 was used for statistical analysis; $P<.05$ was considered significant.

## Results

### Participant Characteristics

Participants were recruited after obtaining Institutional Review Board approval in December 2019, and registration began thereafter. Data collection lasted for up to 6 months from the date of registration. Finally, of the 221 subjects included in the study, 113 were allocated to the enhanced intervention group and 108 were allocated to the standard intervention group. In both groups, more than half of the participants had secondary or higher education (80.9% and 81.4%, respectively), and the mean age was 64.8 (SD 6.3) and 66.3 (SD 6.2) years in the enhanced and standard intervention group, respectively (Table 1). Women accounted for more than 50% of the cohort. The two groups did not differ with respect to educational attainment, marital status, and monthly income. Clinical measures such as BMI, glycated hemoglobin ($HbA_{1c}$), triglycerides, and blood pressure were not different between groups; however, high-density lipoprotein cholesterol levels differed between the two groups ($P=.03$). Additionally, the proportion of patients with MetS at baseline was higher in the standard intervention group ($P=.03$).

**Table 1.** Characteristics of the study participants at baseline.

| Variables | Enhanced intervention group (n=113), n (%) | Standard intervention group (n=108), n (%) | *P* value |
|---|---|---|---|
| **Age group (years)** | | | .07 |
| 35-50 | 0 (0) | 1 (0.9) | |
| 51-70 | 86 (76.1) | 78 (72.9) | |
| >70 | 27 (23.9) | 29 (26.2) | |
| **Gender** | | | .21 |
| Male | 45 (39.8) | 52 (48.1) | |
| Female | 68 (60.2) | 56 (51.9) | |
| **Level of education** | | | .34 |
| No response | 0 (0) | 1 (0.9) | |
| No formal schooling | 2 (1.8) | 1 (0.9) | |
| Elementary school | 19 (16.8) | 18 (16.7) | |
| Middle school | 23 (20.4) | 20 (18.5) | |
| High school | 44 (38.9) | 36 (33.3) | |
| University | 20 (17.7) | 24 (22.2) | |
| Postgraduate | 5 (4.4) | 8 (7.4) | |
| **Marital status** | | | .59 |
| Never married | 1 (0.9) | 0 (0) | |
| Married | 106 (93.8) | 102 (94.4) | |
| Divorced | 1 (0.9) | 0 (0) | |
| Widowed | 5 (4.4) | 6 (5.6) | |
| **Average monthly income (Won[a])** | | | .19 |
| <1 million | 15 (13.3) | 12 (11.1) | |
| 1-2.99 million | 54 (47.8) | 55 (50.9) | |
| 3-4.99 million | 34 (30.1) | 23 (21.3) | |
| ≥5 million | 10 (8.8) | 18 (16.7) | |
| **BMI (kg/m$^2$)** | | | .61 |
| <23 | 22 (19.5) | 16 (14.8) | |
| 23-24.9 | 37 (32.7) | 34 (31.5) | |
| ≥25 | 54 (47.8) | 58 (53.7) | |
| **$HbA_{1c}$[b] (%)** | | | .11 |
| ≤5.6 | 50 (44.3) | 34 (31.5) | |
| 5.7-6.4 | 53 (46.9) | 67 (62.0) | |
| ≥6.5 | 10 (8.8) | 7 (6.5) | |
| **TG[c] (mg/dL)** | | | .15 |
| <200 | 94 (83.2) | 81 (75.0) | |
| 200-239 | 8 (7.1) | 14 (13.0) | |
| ≥240 | 11 (9.7) | 13 (12.0) | |
| **HDL-C[d] (mg/dL)** | | | .03 |
| <60 | 90 (79.6) | 94 (87.0) | |
| ≥60 | 23 (20.4) | 14 (13.0) | |
| **SBP[e] (mmHg)** | | | .32 |

| Variables | Enhanced intervention group (n=113), n (%) | Standard intervention group (n=108), n (%) | *P* value |
|---|---|---|---|
| <120 | 9 (8.0) | 6 (5.6) | |
| 120-139 | 51 (45.1) | 38 (35.2) | |
| 140-159 | 39 (34.5) | 50 (46.3) | |
| ≥160 | 14 (12.4) | 14 (13.0) | |
| **DBP[f] (mmHg)** | | | .60 |
| <80 | 12 (10.6) | 13 (12.0) | |
| 80-89 | 44 (38.9) | 41 (38.0) | |
| 90-99 | 45 (39.8) | 40 (37.0) | |
| ≥100 | 12 (10.6) | 14 (13.0) | |
| **MetS[g] factors** | | | .03 |
| 1 | 16 (14.2) | 10 (9.3) | |
| 2 | 35 (31.0) | 25 (23.1) | |
| 3 | 32 (28.3) | 33 (30.6) | |
| 4 | 26 (23.0) | 33 (30.6) | |
| 5 | 4 (3.5) | 7 (6.5) | |

[a]10 Won=US $0.01.

[b]$HbA_{1c}$: glycated hemoglobin.

[c]TG: triglyceride.

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

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]MetS: metabolic syndrome.

## Change in Participant Measures From Baseline to Follow-Up

Table 2 shows the PA and clinical values measured before and after the intervention. In the enhanced intervention group, body weight, BMI, systolic and diastolic blood pressure, waist circumference, and $HbA_{1c}$ decreased (*P*<.001), whereas in the standard intervention group, only systolic and diastolic blood pressure, waist circumference, and $HbA_{1c}$ decreased (*P*<.001). In both groups, vigorous PA increased and sitting time decreased (*P*<.001), and total PA (*P*=.02 and *P*=.04 in the enhanced and standard intervention group, respectively) and time of walking (*P*<.001) also increased. However, in the enhanced intervention group, vigorous PA, total PA, and walking time increased more than those in the standard intervention group, and the sitting time decreased. The increase of moderate PA was also larger in the enhanced intervention group (Table 2).

**Table 2.** Change in clinical and physical activity outcomes in the study groups following the intervention.

| Variables | Enhanced intervention group (n=113) | | | | Standard intervention group (n=108) | | | |
|---|---|---|---|---|---|---|---|---|
| | Baseline | Follow-up | Change | *P* value | Baseline | Follow-up | Change | *P* value |
| Weight (kg), mean (SD) | 65.5 (10.7) | 64.9 (10.5) | –0.6 (1.8) | <.001 | 65.7 (9.0) | 65.8 (9.0) | 0.1 (3.09) | .55 |
| BMI (kg/m$^2$), mean (SD) | 25.5 (3.0) | 25.3 (2.9) | –0.2 (0.7) | <.001 | 25.7 (2.9) | 25.7 (2.9) | 0.0 (0.9) | .81 |
| SBP[a] (mm Hg), mean (SD) | 139.5 (15.8) | 126.5 (13.9) | –13.0 (14.6) | <.001 | 141.8 (18.1) | 126.1 (21.4) | –15.6 (24.4) | <.001 |
| DBP[b] (mm Hg), mean (SD) | 89.7 (8.6) | 80.0 (9.0) | –9.7 (8.4) | <.001 | 89.1 (8.8) | 79.8 (9.1) | –9.2 (8.1) | <.001 |
| WC[c] (cm), mean (SD) | 91.2 (7.6) | 87.0 (8.0) | –4.1 (4.4) | <.001 | 92.4 (6.7) | 89.5 (7.7) | –2.9 (5.0) | <.001 |
| HDL-C[d] (mg/dL), mean (SD) | 50.5 (10.1) | 49.6 (10.6) | –0.8 (7.9) | .25 | 47.5 (10.7) | 48.5 (10.4) | 0.9 (7.0) | .17 |
| TG[e] (mg/dL), mean (SD) | 146.6 (88.1) | 155.2 (99.5) | 8.5 (63.3) | .15 | 170.4 (150.8) | 150.1 (87.0) | –20.3 (134.5) | 0.11 |
| $HbA_{1c}$[f] (%), mean (SD) | 5.7 (0.4) | 5.6 (0.4) | –0.1 (0.2) | <.001 | 5.8 (0.6) | 5.7 (0.5) | –0.1 (0.2) | <.001 |
| **PA[g], mean (SD)** | | | | | | | | |
| Vigorous PA (hours/week) | 0.8 (2.3) | 2.3 (4.8) | 1.5 (5.3) | <.001 | 0.8 (1.6) | 1.5 (2.2) | 0.7 (2.5) | <.001 |
| Moderate PA (hours/week) | 0.5 (1.3) | 0.7 (2.3) | 0.2 (2.6) | .47 | 0.5 (1.1) | 0.4 (0.8) | –0.1 (1.4) | .27 |
| Walking (hours/week) | 0.6 (1.1) | 1.0 (2.0) | –0.4 (2.2) | .05 | 0.5 (0.6) | 0.7 (1.4) | –0.2 (1.5) | .10 |
| Sedentary behavior (hours/week) | 10.7 (7.8) | 8.3 (4.9) | –2.4 (7.5) | <.001 | 11.8 (8.3) | 9.7 (6.9) | –2.1 (5.7) | <.001 |
| Total PA (minutes/week) | 123.4 (209.0) | 250.5 (535.3) | 127.1 (574.6) | .02 | 117.2 (123.9) | 161.7 (204.3) | 44.4 (224.8) | .04 |
| Steps (n/day) | 6224.2 (2048.5) | 9214.3 (3205.5) | 2990.1 (2892.5) | <.001 | 6272.1 (2024.5) | 8945.5 (3843.3) | 2673.4 (3124.6) | <.001 |
| **MetS[h] factors** | | | | | | | | |
| Number, mean (SD) | 2.7 (1.0) | 2.0 (1.1) | –0.7 (1.1) | <.001 | 3.0(1.0) | 2.1(1.1) | –0.8 (1.1) | <.001 |
| 0, n (%) | 0 (0.0) | 12 (10.6) | N/A[i] | N/A | 0(0.0) | 8(7.5) | N/A | N/A |
| 1, n (%) | 16 (14.2) | 25 (22.1) | N/A | N/A | 10(9.2) | 24(22.2) | N/A | N/A |
| 2, n (%) | 35 (31.0) | 34 (30.1) | N/A | N/A | 25(23.1) | 36(33.3) | N/A | N/A |
| 3, n (%) | 32 (28.3) | 33 (29.2) | N/A | N/A | 33(30.6) | 25(23.1) | N/A | N/A |
| 4, n (%) | 26 (23.0) | 9 (8.0) | N/A | N/A | 33(30.6) | 14(13.0) | N/A | N/A |
| 5, n (%) | 4 (3.5) | 0 (0.0) | N/A | N/A | 7(6.5) | 1(0.9) | N/A | N/A |

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]WC: waist circumference.

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

[e]TG: triglyceride.

[f]$HBA_{1c}$: glycated hemoglobin.

[g]PA: physical activity.

[h]MetS: metabolic syndrome.

[i]N/A: not applicable.

## Changes in the Number of Steps to Meet Guidelines

Table 3 shows the changes in the number of steps before and after the intervention. For the prevention of chronic diseases, according to the guidelines for PA in Korea and the United States, taking at least 10,000 steps per day is recommended [28]. In the enhanced intervention group, there was a significant increase in the number of participants that met the recommended step count after the intervention. Although this number also increased in the standard intervention group, the change was not statistically significant. Table 3 shows the monthly average number of steps from baseline, demonstrating no significant differences in the initial step counts between the two intervention

groups, with the counts increasing in both groups until the fourth month. In the enhanced intervention group, the number of steps continued to increase from the beginning to the end of the trial, whereas in the standard intervention group, the number of steps gradually decreased from the fifth month until the last month.

**Table 3.** Proportion of participants meeting the step guidelines before and after the intervention.

| Steps per day | Baseline | Follow-up | *P* value |
|---|---|---|---|
| **Enhanced intervention group (n=113), n (%)** | | | .002 |
| <10,000 | 104 (92.0) | 87 (76.9) | |
| ≥10,000 | 9 (8.0) | 26 (23.1) | |
| **Standard intervention group (n=108), n (%)** | | | .72 |
| <10,000 | 100 (92.6) | 92 (85.2) | |
| ≥10,000 | 8 (7.4) | 16 (14.8) | |

# Discussion

## Principal Findings and Comparison With Prior Studies

The results of this study showed that the app of the wearable device itself was effective in reducing blood pressure, waist circumference, and $HbA_{1c}$. The addition of a mobile app along with the wearable device to the intervention also led to reduced weight and BMI. This result is consistent with a previous study that used a mobile app and a wearable device in an overweight cohort, and found that increasing PA reduced the risk of type 2 diabetes and lowered blood pressure through clinically significant weight loss [29]. In addition, the results of a study measuring the number of steps and activity levels in the elderly were similar to those obtained in the enhanced intervention group of this study regarding significant reductions in their weight and BMI [30]. Rowley et al [31] also found that an intervention in which participants were provided feedback via websites, along with education and self-management had more activity than that of the group that received only an activity meter. These results are attributed to the additional education on MetS disease, normal reference values, correct eating habits, and the importance of increased activity. However, this result is in contrast to the findings of another study showing that monitoring and smartphone apps did not significantly improve PA in patients with chronic obstructive pulmonary disease [32].

Increased PA has been shown to prevent or delay the onset of diabetes, heart disease, and chronic diseases [33]; thus, increasing the PA among groups with at least one major risk factor should be considered a priority for disease prevention.

In this study, the IPAQ was used to evaluate the aspects of PA that cannot be measured using wearable devices alone, while a wearable device was used to quantify movement data to assess PA. Accordingly, the number of steps was measured using wearable devices, and the amount of personal PA, including vigorous exercise and sitting time, was investigated by the IPAQ. Text messages and mobile apps are currently being used as interventions to increase PA owing to their affordability and convenience, and have been established as important strategies that can change the lifestyle of the wearer. However, few studies have investigated how wearable devices and apps can be used to prevent MetS through changes in PA in the middle-aged population living in rural areas of Korea. Wang et al [34] reported that using a wearable device and providing a text message for 6 weeks resulted in participants with obesity significantly increasing their activity by 1266 steps in 1 week; however, this was reduced to a mere increase of 24 steps in the consecutive weeks. Additionally, the total PA time decreased by 15 minutes. In this study, the number of steps increased by 2990 by the end of the clinical trial, and the total PA time increased by 127 minutes. In previous studies, a minimum of 150 minutes of moderate to vigorous physical activity (MVPA) per week was recommended to prevent cardiovascular disease and chronic diseases, which is also the recommended PA level in the Korean and US guidelines [7,35,36]. In this study, only the enhanced intervention group reached the recommended MVPA (188 minutes per week vs 118 minutes per week in the standard intervention group). Owolabi et al [37] reported that adherence to the recommended PA was low even after a text message intervention; however, in this study, intervention through the app and text messages had an effect on participants reaching the recommended goal. In addition, the enhanced intervention group showed higher persistence, whereas in the standard intervention group, the number of steps initially increased and then decreased from the middle to late part of the study period.

## Recommendations and Limitations

To our knowledge, this is one of the few studies that analyzed the impact of wearable devices on the promotion of PA in middle-aged people living in rural areas of Korea. Recently, the accumulation of clinical data through connected devices with mobile apps has soared. However, evaluation of the clinical measures obtained by using the intervention service has not yet reached the pace of development. In both the enhanced intervention group using wearable devices and mobile apps and the standard intervention group using only wearable devices, the PA of the participants improved. This result suggests that the improvement of PA in the standard intervention group may have been a psychological effect of participating in the study. However, the knowledge and information provided in text messages and the app served as a major factor in preventing MetS. The distribution of participants without any MetS factors was confirmed to increase by 12 (10.6%) among the 113 participants in the enhanced intervention group and by only 8 (7.5%) among the 108 participants in the standard intervention group. This finding is in contrast to the study by Jakicic et al

[38], who found that the protocol for monitoring PA and providing feedback did not have significant effects. According to Patel et al [39], wearable devices can promote health behavior change, but their successful use and potential health benefits depend more on the design of engagement strategies than on the characteristics of the technology. This highlights the need for personal encouragement, competitive spirit and collaboration, and effective feedback that are linked to human behavior.

A limitation of this study is that the participants were only recruited from rural areas in Korea; hence, the cohort was fairly homogeneous with a similar lifestyle and infrastructure. In the future, it will be necessary to provide information on cultural factors and rural geographical characteristics suitable for Koreans, and that proper diet and lifestyle modifications are needed to prevent chronic diseases. In addition, comparative studies between urban and rural residents in Korea are needed, along with examining the trends of lifestyle and clinical indicators, including various diseases and patient groups, through various wearable devices and advanced mobile apps.

## Conclusions

PA monitoring and intervention using a wearable device for 6 months effectively prevented MetS in rural participants in Korea. Moreover, blood pressure, waist circumference, and $HbA_{1c}$ levels improved in both intervention groups, which were effective in reducing MetS factors. However, there was a difference in the persistence of PA between the two groups. The enhanced intervention group continuously increased the amount of PA above the recommended level to prevent MetS, and as a result, body weight and BMI were further improved. Since the clinical values that can confirm the improvement of MetS may not improve in a short time, a longer-term study is needed.

## Acknowledgments

This research was supported by the Ministry of Trade, Industry & Energy, Korea Institute for Advancement of Technology through the Next-generation Life and Health Industrial Ecosystem Creation Project (R0006229).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1080 KB - mhealth_v10i2e34059_app1.pdf ]

## References

1. World Health Organization Department of Noncommunicable Disease Surveillance. Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. World Health Organization. 1999. URL: https://apps.who.int/iris/bitstream/handle/10665/66040/?sequence=1 [accessed 2021-12-15]
2. Kiani IG, Khan AN, Yasir S, Baluch UT. Frequency of metabolic syndrome in type-2 diabetes mellitus. J Ayub Med Coll Abbottabad 2016;28(1):59-62. [Medline: 27323564]
3. World Health Organization. Meeting report. 2018 Presented at: General meeting of the WHO global coordination mechanism on the prevention and control of noncommunicable diseases; November 5-6, 2018; Geneva, Switzerland URL: https://apps.who.int/iris/bitstream/handle/10665/310961/WHO-NMH-NMA-GCM-19.02-eng.pdf?sequence=1&isAllowed=y
4. Lee J, Kang M, Lee S. Effects of the e-Motivate4Change program on metabolic syndrome in young adults using health apps and wearable devices: quasi-experimental study. J Med Internet Res 2020 Jul 30;22(7):e17031 [FREE Full text] [doi: 10.2196/17031] [Medline: 32729838]
5. O'Donoghue G, Perchoux C, Mensah K, Lakerveld J, van der Ploeg H, Bernaards C, DEDIPAC Consortium. A systematic review of correlates of sedentary behaviour in adults aged 18-65 years: a socio-ecological approach. BMC Public Health 2016 Feb 17;16(1):163 [FREE Full text] [doi: 10.1186/s12889-016-2841-3] [Medline: 26887323]
6. Hancock S, Winterton R, Wilding C, Blackberry I. Understanding ageing well in Australian rural and regional settings: applying an age-friendly lens. Aust J Rural Health 2019 Aug;27(4):298-303. [doi: 10.1111/ajr.12497] [Medline: 31429143]
7. Jensen L, Monnat S, Green J, Hunter L, Sliwinski M. Rural population health and aging: toward a multilevel and multidimensional research agenda for the 2020s. Am J Public Health 2020 Sep;110(9):1328-1331. [doi: 10.2105/AJPH.2020.305782] [Medline: 32673118]
8. Trivedi T, Liu J, Probst J, Martin A. The metabolic syndrome: are rural residents at increased risk? J Rural Health 2013;29(2):188-197. [doi: 10.1111/j.1748-0361.2012.00422.x] [Medline: 23551649]
9. Siepmann C, Kowalczuk P. Understanding continued smartwatch usage: the role of emotional as well as health and fitness factors. Electron Markets 2021 Feb 18;31(4):795-809. [doi: 10.1007/s12525-021-00458-3]
10. Lee SM, Lee D. Healthcare wearable devices: an analysis of key factors for continuous use intention. Serv Bus 2020 Oct 15;14(4):503-531. [doi: 10.1007/s11628-020-00428-3]

11. Mizuno A, Changolkar S, Patel M. Wearable devices to monitor and reduce the risk of cardiovascular disease: evidence and opportunities. Annu Rev Med 2021 Jan 27;72:459-471. [doi: 10.1146/annurev-med-050919-031534] [Medline: 32886543]
12. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. "I must try harder": Design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019;10:2388. [doi: 10.3389/fpsyg.2019.02388] [Medline: 31749733]
13. Chong K, Guo J, Deng X, Woo B. Consumer perceptions of wearable technology devices: retrospective review and analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
14. Kim H, Park E, Lee S, Kim M, Park EJ, Hong S. Self-management of chronic diseases among older Korean adults: an mHealth training, protocol, and feasibility study. JMIR Mhealth Uhealth 2018 Jun 29;6(6):e147 [FREE Full text] [doi: 10.2196/mhealth.9988] [Medline: 29959109]
15. Lee A, Chavez S, Bian J, Thompson L, Gurka M, Williamson V, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
16. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
17. Sequi-Dominguez I, Alvarez-Bueno C, Martinez-Vizcaino V, Fernandez-Rodriguez R, Del Saz Lara A, Cavero-Redondo I. Effectiveness of mobile health interventions promoting physical activity and lifestyle interventions to reduce cardiovascular risk among individuals with metabolic syndrome: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e17790 [FREE Full text] [doi: 10.2196/17790] [Medline: 32865503]
18. Park E. Research in Brief. Changes in health behaviors among adults: a look through the Korean Health Panel Survey 2010~ 2018. Korea Institute for Health and Social Affairs (KIHASA). 2021 Aug 11. URL: http://repository.kihasa.re.kr/bitstream/201002/38150/1/20210811__Research_in_Brief_11.pdf [accessed 2022-01-02]
19. Lee D, Kim Y, Jekal Y, Park S, Kim K, Naruse M, et al. Low levels of physical activity are associated with increased metabolic syndrome risk factors in Korean adults. Diabetes Metab J 2013 Apr;37(2):132-139 [FREE Full text] [doi: 10.4093/dmj.2013.37.2.132] [Medline: 23641354]
20. King AC, Whitt-Glover MC, Marquez DX, Buman MP, Napolitano MA, Jakicic J, 2018 Physical Activity Guidelines Advisory Committee. Physical activity promotion: highlights from the 2018 Physical Activity Guidelines Advisory Committee systematic review. Med Sci Sports Exerc 2019 Jun;51(6):1340-1353. [doi: 10.1249/MSS.0000000000001945] [Medline: 31095090]
21. Duncan M, Vandelanotte C, Trost S, Rebar A, Rogers N, Burton N, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Jul 30;16(1):670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
22. Brickwood K, Watson G, O'Brien J, Williams A. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
23. Yi J, Kim Y, Cho Y, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
24. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
25. Wani R, Nabi S. Reliability and validity of a culturally adaptive version of the international physical activity questionnaire in indian subcontinent: A cross-sectional study. Int J Prev Med 2020;11(1):40-23. [doi: 10.4103/ijpvm.ijpvm_120_19]
26. Carsin A, Fuertes E, Schaffner E, Jarvis D, Antó JM, Heinrich J, et al. Restrictive spirometry pattern is associated with low physical activity levels. A population based international study. Respir Med 2019 Jan;146:116-123 [FREE Full text] [doi: 10.1016/j.rmed.2018.11.017] [Medline: 30665509]
27. IPAQ scoring protocol. The IPQA group. URL: http://www.ipaq.ki.se [accessed 2022-01-02]
28. Lee SH, Ha Y, Jung M, Yang S, Kang WS. The effects of a mobile wellness intervention with Fitbit use and goal setting for workers. Telemed J E Health 2019 Nov;25(11):1115-1122. [doi: 10.1089/tmj.2018.0185] [Medline: 30648920]
29. Fukuoka Y, Gay C, Joiner K, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
30. Suboc T, Strath S, Dharmashankar K, Harmann L, Couillard A, Malik M, et al. The impact of moderate intensity physical activity on cardiac structure and performance in older sedentary adults. Int J Cardiol Heart Vessel 2014 Sep 01;4:19-24 [FREE Full text] [doi: 10.1016/j.ijchv.2014.08.007] [Medline: 25530947]

31. Rowley T, Lenz E, Swartz A, Miller N, Maeda H, Strath S. Efficacy of an individually tailored, internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]
32. Vorrink SNW, Kort HSM, Troosters T, Zanen P, Lammers JJ. Efficacy of an mHealth intervention to stimulate physical activity in COPD patients after pulmonary rehabilitation. Eur Respir J 2016 Oct;48(4):1019-1029 [FREE Full text] [doi: 10.1183/13993003.00083-2016] [Medline: 27587557]
33. Jang I, Kim H, Lee E, Jung H, Park H, Cheon SH, et al. Impact of a wearable device-based walking programs in rural older adults on physical activity and health outcomes: cohort study. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e11335 [FREE Full text] [doi: 10.2196/11335] [Medline: 30463838]
34. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
35. Tucker J, Welk G, Beyler N. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]
36. Kim J. Longitudinal trend of prevalence of meeting physical activity guidelines among Korean adults. Exerc Med 2017 Jun 04;1:2. [doi: 10.26644/em.2017.002]
37. Owolabi E, Goon DT, Ajayi A. Impact of mobile phone text messaging intervention on adherence among patients with diabetes in a rural setting: A randomized controlled trial. Medicine 2020 Mar;99(12):e18953. [doi: 10.1097/MD.0000000000018953] [Medline: 32195927]
38. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
39. 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]


## Abbreviations

**$HbA_{1c}$:** glycated hemoglobin
**IPAQ:** International Physical Activity Questionnaire
**MET:** metabolic equivalents of task
**MetS:** metabolic syndrome
**mHealth:** mobile health
**MVPA:** moderate to vigorous physical activity
**PA:** physical activity


*Edited by L Buis; submitted 12.11.21; peer-reviewed by E Urtnasan, JU Park; comments to author 06.12.21; revised version received 20.12.21; accepted 21.12.21; published 24.02.22.*

*Please cite as:*
*Kim HJ, Lee KH, Lee JH, Youk H, Lee HY*
*The Effect of a Mobile and Wearable Device Intervention on Increased Physical Activity to Prevent Metabolic Syndrome: Observational Study*
*JMIR Mhealth Uhealth 2022;10(2):e34059*
*URL: https://mhealth.jmir.org/2022/2/e34059*
*doi:10.2196/34059*
*PMID:35200145*

Original Paper

# A Smartphone-Based Model of Care to Support Patients With Cardiac Disease Transitioning From Hospital to the Community (TeleClinical Care): Pilot Randomized Controlled Trial

Praveen Indraratna[1,2], MBBS; Uzzal Biswas[3], PhD; James McVeigh[1], MClinN; Andrew Mamo[1], MBBS; Joseph Magdy[1,4], MD; Dominic Vickers[1], MD; Elaine Watkins[1], BScN; Andreas Ziegl[5], MSc; Hueiming Liu[6], PhD; Nicholas Cholerton[1], MPT; Joan Li[1], MBBS; Katie Holgate[1], BASc; Jennifer Fildes[1], MPsych, RN; Robyn Gallagher[7], PhD; Cate Ferry[8], RN; Stephen Jan[6], PhD; Nancy Briggs[9], PhD; Guenter Schreier[5], PhD; Stephen J Redmond[3,10], PhD; Eugene Loh[4], MBBS; Jennifer Yu[1,2], PhD; Nigel H Lovell[3], PhD; Sze-Yuan Ooi[1,2], MD

[1]Department of Cardiology, Prince of Wales Hospital, Randwick, Australia
[2]Prince of Wales Clinical School, UNSW Sydney, Sydney, Australia
[3]Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, Australia
[4]Department of Cardiology, The Sutherland Hospital, Sydney, Australia
[5]Center for Health and Bioresources, Austrian Institute of Technology, Graz, Austria
[6]The George Institute for Global Health, Sydney, Australia
[7]Susan Wakil School of Nursing and Midwifery, Charles Perkins Centre, University of Sydney, Sydney, Australia
[8]National Heart Foundation of Australia, Sydney, Australia
[9]Stats Central, Mark Wainwright Analytical Centre, UNSW Sydney, Sydney, Australia
[10]School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland

**Corresponding Author:**
Praveen Indraratna, MBBS
Department of Cardiology
Prince of Wales Hospital
Barker Street
Randwick, 2031
Australia
Phone: 61 2 9382 2222
Email: praveen@unsw.edu.au

## Abstract

**Background:** Patients hospitalized with acute coronary syndrome (ACS) or heart failure (HF) are frequently readmitted. This is the first randomized controlled trial of a mobile health intervention that combines telemonitoring and education for inpatients with ACS or HF to prevent readmission.

**Objective:** This study aims to investigate the feasibility, efficacy, and cost-effectiveness of a smartphone app–based model of care (TeleClinical Care [TCC]) in patients discharged after ACS or HF admission.

**Methods:** In this pilot, 2-center randomized controlled trial, TCC was applied at discharge along with usual care to intervention arm participants. Control arm participants received usual care alone. Inclusion criteria were current admission with ACS or HF, ownership of a compatible smartphone, age ≥18 years, and provision of informed consent. The primary end point was the incidence of unplanned 30-day readmissions. Secondary end points included all-cause readmissions, cardiac readmissions, cardiac rehabilitation completion, medication adherence, cost-effectiveness, and user satisfaction. Intervention arm participants received the app and Bluetooth-enabled devices for measuring weight, blood pressure, and physical activity daily plus usual care. The devices automatically transmitted recordings to the patients' smartphones and a central server. Thresholds for blood pressure, heart rate, and weight were determined by the treating cardiologists. Readings outside these thresholds were flagged to a monitoring team, who discussed salient abnormalities with the patients' usual care providers (cardiologists, general practitioners, or HF outreach nurses), who were responsible for further management. The app also provided educational push notifications. Participants were followed up after 6 months.

**Results:** Overall, 164 inpatients were randomized (TCC: 81/164, 49.4%; control: 83/164, 50.6%; mean age 61.5, SD 12.3 years; 130/164, 79.3% men; 128/164, 78% admitted with ACS). There were 11 unplanned 30-day readmissions in both groups (*P*=.97). Over a mean follow-up of 193 days, the intervention was associated with a significant reduction in unplanned hospital readmissions (21 in TCC vs 41 in the control arm; *P*=.02), including cardiac readmissions (11 in TCC vs 25 in the control arm; *P*=.03), and higher rates of cardiac rehabilitation completion (20/51, 39% vs 9/49, 18%; *P*=.03) and medication adherence (57/76, 75% vs 37/74, 50%; *P*=.002). The average usability rating for the app was 4.5/5. The intervention cost Aus $6028 (US $4342.26) per cardiac readmission saved. When modeled in a mainstream clinical setting, enrollment of 237 patients was projected to have the same expenditure compared with usual care, and enrollment of 500 patients was projected to save approximately Aus $100,000 (approximately US $70,000) annually.

**Conclusions:** TCC was feasible and safe for inpatients with either ACS or HF. The incidence of 30-day readmissions was similar; however, long-term benefits were demonstrated, including fewer readmissions over 6 months, improved medication adherence, and improved cardiac rehabilitation completion.

**Trial Registration:** Australian New Zealand Clinical Trials Registry ACTRN12618001547235; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375945

***(JMIR Mhealth Uhealth 2022;10(2):e32554)*** doi:10.2196/32554

**KEYWORDS**

digital health; telemedicine; mHealth; heart failure; ischemic heart disease; mobile phone

## Introduction

### Cardiovascular Disease

Cardiovascular disease remains the most prevalent cause of morbidity and mortality in high-income countries despite significant advances in treatment over the last 5 decades. Myocardial infarction is responsible for 15% of worldwide mortality [1], and heart failure (HF) affects >26 million people worldwide [2]. Recent epidemiological data show that cardiovascular mortality is no longer declining and is indeed rising in some communities [3], and hospitalization rates are universally increasing [4,5]. The principal drivers include an aging population and rising prevalence of adult and childhood obesity [6,7]. Coupled with increasing health care costs, these trends raise concerns regarding the sustainability of the already overburdened traditional model of health care.

Cardiac readmissions are a potential target for system improvement. Readmission rates for both acute coronary syndrome (ACS) and HF approach 20% for patients in the first month after discharge [8-10]. Readmissions are associated with increased mortality and costs for the health care system [11]. In Australia, the estimated annual cost of readmissions for HF exceeds Aus $600 million (US $463.7 million) [12]. A recent audit of 3 hospitals in the state of New South Wales reported that 27% of angina pectoris admissions and 63% of HF admissions were preventable [13].

Up to 45% of mortality from recurrent myocardial infarction is preventable [14]. Secondary prevention for both conditions (ACS and HF) is critical and involves maximizing medication compliance, self-care, and optimization of modifiable risk factors, including weight and blood pressure (BP). However, secondary prevention programs have suboptimal uptake. For ACS, the cornerstone of secondary prevention is cardiac rehabilitation (CR), which is only attended by 20% to 30% of eligible participants because of competing demands such as employment and family responsibilities as well as travel time and costs [15]. For HF, management using community HF teams is resource-intensive and not uniformly available.

### Telehealth

Telehealth, the provision of health care by means of telecommunication technology, is a valuable adjunct in the management of chronic diseases. Within the scope of telehealth is mobile health (mHealth), which uses ubiquitous mobile phone technology for service delivery. Broadly, mHealth interventions encompass SMS text messaging strategies and telemonitoring systems in the form of smartphone apps. Telemonitoring is the practice of remote transmission and receipt of physical parameters such as pulse rate, BP, and weight. A recent meta-analysis found that the use of mHealth interventions in cardiovascular disease was associated with an improvement in BP and HF hospitalization rates [16]. The most successful interventions included several key factors: a method of *flagging* abnormal results, involvement of the patients' usual health care providers, and automatic data transmission as opposed to manual data entry by the patients. Thus, from a collaboration between a team of hospital-based clinicians and biomedical engineers, the TeleClinical Care (TCC) smartphone app was developed to include all these factors. Crucially, the app contains an educational component in addition to telemonitoring, making it a rare multifunctional mHealth intervention to undergo a randomized controlled trial (RCT). The app was designed to be used by patients diagnosed with either ACS or HF to maximize uptake. It is the first mHealth telemonitoring intervention to be trialed in Australian patients with HF.

### Objectives

The primary objective is to examine the efficacy of the TCC model compared with usual care alone on the incidence of 30-day hospital readmission rates in patients recently discharged with ACS or HF. Secondary objectives include: (1) to describe the compliance rate with the intervention as well as the frequency of alerts and actions subsequently undertaken, (2) to examine the impact of the intervention on clinical outcomes,

(3) to examine the cost-effectiveness of the intervention, and (4) to measure patient satisfaction with the intervention.

## Methods

### Participants

Patients were recruited between February 2019 and March 2020 from 2 hospitals in Sydney, New South Wales, Australia (Prince of Wales Hospital and The Sutherland Hospital). Patients were eligible if they were being discharged after an admission for either HF or ACS, were aged ≥18 years, and owned a compatible smartphone (defined as operating either Apple iOS 9.0 or above, or Android 7.0 Nougat or above). The exclusion criteria were inability or unwillingness to provide informed consent, inability to operate the app because of physical or cognitive limitations, inability to attend in-person follow-up (such as participants who normally resided outside of Sydney) or travel overseas for any duration within the first 30 days after discharge or for a period of >1 month, or expected discharge to another hospital or a nursing home. Advanced age, comorbidities, and familiarity with smartphone apps were not used as inclusion or exclusion criteria. All patients who met the inclusion criteria were approached for participation. The participants did not receive any financial compensation during the trial.

### Ethics

This study received ethical approval from the South Eastern Sydney Local Health District Human Research Ethics Committee (approval number 2019/ETH11442). The study was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12618001547235).

### Enrollment

Patients were enrolled during the index admission after providing written informed consent. All baseline data were collected before discharge and before randomization. BP and weight were measured using an automatic digital sphygmomanometer (A&D Medical UA-651BLE) and a digital weighing scale (A&D Medical UC-352BLE). These same devices were provided to the participants assigned to the intervention arm. BP was measured in the seated position. A total of 2 measurements were taken 1 to 5 minutes apart and averaged. If the 2 systolic readings differed by >15 mm Hg, a third measurement was taken, and the 2 closest readings were averaged. Height was measured using a wall-mounted stadiometer. Waist circumference was measured halfway between the costal margin and iliac crest as per World Health Organization guidelines [17]. A 6-minute walk distance test was performed using a graduated 25-meter track with standardized encouragement according to the protocol described by the Lung Foundation of Australia [18]. The test was not performed on those unsafe to complete it because of frailty, unsteadiness, or physical limitations. Serum low-density lipoprotein cholesterol levels were measured in blood samples previously obtained during hospitalization. A written questionnaire was provided to the participants containing the Morisky–Green–Levine 4-item medication compliance (MGL) score [19], the 5-level EuroQol 5-dimension quality of life assessment [20], and the Patient Activation Measure [21]. All baseline data were collected by study investigators: PI and JMc at Prince of Wales Hospital, and AM and JM at Sutherland Hospital.

### Randomization

Before discharge, the participants were randomized 1:1 into either TCC plus usual care or usual care alone. Block randomization was performed using a randomization schedule created by an independent statistician, which was subsequently deployed within a web-based system (Research Electronic Data Capture) [22]. Randomization strata included hospital and primary diagnosis (ACS or HF). Randomization was performed by the investigator who collected the baseline data (PI and JMc at Prince of Wales Hospital, and AM and JM at Sutherland Hospital).

### Intervention

The participants assigned to the intervention arm received the TCC app (Figure 1) on their smartphone and connected Bluetooth peripheral devices at the time of discharge: a digital sphygmomanometer, a digital weighing scale, and a fitness band (Xiaomi MiBand 2; Figure 2). The participants were instructed to measure BP and pulse rate via the sphygmomanometer, as well as weight, daily. Before discharge, the participants were shown how to use the devices and performed 1 measurement with each device under the supervision of the research team member to ensure the correct technique. The participants were also provided with a pamphlet that described the correct technique for using the devices and some basic troubleshooting advice. Activity data were obtained either via the smartphone or the fitness band as minutes of activity per day. Readings could be performed at any time relative to medication dosing. Readings were automatically transmitted from the peripheral devices to the smartphone app via Bluetooth and subsequently to a web-based server (KIOLA; Figure 3) developed at the Austrian Institute of Technology and adapted for the Australian context by the technical members of our team. Readings could be displayed within the app in graphical form for viewing by the patient. These graphs could be presented to the patient's general practitioner (GP) or cardiologist at follow-up visits, but this was not mandated. The app provided 3 weekly educational push notifications to promote healthy behavior choices, including dietary advice, physical exercise, and smoking cessation. The text for these notifications was based on the National Heart Foundation of Australia's *Managing My Heart Health* consumer resource [23].

If readings had not been received by the server for >48 hours, the participant was contacted by a biomedical engineer to ascertain if there had been a technical issue. If the patient admitted noncompliance with the program on 3 separate occasions, they were not contacted again for absent readings.

**Figure 1.** Screenshots of the TeleClinical Care (TCC) app. From left to right: the TCC app home screen, the appearance of an educational notification, weekly record of blood pressure readings, and weekly record of weight readings.

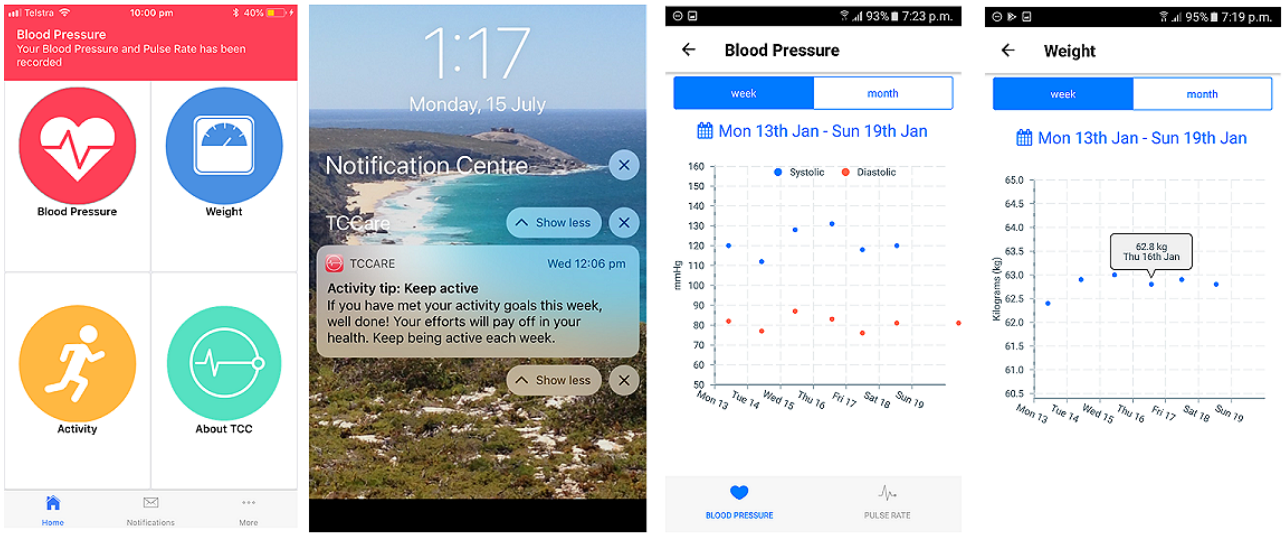
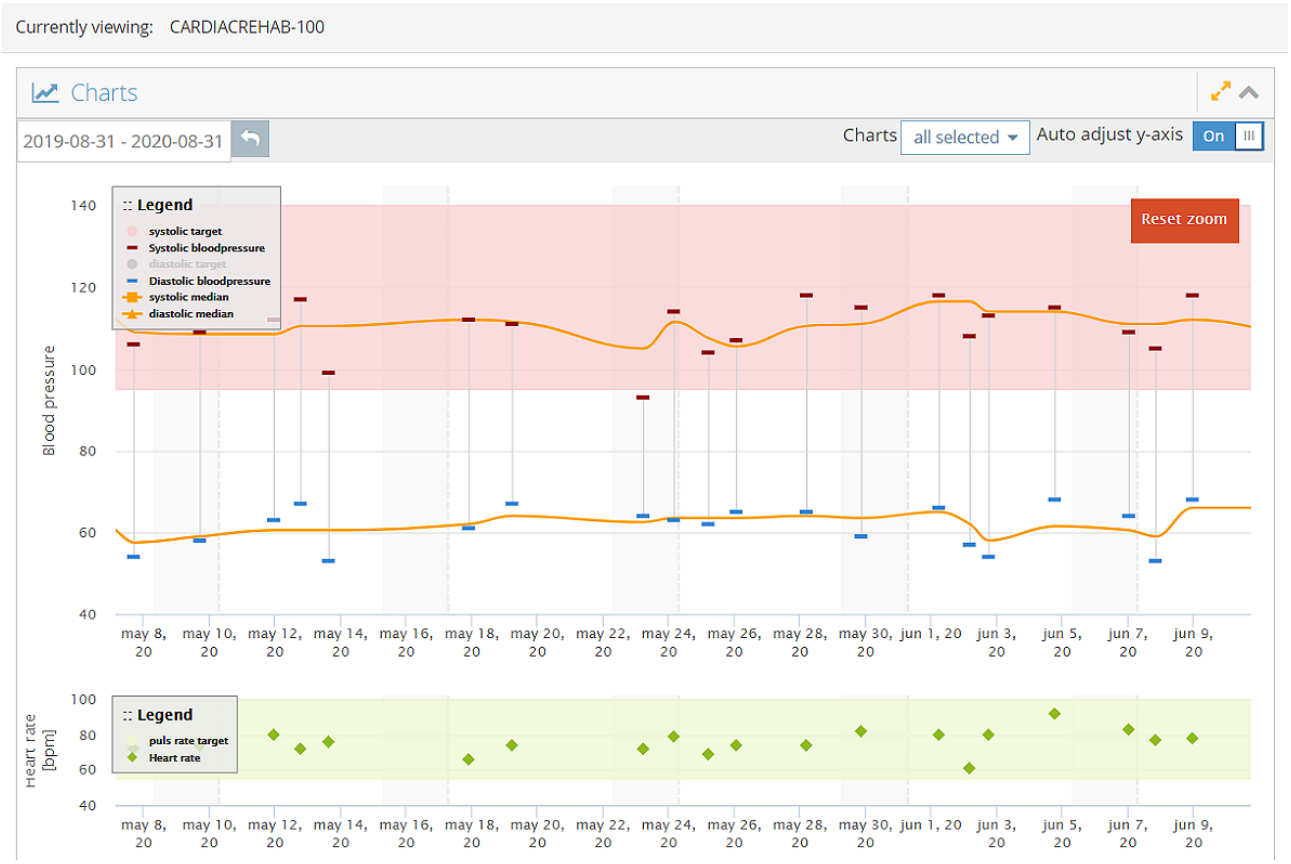



**Figure 2.** Bluetooth-enabled peripheral devices. From left to right: sphygmomanometer (A&D Medical UA-651BLE), weighing scale (A&D Medical UC-352BLE), and activity monitor (Xiaomi MiBand 2).

**Figure 3.** Screenshot of the KIOLA back-end, which is visible to monitoring clinicians. Blood pressure and pulse rate are recorded when the data are sent from the Bluetooth-enabled sphygmomanometer. Readings outside the shaded zone automatically trigger an email alert to the monitoring clinicians. Bpm: beats per minute.




For each patient, customizable limits for BP, pulse rate, and weight gain were defined at the time of discharge in consultation with the treating cardiologist. This was a 2-tier system of *yellow* (low priority) and *red* (high priority) alerts. For example, for a particular patient, a systolic BP >180 mm Hg could be defined as a red alert, and a systolic BP of 160-179 mm Hg could be defined as a yellow alert. The limits could be modified during the trial at the discretion of the monitoring team. If a reading returned outside of the defined limits, an alert was delivered by email to the monitoring team, which consisted of a cardiologist and a cardiac nurse practitioner who alternated monitoring duties. Emails were monitored from 8 AM to 5 PM on weekdays. Alerts delivered after hours, on weekends, or on public holidays were assessed the following weekday. Upon reviewing an alert, the monitoring clinician would decide whether to contact the patient and, upon doing so, assess whether the alert required escalation to the patient's GP or cardiologist. Patients were mandatorily contacted following receipt of any red alert. For yellow alerts, the monitoring team contacted the patients based on their own discretion. For example, alerts that were clearly erroneous (eg, a weight reading of 150 kg in a patient who normally weighed 75 kg) or those that rapidly normalized or were only marginally above the threshold and not considered clinically significant did not mandatorily require patient contact. Decisions to alter management or order investigations were made by the patient's GP or cardiologist and not by the monitoring team. All alerts, response details, and outcomes were recorded. Usual care, provided in both arms, included a recommendation to follow up with the GP within 1 week of discharge and with the treating cardiologist, who determined the timing of this visit. Patients with ACS were referred to CR, and patients with HF were referred to the local HF outreach service.

## Outcome Parameters

The participants were followed up at 6 months. This occurred in person until March 2020 and then by telephone after COVID-19 was defined as a global pandemic. The primary outcome was the number of readmissions at 30 days, which was chosen because early readmissions are designated as a key hospital performance indicator by the state government. The occurrence of readmissions as well as the length of stay were confirmed by patient interviews, review of the local electronic medical records, and the Australian national health database (MyHealthRecord). A readmission was defined as an unplanned return to hospital, either via the emergency department (ED) or direct admission, resulting in the acceptance of care of the patient by any inpatient medical team. Planned admissions, ED presentations that resulted in discharge without inpatient admission, and admissions to the ED short stay unit were not considered readmissions for the purpose of this study. The cause of the readmission was determined by the summary diagnosis given in the discharge summary and was classified as noncardiac

or cardiac. CR attendance was defined as presence during at least one session. CR completion was defined as attendance to ≥10 sessions or formal discharge by the CR staff. CR attendance was routinely recorded in the patient's electronic medical record by the CR staff at both hospitals. Only patients with ACS were included in this analysis as patients with HF are not routinely referred to CR at either institution. The analysis was limited to those enrolled ≥2 months before the closure of CR for COVID-19. For in-person visits, physical parameters were measured by blinded investigators. During the COVID-19 pandemic, the final BP (average of the last 2 readings) and weight were obtained from the readings submitted via the app for those in the intervention arm; however, the corresponding values were not obtained from the participants in the control arm. Follow-up blood tests were not mandated during the pandemic. The participants completed the same questionnaires at baseline by either written or telephone means depending on whether the follow-up date was before or during the pandemic. The participants in the intervention arm completed an evaluation of the TCC program (user experience questionnaire) in either written, telephone, or web-based form (Multimedia Appendix 1). This questionnaire was designed specifically for this study. Alerts were defined as clinically significant alerts if they led to a change in investigation or management or led directly to a consultation with a health care professional. Major adverse cardiovascular events were defined as a composite of all-cause death, nonfatal myocardial infarction, and nonfatal stroke.

### Statistical Analysis

As this was a pilot study, the sample size was not determined by a formal power calculation. Readmission analysis was performed using the Andersen–Gill Cox regression model. Single continuous variables were analyzed using the 2-tailed *t* test. Repeated measures were analyzed using linear mixed models. Nonparametric variables were analyzed using the Mann–Whitney *U* test. Single categorical variables were analyzed using the Pearson chi-square test. Repeated categorical variables were analyzed using generalized linear mixed models. Linear and generalized linear mixed models generated both a time interaction (change in parameters from baseline to follow-up) and a group-by-time interaction (change in parameters over time and between groups). Statistical analysis was performed using Stata Statistical Software: Release 16 (StataCorp LLC) and IBM SPSS Statistics for Windows, version 26.0. All analyses applied the intention-to-treat principle.

### Cost-Effectiveness

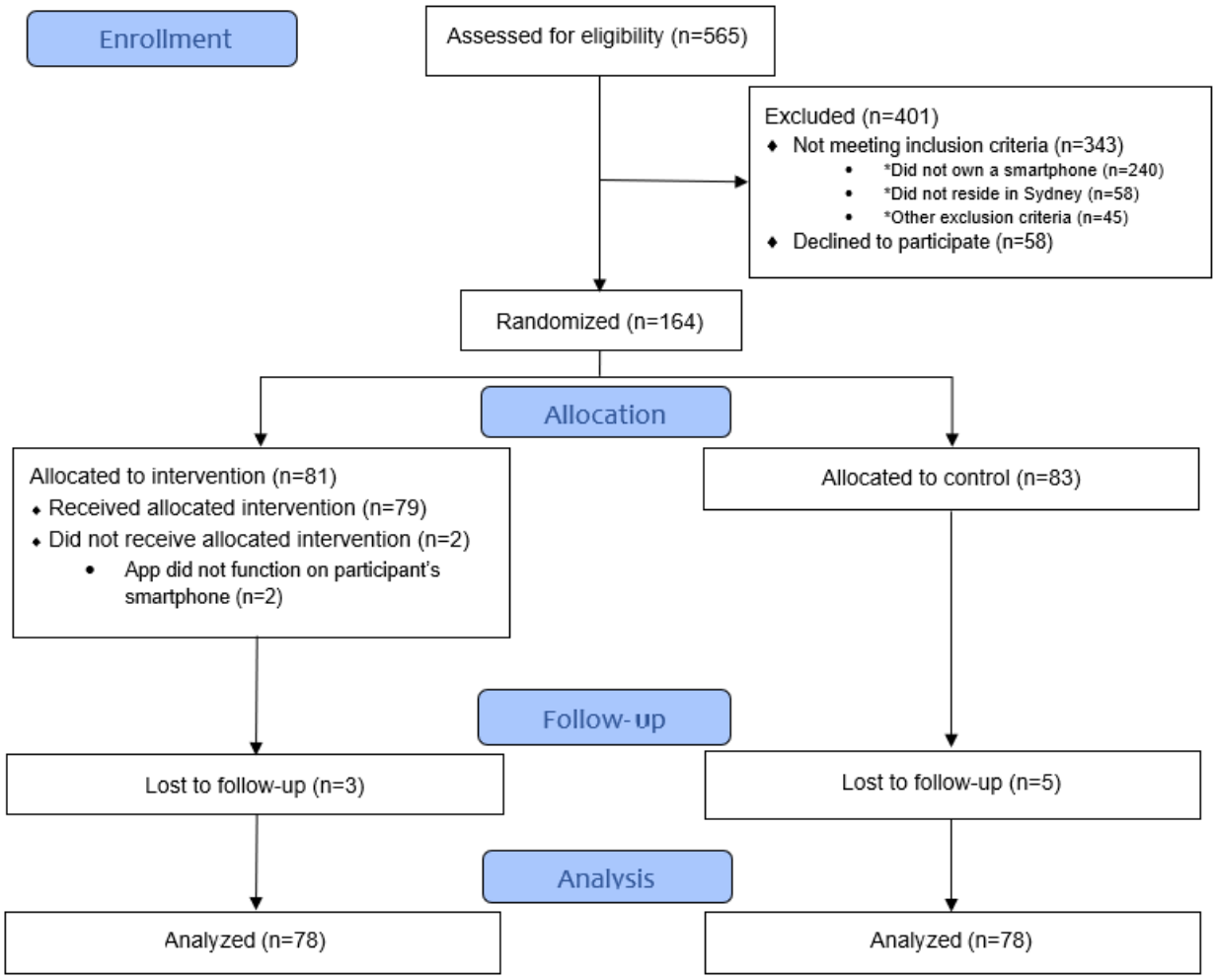
Running costs were recorded over the duration of the trial. Components of the running costs included the cost of equipment, staffing costs, server maintenance costs, and the cost of health care consultations generated by the system. A figure of cost per cardiac readmission saved was calculated by dividing the total cost incurred by the difference in cardiac readmission rates between the 2 groups. As costs in the research setting were unlikely to reflect mainstream clinical practice, a 12-month *real world* cost-effectiveness model was undertaken (Multimedia Appendix 2).

## Results

### Screening, Enrollment, and Follow-up

Between February 2019 and March 2020, 565 potential participants were screened for eligibility, of which 240 (42.5%) did not own a compatible smartphone, which was the most common reason for exclusion. Approximately 28.5% (161/565) of patients met ≥1 of the remaining exclusion criteria, the most common reasons being unwillingness to participate and living outside Sydney (and being unable to return for in-person follow-up; 58/161, 36% of patients in each case). A total of 128 patients with ACS and 36 patients with HF were enrolled for a total of 164 participants (Figure 4). Enrollment was terminated early at the onset of the COVID-19 pandemic. The app did not operate on the smartphones of 2 patients in the intervention arm (2/164, 1.2%). Another patient chose not to use the app after randomization because of a new diagnosis of lung cancer, although he did not withdraw from the study. These 3 patients (3/164, 1.8%) and all others randomized into the intervention regardless of compliance were included as part of the intention-to-treat analysis. The mean follow-up time was 193 days. Of the 164 patients, 8 (4.9%) were lost to follow-up.

**Figure 4.** Enrollment flowchart.




## Baseline Characteristics

The mean age was 61.5 years, and 79.3% (130/164) of patients were men (Table 1). Approximately 25.6% (42/164) of patients had moderate or severe left ventricular dysfunction. Most patients received guideline-directed medical therapy at baseline (Multimedia Appendix 3).

**Table 1.** Baseline characteristics of the enrolled cohort (N=164).

| Characteristic | TCC[a] (n=81) | Control (n=83) |
|---|---|---|
| Age (years), mean (SD) | 61.3 (12.3) | 61.7 (12.6) |
| **Gender, n (%)** | | |
| Male | 65 (80) | 65 (78) |
| Female | 16 (20) | 18 (22) |
| **Clinical characteristics** | | |
| ACS,[b] n (%) | 63 (78) | 65 (78) |
| HF,[c] n (%) | 18 (22) | 18 (22) |
| Moderate or severe LV[d] dysfunction, n (%) | 21 (26) | 21 (25) |
| Current smoker, n (%) | 18 (22) | 21 (25) |
| Atrial fibrillation, n (%) | 16 (20) | 20 (24) |
| Hypertension, n (%) | 39 (48) | 46 (55) |
| Diabetes, n (%) | 21 (26) | 22 (27) |
| Chronic kidney disease, n (%) | 11 (14) | 12 (14) |
| Systolic BP[e] (mm Hg), mean (SD) | 119 (18) | 121 (18) |
| Weight (kg), mean (SD) | 85.0 (16.8) | 87.9 (22.3) |
| BMI (kg/m$^2$), mean (SD) | 28.5 (4.5) | 30.1 (6) |
| Waist circumference (cm), mean (SD) | 100 (13) | 104 (16) |
| 6-minute walk test distance (m), mean (SD) | 385 (119) | 353 (124) |
| LDL-C[f] (mmol/L), mean (SD) | 2.33 (0.9) | 2.26 (1.05) |
| 5-level EuroQol 5-dimension calculated score (−0.10 to 1.00), mean (SD) | 0.84 (0.17) | 0.80 (0.17) |
| Self-reported quality of life score (0-100), mean (SD) | 66.7 (18) | 63.1 (21) |
| MGL[g] score (0-4) | 3.11 | 3.29 |
| Patient Activation Measure (0-100), mean (SD) | 64.5 (15) | 63 (16) |

[a]TCC: TeleClinical Care.

[b]ACS: acute coronary syndrome.

[c]HF: heart failure.

[d]LV: left ventricular.

[e]BP: blood pressure.

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

[g]MGL: Morisky–Green–Levine 4-item medication compliance.

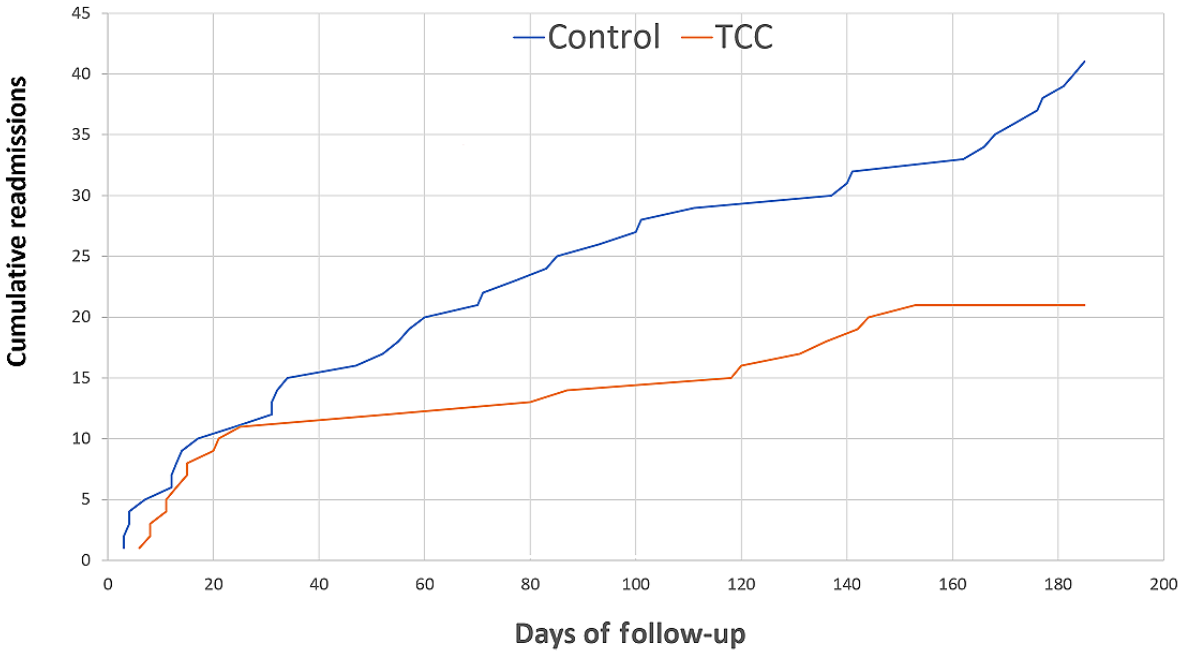
## Readmissions at 30 Days

All-cause, unplanned readmissions at 30 days were similar in the 2 groups (11 in the intervention arm and 11 in the control arm; *P*=.97).

## Total Readmissions

At 6 months, the intervention was associated with a reduction in all-cause, unplanned readmissions, with a total of 21 readmissions in the intervention arm and 41 readmissions in the control arm (hazard ratio [HR] 0.51, 95% CI 0.31-0.88; *P*=.02; Figure 5). Cardiac readmissions were also less common in the intervention arm (11 in the intervention arm vs 25 in the control arm; HR 0.44, 95% CI 0.22-0.90; *P*=.03). There was a numeric reduction in noncardiac readmissions during the study period, which did not reach statistical significance (10 in the intervention arm vs 16 in the control arm; HR 0.64, 95% CI 0.29-1.40; *P*=.26). Among patients with HF, there were 5 cardiac readmissions in the intervention arm and 18 readmissions in the control arm; however, this difference did not reach statistical significance, likely because of the smaller patient population.

**Figure 5.** Cumulative readmissions over the course of the trial. TCC: TeleClinical Care.




## Compliance With the Intervention

The 2 patients for whom the app did not function on their smartphones were excluded from the compliance analysis. The average percentage of days that the participants transmitted data was 64.2% (SD 27.5%). BP and weight transmissions occurred at equal frequencies (64.2% of days each). Of the 79 patients, 60 (76%) transmitted data on >50% of days. Approximately 52% (42/81) of patients transmitted data on an average of ≥5 days per week (ie, more than 71% of all days). Approximately 20% (16/79) of patients transmitted data for less than an average of 3 days per week.

## Alerts

A total of 585 (2.5%) alerts were generated out of 23,401 transmissions, of which 419 (71.6%) were for the 63 patients with ACS (mean 6.7 alerts per patient), and 166 (28.4%) were for the 18 patients with HF (mean 9.2 alerts per patient; Multimedia Appendix 4). Of the 79 patients, 11 (14%) did not generate any alerts. On the basis of their interpretation of the alerts, the monitoring clinicians chose to contact patients after 30.9% (181/585) of alerts, with a mean and median response time of 12.5 hours and 5.0 hours, respectively. Approximately 12.5% (73/585) of alerts required discussion with one of the patient's health care professionals. The remaining alerts were either erroneous, rapidly normalized, or not of clinical concern (Multimedia Appendix 4). A total of 54 health care consultations were generated from the alerts. The timing of 83% (45/54) of these consultations was known, and the mean time to consultation was 56 hours (median 26 hours). Approximately 16.1% (94/585) of alerts were clinically significant. Of the 79 patients, 42 (53%) did not generate any clinically significant alerts. The causes of the alerts can be found in Multimedia Appendix 4.

## Clinical Outcomes

A total of 4 deaths occurred in the control arm (4/83, 5%) and 1 in the intervention arm (1/81, 1%). All deaths were of cardiovascular causes. There was 1 nonfatal myocardial infarction in the intervention arm (1/81, 1%) and none in the control arm. No strokes occurred in either group. There was no statistically significant difference in mortality or major adverse cardiovascular events (Table 2).

**Table 2.** Major adverse cardiovascular events (MACEs; N=164).

| Clinical outcome | TCC[a] (n=81) | Control (n=83) | Relative risk (95% CI) | *P* value |
|---|---|---|---|---|
| Mortality | 1 | 4 | 0.25 (0.03-2.24) | .22 |
| Nonfatal MI[b] | 1 | 0 | 3.07 (0.13-74.3) | .49 |
| Nonfatal stroke | 0 | 0 | —[c] | — |
| MACEs | 2 | 4 | 0.51 (0.10-2.72) | .43 |

[a]TCC: TeleClinical Care.

[b]MI: myocardial infarction.

[c]Not possible to calculate as there were no events.

## CR Attendance and Completion

There was no significant difference in CR attendance rates. However, there were statistically significant differences in CR completion rates, both as a proportion of participants who attended and as a proportion of the total group population (Table 3).

**Table 3.** Cardiac rehabilitation completion rates (N=100).

| Parameter | TCC[a] (n=51), n (%) | Control (n=49), n (%) | Statistical analysis | |
|---|---|---|---|---|
| | | | OR[b] (95% CI) | *P* value |
| Attendance rate | 28 (55) | 21 (43) | 1.62 (0.74-3.58) | .23 |
| Completion rate (attendees only) | 20 (71)[c] | 9 (43)[d] | 3.30 (1.01-11) | .04 |
| Completion rate | 20 (39) | 9 (18) | 2.90 (1.15-7.17) | .02 |

[a]TCC: TeleClinical Care.

[b]OR: odds ratio.

[c]n=28.

[d]n=21.

## Physical Parameters

Because of the cancellation of in-person visits, these outcomes could not be assessed for many participants. The results are summarized in Multimedia Appendix 5.

## Questionnaire Results

At baseline, 10 patients in the intervention arm (10/81, 12%) and 14 in the control arm (14/83, 17%) did not use regular medications. These patients did not complete the MGL questionnaire at baseline but were instructed to complete it at follow-up. The proportion of patients who reported good adherence (defined by an MGL score of 4/4) improved significantly in the intervention arm (34/71, 48% to 57/76, 75%; *P*<.001). In the control arm, this proportion fell from 61% (42/69) to 50% (37/74; *P*=.19). Overall, there was a significant interaction favoring the intervention arm (*P*=.002).

The self-reported quality of life score from the 5-level EuroQol 5-dimension questionnaire improved significantly in both groups, but there was no difference between groups. The Patient Activation Measure score improved significantly in both groups, but there was no difference between groups (Multimedia Appendix 6).

## User Experience

Of the 81 participants, 66 (81%) completed the questionnaire. Reasons for noncompletion included limited use of the app, inadequate understanding of English, or declining to participate. The average rating out of 5 given for the app was 4.56. Approximately 96% (64/67) of users rated it as *easy* or *very easy* to use.

## Cost-Effectiveness

### *Trial Costs*

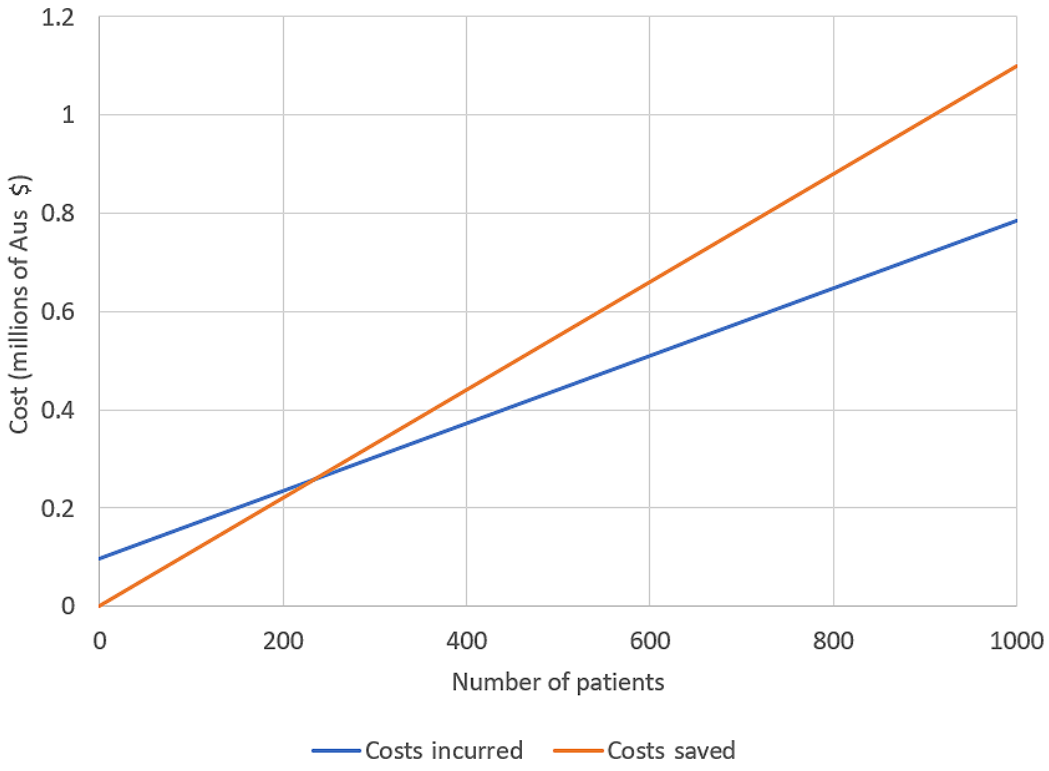
The trial ran for 20 months. Staffing costs were Aus $53,435 (US $38,491.80), which comprised total remuneration for staff responsible for enrolling and monitoring of participants. Technical support was provided as in-kind support. Equipment for the 81 participants in the intervention arm had a total cost of Aus $18,630 (US $13,420.10), and server maintenance costs totaled Aus $9000 (US $6483.14). The trial generated 18 additional GP visits with a total cost of Aus $698 (US $502.80) and 17 cardiologist visits with a total cost of Aus $1343 (US $967.43). HF outreach services do not have a defined per-visit cost; thus, no additional costs were included. Thus, we calculated the total cost of the intervention as Aus $82,408 (US $59,362.50). There were 14 fewer cardiac readmissions in the control arm, which was adjusted to 13.67 given the slightly higher number of patients in the control arm. Thus, the cost per cardiac readmission saved was Aus $6028 (US $4342.26).

In the control arm, the total cost of cardiac readmissions for all patients combined was Aus $85,213 (US $61,383). In the intervention arm, the equivalent cost was Aus $38,640 (US $27,834.30). The reduction in costs from readmission avoidance was Aus $550 (US $396.19) per patient for a 6-month participation, which was doubled to Aus $1100 (US $792.38) for the projected 12-month real-world model.

### *Model for 12 Months*

In this model, it was calculated that each patient would require approximately 5.8 hours of attention from the monitoring team if enrolled for 12 months (Multimedia Appendix 7). The standard per-hour nursing cost was Aus $49.85 (US $35.91), thus generating a per-patient nursing cost of Aus $289 (US $208.18). For a single nurse working 40 hours per week, it was estimated that they could simultaneously monitor up to 358 patients. Equipment costs were revised to Aus $200 (US $144.07) per patient as the MiBand 2 activity monitor was not intended for future use. Costs generated because of medical consultations were doubled to reflect a 12-month participation, as were costs saved by avoiding readmissions. Outside of the research setting, technical costs were projected to be higher because of the requirement of a commercial license for the KIOLA platform and technical support. A baseline cost of Aus $92,388 (US $66,551.50) was applied, as well as an annual cost of Aus $184 (US $132.54) per patient. A graph of the cost-effectiveness model is shown in Figure 6. According to this model, when the number of enrolled patients is ≥237, the costs saved from the prevention of readmissions will surpass all incurred costs.

**Figure 6.** Cost-Effectiveness of the TeleClinical Care model as described by total costs incurred by the system and total costs saved by projected cardiac readmission prevention. The x-axis represents the number of patients enrolled, and the y-axis represents the cost in millions of Aus $.




## Discussion

### Principal Findings

The TCC program, which combined telemonitoring and educational messaging within a smartphone app, was not associated with a reduction in readmissions at 30 days in patients discharged after an admission with ACS or HF, although event rates were low, particularly for cardiac readmissions. However, the intervention showed benefit with respect to reducing the incidence of cardiac and all-cause readmissions over the 6-month study period, as well as an improvement in CR completion rate and medication compliance.

There is a paucity of data for mHealth interventions targeting patients with ACS or a general inpatient population such as the one examined in this study. Telemedical interventions for HF have yielded varying results, which can be explained by the heterogeneous nature of the interventions. A variety of protocols and patient populations have been previously described and, therefore, drawing parallels between trials is often troublesome.

A large early trial, *Tele-HF* (2010), used a voice-interactive system and included body weight as the sole physical parameter without measurement of pulse or BP [24]. Compliance was poor, with approximately only 50% of participants taking part ≥3 times per week. This trial was negative for its primary end point (all-cause death or readmission within 180 days of enrollment) but demonstrated possible improvements for future telemedicine systems.

An example of how varying intervention design and delivery may influence results is seen by comparison of the Telemedical Interventional Monitoring in Heart Failure (*TIM-HF*) [25] and *TIM-HF2* [26] trials, both conducted by the same German group. In the *TIM-HF*, 710 stable outpatients were enrolled, and electrocardiogram, BP, and body weight results were transmitted via a PDA and mobile phone service where they were reviewed by an independent clinician who communicated with the patients' usual practitioner every 3 months. No difference in mortality or readmission was observed. In the *TIM-HF2*, a study of 1571 patients, a similar system was applied but with the addition of regular interaction with the patient's GP and cardiologist. Here, a reduction in percentage days lost because of cardiovascular readmissions and all-cause death was observed (4.88% vs 6.64%; *P*=.046). The authors identified the ability to guide the patient's management through their usual provider as a contributor to the success of the trial.

In comparison with these interventions, TCC had several advantages. The participants were required to measure BP, weight, and activity daily and, unlike several other mHealth studies in patients with cardiac disease [27-29], the results were automatically transmitted, thus removing the need for participants to manually enter data, which is burdensome and potentially error-prone. By aiming for daily transmission, there was a significant volume of data to detect trends in readings and to contextualize abnormalities. Weekly data entry, as has previously been described [29], is unlikely to be sensitive enough to detect clinical deterioration and thus prevent readmissions.

The alerts were automated, which allowed the monitoring team to efficiently identify the patients that required attention. The monitoring team, which consisted of a cardiologist and cardiac nurse practitioner, had significant clinical experience and was comfortable in deciding which alerts were clinically significant and which were not. As a result, <10% of all alerts received (54/585, 9.2%) led directly to a health care consultation. The involvement of the patients' usual health care providers was also important as it is assumed that knowledge of the patient and their medical background is key to the interpretation and management of alerts. To date, only 1 individual mHealth RCT has demonstrated a reduction in readmissions in patients with HF. This study by Dendale et al [30] randomized 160 patients with HF and similarly used automated data transmission combined with interaction with the patients' usual health care providers. On average, the patients generated 27 alerts over the 6-month period, which was higher than what was observed in this study.

Compliance with TCC was reasonable, with participants transmitting data on approximately 64% of study days. This is lower than in other mHealth studies, which have reported compliance rates of 80% to 95% [25,28-31]. Although 1 study reported using automated phone calls to improve compliance [31], others did not report on the level of encouragement participants were given to transmit data daily. Our study used a *three strikes* policy and, beyond that, noncompliant patients were not reminded to perform measurements. An automated system may have improved compliance and, thus, the overall results.

Patients with ACS in the intervention arm were more likely to complete CR, which is consistent with previous studies [32]. Although the intervention did not directly encourage patients to attend, it is hypothesized that the daily routine of taking measurements and the educational notifications of TCC helped engage patients and promote self-care. Thus, the benefits of TCC may have been amplified by the benefits of attending a full course of CR.

A similar principle may explain the improvement in medication adherence observed in TCC; that is, that patient education and increased engagement reinforced the importance of medication adherence in the management of their cardiac condition. It is also hypothesized that the daily requirement to measure BP was a memory trigger for taking medications.

Thus, the reduction in cardiac readmissions observed at 6 months is likely a consequence of a multifactorial mechanism. Deteriorations in the patients' physical condition were identified and managed in the outpatient sector, and improved self-care leading to higher engagement with CR and medication adherence is likely to have made such deteriorations less likely to occur in the first place. There was no significant reduction in the incidence of 30-day readmissions, suggestive of a medium- to long-term benefit of the intervention rather than an immediate one.

The participants generally found TCC easy to use, which likely improved app compliance. Although the app did not function for 2 participants (2/164, 1.2%), all other technical issues were remedied during the trial, and no discontinuations because of technical issues occurred. The app design was optimized for older adult patients with features such as large buttons and graphical displays. The use of Bluetooth synchronization eliminated the need for manual data entry, thereby easing the work burden on the patients. In the cardiovascular mHealth space, usability data have generally been underreported; thus, there is minimal scope for comparison with other apps.

During the trial itself, the costs incurred outweighed the costs of cardiac readmissions saved. The primary contributor to this were the constant staffing costs incurred regardless of patient load. For example, at the commencement of the trial, when a small number of enrollments had occurred, and at the end of the trial, after recruitment had ceased because of COVID-19, staffing costs were incurred at the same rate as during the peak of the trial.

The 12-month cost-effectiveness model demonstrated that costs saved will exceed costs incurred when >237 patients are recruited. There are several assumptions in this model; however, they generally underestimate the cost-effectiveness of TCC. For example, it was assumed that the rate of alerts, medical consultations, and readmissions would continue unchanged from months 6-12 of a participant's enrollment. This might not be the case if the participant's condition stabilized, and the readmission rates compared with the control group could potentially fall further. It was also assumed that the time taken to perform certain duties by staff would remain constant when, realistically, this should reduce as the staff members become more experienced. The cost-effectiveness models also did not consider the impact of noncardiac readmissions, which were not shown to be significantly different between the intervention and control arms in this study. Ultimately, the results of a larger RCT of this program will further inform cost-effectiveness models. For cardiovascular mHealth models, only 1 cost-effectiveness analysis has been published; however, it was in the context of an SMS text messaging intervention [33] rather than a telemonitoring system.

This study is limited by its relatively small sample size and the loss of data for several physical parameters, both of which were influenced by the COVID-19 pandemic. Thus, few conclusions can be drawn on several end points, including anthropometric measurements, 6-minute walk distance, and low-density lipoprotein cholesterol. However, because of the randomized nature of the trial, participants in both arms were equally affected by the pandemic, and the end points of 30-day and 6-month readmissions, as well as medication adherence, were not affected disproportionately in either arm. The generalizability of the results should be considered with caution. The intervention was only offered to participants who were smartphone owners. If a family member offered to share the use of their own phone for the study, this was not permitted. In the HF cohort, where the mean age was 79 years, smartphone ownership rates were low (41/224, 18.3%), thus limiting enrollment. It is not known whether similar results would have been achieved if patients lacking smartphones were provided with them. However, it is anticipated that smartphone ownership rates will continue to increase in all age groups; thus, future studies may enroll a higher proportion of older adult patients.

Whether the positive outcomes identified in the TCC study have continued ongoing long-term benefits after the completion of the trial remains unknown. There are limited data regarding residual benefits of mHealth interventions, although 1 study has reported improved BMI in patients 4 years after concluding the intervention, suggesting that learned behaviors may continue in the long term [34]. As most participants in this study consented to long-term data linkage analysis, it is possible that this question can be addressed in the future. Furthermore, a large, multicenter RCT of a modified TCC program powered for clinical end points is scheduled to commence in 2021. The primary end point will be unplanned hospital readmissions at 6 months, the participants will be followed for 12 months, and an additional SMS text messaging arm will be used for patients who own a mobile phone that is not capable of operating the new app.

## Conclusions

The TCC program is a novel and innovative model of care based on a smartphone app that facilitates telemonitoring and patient education. The system was demonstrated to be safe, feasible, patient-friendly, and cost-effective when applied to patients with ACS and HF at the time of discharge. Clinical benefits were observed regarding the rate of cardiac and all-cause readmissions, medication compliance, and CR completion. These results are promising, but confirmation with a larger trial is necessary before implementing widespread adoption of the model.

## Acknowledgments

This work was funded by the University of New South Wales, Prince of Wales Hospital, Prince of Wales Hospital Foundation, and Sutherland Hospital. PI is supported by a postgraduate research scholarship cofunded by the National Health and Medical Research Council and the National Heart Foundation of Australia. The authors acknowledge the assistance in the Research Electronic Data Capture survey and database design by David Jung of Research Technology Services at the University of New South Wales, Sydney.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
App user evaluation questionnaire.
[DOCX File , 122 KB - mhealth_v10i2e32554_app1.docx ]

Multimedia Appendix 2
Methods for the 12-month cost-effectiveness model.
[DOCX File , 13 KB - mhealth_v10i2e32554_app2.docx ]

Multimedia Appendix 3
Medications taken at baseline for patients in the intervention and control arms.
[DOCX File , 13 KB - mhealth_v10i2e32554_app3.docx ]

Multimedia Appendix 4
Summary of alerts generated and their designation after investigation by the monitoring team, alert classification including examples, and summary of the causes of total and clinically significant alerts received by patients with acute coronary syndrome or heart failure.
[DOCX File , 15 KB - mhealth_v10i2e32554_app4.docx ]

Multimedia Appendix 5
Physical parameters at follow-up.
[DOCX File , 14 KB - mhealth_v10i2e32554_app5.docx ]

Multimedia Appendix 6
Quality of life and Patient Activation Measure results.
[DOCX File , 13 KB - mhealth_v10i2e32554_app6.docx ]

Multimedia Appendix 7
Monitoring duties.
[DOCX File , 13 KB - mhealth_v10i2e32554_app7.docx ]

Multimedia Appendix 8
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1234 KB - mhealth_v10i2e32554_app8.pdf ]

## References


1. Jayaraj JC, Davatyan K, Subramanian S, Priya J. Epidemiology of Myocardial Infarction. London, UK: Intech Open; 2018.
2. Savarese G, Lund LH. Global public health burden of heart failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]
3. Zou Z, Cini K, Dong B, Ma Y, Ma J, Burgner DP, et al. Time trends in cardiovascular disease mortality across the BRICS: an age-period-cohort analysis of key nations with emerging economies using the global burden of disease study 2017. Circulation 2020 Mar 10;141(10):790-799. [doi: 10.1161/CIRCULATIONAHA.119.042864] [Medline: 31941371]
4. Lawson CA, Zaccardi F, Squire I, Ling S, Davies MJ, Lam CS, et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: a population-based study. Lancet Public Health 2019 Aug;4(8):406-420 [FREE Full text] [doi: 10.1016/S2468-2667(19)30108-2] [Medline: 31376859]
5. Agarwal MA, Fonarow GC, Ziaeian B. National Trends in Heart Failure Hospitalizations and Readmissions From 2010 to 2017. JAMA Cardiol 2021 Aug 01;6(8):952-956. [doi: 10.1001/jamacardio.2020.7472] [Medline: 33566058]
6. Lightwood J, Bibbins-Domingo K, Coxson P, Wang YC, Williams L, Goldman L. Forecasting the future economic burden of current adolescent overweight: an estimate of the coronary heart disease policy model. Am J Public Health 2009 Dec;99(12):2230-2237. [doi: 10.2105/AJPH.2008.152595] [Medline: 19833999]
7. Di Cesare M, Sorić M, Bovet P, Miranda JJ, Bhutta Z, Stevens GA, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019 Nov 25;17(1):212 [FREE Full text] [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
8. Al-Omary MS, Davies AJ, Evans T, Bastian B, Fletcher PJ, Attia J, et al. Mortality and readmission following hospitalisation for heart failure in Australia: a systematic review and meta-analysis. Heart Lung Circ 2018 Aug;27(8):917-927. [doi: 10.1016/j.hlc.2018.01.009] [Medline: 29519691]
9. Rana S, Tran T, Luo W, Phung D, Kennedy RL, Venkatesh S. Predicting unplanned readmission after myocardial infarction from routinely collected administrative hospital data. Aust Health Rev 2014 Sep;38(4):377-382. [doi: 10.1071/AH14059] [Medline: 25001433]
10. Labrosciano C, Horton D, Air T, Tavella R, Beltrame JF, Zeitz CJ, et al. Frequency, trends and institutional variation in 30-day all-cause mortality and unplanned readmissions following hospitalisation for heart failure in Australia and New Zealand. Eur J Heart Fail 2021 Jan;23(1):31-40. [doi: 10.1002/ejhf.2030] [Medline: 33094886]
11. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation 2012 Jul 24;126(4):501-506 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.125435] [Medline: 22825412]
12. Chan Y, Tuttle C, Ball J, Teng TK, Ahamed Y, Carrington MJ, et al. Current and projected burden of heart failure in the Australian adult population: a substantive but still ill-defined major health issue. BMC Health Serv Res 2016 Sep 21;16(1):501 [FREE Full text] [doi: 10.1186/s12913-016-1748-0] [Medline: 27654659]
13. Johnston J, Longman J, Ewald D, King J, Das S, Passey M. Study of potentially preventable hospitalisations (PPH) for chronic conditions: what proportion are preventable and what factors are associated with preventable PPH? BMJ Open 2020 Nov 09;10(11):e038415 [FREE Full text] [doi: 10.1136/bmjopen-2020-038415] [Medline: 33168551]
14. Prescott E, Piepoli M, Stepinska J, Fitzsimons D, Kerins M. ESC Prevention of cardiovascular disease programme. European Association of Preventive Cardiology. 2015. URL: https://www.escardio.org/static-file/Escardio/Subspecialty/EAPC/Documents/EAPC-SPMI%20Summary%20Report-HD.pdf [accessed 2020-10-11]
15. Ritchey MD, Maresh S, McNeely J, Shaffer T, Jackson SL, Keteyian SJ, et al. Tracking cardiac rehabilitation participation and completion among medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020 Jan;13(1):e005902. [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
16. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e16695 [FREE Full text] [doi: 10.2196/16695] [Medline: 32628615]
17. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1-452. [Medline: 8594834]
18. Six-minute walk test. Lung Foundation of Australia. 2016. URL: https://pulmonaryrehab.com.au/patient-assessment/assessing-exercise-capacity/six-minute-walk-test/ [accessed 2021-09-11]
19. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [doi: 10.1097/00005650-198601000-00007] [Medline: 3945130]
20. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
21. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Managing my heart health: for people with, or at high risk of, coronary heart disease. National Heart Foundation of Australia, Melbourne, Australia. 2012. URL: https://www.yumpu.com/en/document/view/21769999/managing-my-heart-health-at-a-glance-record-card-heart-foundation [accessed 2022-01-27]
24. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
25. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure. Circulation 2011 May 03;123(17):1873-1880. [doi: 10.1161/circulationaha.111.018473]
26. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Dec 22;392(10152):1047-1057 [FREE Full text] [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
27. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
28. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
29. Vuorinen A, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
30. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340. [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
31. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
32. Rosario MB, Lovell NH, Fildes J, Holgate K, Yu J, Ferry C, et al. Evaluation of an mHealth-based adjunct to outpatient cardiac rehabilitation. IEEE J Biomed Health Inform 2018 Nov;22(6):1938-1948. [doi: 10.1109/jbhi.2017.2782209]
33. Burn E, Nghiem S, Jan S, Redfern J, Rodgers A, Thiagalingam A, et al. Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. Heart 2017 Dec;103(12):893-894. [doi: 10.1136/heartjnl-2016-310195] [Medline: 28235776]
34. Bernabe-Ortiz A, Pauschardt J, Diez-Canseco F, Miranda JJ. Sustainability of mHealth effects on cardiometabolic risk factors: five-year results of a randomized clinical trial. J Med Internet Res 2020 Apr 21;22(4):e14595 [FREE Full text] [doi: 10.2196/14595] [Medline: 32314970]


## Abbreviations

**ACS:** acute coronary syndrome
**BP:** blood pressure
**CR:** cardiac rehabilitation
**ED:** emergency department
**GP:** general practitioner
**HF:** heart failure
**HR:** hazard ratio
**MGL:** Morisky–Green–Levine 4-item medication compliance
**mHealth:** mobile health
**RCT:** randomized controlled trial
**TCC:** TeleClinical Care
**TIM-HF:** Telemedical Interventional Monitoring in Heart Failure

*Edited by L Buis; submitted 02.08.21; peer-reviewed by B Dong, N Freene, B Dinesen; comments to author 10.09.21; revised version received 13.09.21; accepted 09.12.21; published 28.02.22.*

*Please cite as:*
*Indraratna P, Biswas U, McVeigh J, Mamo A, Magdy J, Vickers D, Watkins E, Ziegl A, Liu H, Cholerton N, Li J, Holgate K, Fildes J, Gallagher R, Ferry C, Jan S, Briggs N, Schreier G, Redmond SJ, Loh E, Yu J, Lovell NH, Ooi SY*
*A Smartphone-Based Model of Care to Support Patients With Cardiac Disease Transitioning From Hospital to the Community (TeleClinical Care): Pilot Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2022;10(2):e32554*
*URL: https://mhealth.jmir.org/2022/2/e32554*
*doi:10.2196/32554*
*PMID:35225819*

Original Paper

# Effects of a Mindfulness App on Employee Stress in an Australian Public Sector Workforce: Randomized Controlled Trial

Larissa Bartlett[1,2], PhD; Angela J Martin[2], PhD; Michelle Kilpatrick[2], PhD; Petr Otahal[2], BSc; Kristy Sanderson[2,3], PhD; Amanda L Neil[2], PhD

[1]Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Australia

[2]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

[3]School of Health Sciences, University of East Anglia, Norwich, United Kingdom

**Corresponding Author:**
Larissa Bartlett, PhD
Wicking Dementia Research and Education Centre
University of Tasmania
19 Liverpool Street
Hobart, 7000
Australia
Phone: 61 409791606
Email: larissa.bartlett@utas.edu.au

## Abstract

**Background:** Workplace-based mindfulness programs have good evidence for improving employee stress and mental health outcomes, but less is known about their effects on productivity and citizenship behaviors. Most of the available evidence is derived from studies of mindfulness programs that use class-based approaches. Mindfulness apps can increase access to training, but whether self-directed app use is sufficient to realize benefits equivalent to class-based mindfulness programs is unknown.

**Objective:** We assessed the effectiveness of a mindfulness app, both with and without supporting classes, for reducing employees' perceived stress. Changes in mindfulness, mental health, quality of life, perceptions of job demand, control and support, productivity indicators, organizational citizenship, and mindful behaviors at work were also investigated.

**Methods:** Tasmanian State Service employees were invited by the Tasmanian Training Consortium to a 3-arm randomized controlled trial investigating the effects of a mindfulness app on stress. The app used in the Smiling Mind Workplace Program formed the basis of the intervention. The app includes lessons, activities, and guided meditations, and is supported by 4 instructional emails delivered over 8 weeks. Engagement with the app for 10-20 minutes, 5 days a week, was recommended. Reported data were collected at baseline (time point 0), 3 months from baseline (time point 1 [T1]), and at 6-month follow-up (time point 2). At time point 0, participants could nominate a work-based observer to answer surveys about participants' behaviors. Eligible participants (n=211) were randomly assigned to self-guided app use plus four 1-hour classes (app+classes: 70/211, 33.2%), self-guided app use (app-only: 71/211, 33.6%), or waitlist control (WLC; 70/211, 33.2%). Linear mixed effects models were used to assess changes in the active groups compared with the WLC at T1 and for a head-to-head comparison of the app+classes and app-only groups at follow-up.

**Results:** App use time was considerably lower than recommended (app+classes: 120/343 minutes; app-only: 45/343 minutes). Compared with the WLC at T1, no significant change in perceived stress was observed in either active group. However, the app+classes group reported lower psychological distress (β=−1.77, SE 0.75; *P*=.02; Cohen *d*=–0.21) and higher mindfulness (β=.31, SE 0.12; *P*=.01; Cohen *d*=0.19). These effects were retained in the app+classes group at 6 months. No significant changes were observed for the app-only group or for other outcomes. There were no significant changes in observer measures at T1, but by time point 2, the app+classes participants were more noticeably mindful and altruistic at work than app-only participants.

**Conclusions:** Including classes in the training protocol appears to have motivated engagement and led to benefits, whereas self-guided app use did not realize any significant results. Effect sizes were smaller and less consistent than meta-estimates for class-based mindfulness training.

**Trial Registration:** Australian New Zealand Clinical Trials Register ACTRN12617001386325; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=372942&isReview

*(JMIR Mhealth Uhealth 2022;10(2):e30272)* doi:10.2196/30272

**KEYWORDS**

mindfulness; stress; apps; smartphone app; employee; workplace; performance; mobile phone

# Introduction

## Workplace Mindfulness Training

There is growing evidence in support of workplace-based mindfulness programs for increasing employee mindfulness, reducing stress, and improving mental health and well-being [1,2]. In the workplace literature, mindfulness correlates positively with psychological capital, organizational citizenship, and perceived job control and inversely with perceived job demands [3-5]. Accordingly, it is theorized that increasing employee mindfulness through training may help protect against stress, poor mental health, and work-based psychosocial risks. However, few randomized controlled trials (RCTs) have examined the intervention effects of workplace-based mindfulness programs on psychosocial risk factors or organizational outcomes such as employee productivity or performance [1].

Unmanaged stress is known to lead to psychological distress, depression, and anxiety [6,7], which are well-evidenced contributors to lost productivity via higher levels of employee absenteeism and presenteeism [8]. In Australia, the combined annual cost of absenteeism and presenteeism attributable to poor mental health is >US $11 billion, representing a significant economic burden [9]. Furthermore, the consequences of chronic stress include inattentiveness and antisocial or aggressive behavior that can be detrimental to work-based relationships and performance [10].

The occupational health psychology and workplace management literature points to the importance of considering factors that affect employee stress at both the organizational and individual levels [11]. A combined focus on minimizing work-related risk factors for mental health problems, promoting positive aspects of work and fostering employee strengths, and providing tertiary support to address presenting problems is considered best practice [12]. Although redressing adverse working conditions and improving management practices are vital components of workplace well-being strategies, supporting staff to access and develop personal coping strategies is also an important aspect of a healthy work environment [13]. Mindfulness training can provide personal support for employees as it actively cultivates adaptive coping skills that can buffer the effects of stress on employee health and well-being [14,15]. It may also help redress the organizational burden of health-related lost productive time (LPT) by improving mental health [16].

Mindfulness meditation involves the sustained practice of intentionally applying nonjudgmental attention to the current experience. There is some evidence that this practice improves attentional capacities [17], prosocial acting [18], and qualities that influence interpersonal relationships, such as gratitude and forgiveness [19]. Aggression has also been shown to reduce by following mindfulness training [20]. Amassing evidence suggests that increasing mindfulness through training can improve workplace performance, relationships, and well-being [21,22].

## Mindfulness Apps

Smartphone apps are an increasingly popular and accessible mode of delivery for mindfulness training and practice [23]. App functionality enables high-quality multimedia delivery of learning content that can be entirely preprogrammed to maximize intervention integrity and support self-guided learning [24]. For behavioral research, apps also have the ability to record engagement and use data. These data offer a more accurate measure of program engagement than participant recall, which is often used in mindfulness studies [25].

According to a review of 23 mindfulness apps against the Mobile App Rating Scale [26], the top 4 were Headspace, Smiling Mind, iMindfulness, and Mindfulness Daily [23]. The review by Mani [23] noted an absence of RCT evidence for the efficacy of mindfulness apps. Several trials of mindfulness apps have since been published, reporting results for stress, anxiety, depression, and well-being [27-32]. Only one of these RCTs was conducted in a workforce sample [27] in which self-guided use of the Headspace app gave rise to significant small- to moderate-sized effects for well-being, anxiety, depression, and psychosocial risk factors (job control and social support). Thus, this study supports the potential of an app-based workplace-based mindfulness program to positively influence job-related and affect-related variables associated with employee stress [33,34]. However, the effects of app-based workplace-based mindfulness programs have not yet been assessed for changing employee stress appraisals; chronic stress symptomology; or organizational performance outcomes such as productivity, citizenship behaviors, and social interactions [21,22].

## Study Aims

This study examines the efficacy of an app-based, low-dose workplace-based mindfulness program in a large, geographically and occupationally diverse Australian public service workforce. The trial followed an earlier pilot RCT of a 5-week Mindfulness at Work Program within the same workforce [35]. The Mindfulness at Work Program involved five 90-minute in-person classes and prescribed 20 minutes of daily meditation practice. Results of the pilot showed strong effects on stress reduction, mental health, and well-being but no significant improvements in health-related productivity. In-person class attendance at work time was found to be unfeasible for a high proportion of employees due to scheduling and geographical barriers. This study was conceived to examine whether low-dose mindfulness training using a mindfulness app could overcome accessibility challenges and realize beneficial outcomes for employee stress observed in face-to-face programs. The app that underpins the Smiling Mind Workplace Program [36] was selected, as it is already established in the Australian market and ranks highly against the Mobile App Rating Scale criteria [23].

The primary aim of this study is to assess the efficacy of the Smiling Mind Workplace Program app, offered both with and without supporting classes, in reducing employee stress (aim 1). We hypothesize that employees using the Smiling Mind Workplace Program app in conjunction with a series of four 1-hour classes (app+classes group) or using the Smiling Mind Workplace Program app self-guided without supporting classes (app-only group) would each report a consistent moderate-sized reduction in perceived stress when compared with a waitlist control (WLC) group.

The secondary aims are to explore the effects of this low-dose mindfulness intervention on psychological distress, mindfulness, health-related quality of life, perceived job demands, control, and resources (aim 2); explore changes in health-related LPT (aim 3); and explore observer-reported changes in participants' organizational citizenship and mindful behaviors (aim 4). The effect retention was also investigated (aim 5).

## Methods

### Overview

A 3-arm, open-label, parallel-group RCT was conducted between February 2018 and April 2019. The study was approved by the University of Tasmania health and medical human research ethics committee (H0016587) and registered with the Australian and New Zealand Clinical Trials Register in February 2018 (12617001386325). Baseline data were collected using web-based surveys administered in February 2018 (time point 0 [T0]). Postintervention surveys were conducted 3 months from baseline in May 2018 (time point 1 [T1]), with a 6-month follow-up in July 2018 (time point 2 [T2]). App use data were obtained at T1 and T2. The active intervention groups completed their training between T0 and T1. The control group was invited to access the intervention between T1 and T2. A further data collection wave was conducted 14 months from baseline (time point 3); however, analyses were not conducted because of high (85%) attrition (data not reported).

### Participants

#### *Overview*

The study sample was drawn from the Tasmanian State Service (TSS). The TSS employs approximately 18,000 people from 18 service agencies and centers across the island state of Tasmania, Australia. TSS employees work in a wide variety of roles (eg, frontline service and professional, administration, information, and asset management and maintenance). An invitation was widely disseminated via email and staff newsletters to express interest in joining a study of app-based mindfulness training for employee stress protection (Multimedia Appendix 1 [37]). The Tasmanian Training Consortium (TTC), which provides TSS staff development and training services, coordinated the invitation dissemination and collated the responses.

Participants needed to have access to a smartphone of any brand for personal use, permission from their supervisor to attend four 1-hour seminars in person or via videoconferencing, and make a commitment to complete the surveys. Eligibility was assessed after baseline based on no concurrent mindfulness or stress-management program of any type, including the use of other mindfulness apps, and not having unmanaged depression or other mental health conditions that might be exacerbated with unsupervised meditation. Mental health eligibility was assessed using baseline survey data from the Patient Health Questionnaire-9 (PHQ-9; [38]) and 2 questions about current and past mental health diagnoses. If respondents indicated a current or previously diagnosed mental health condition or their PHQ-9 score exceeded 15, indicating moderate-to-severe depression symptoms, their study eligibility was subject to review by a registered psychologist.

In the baseline surveys (T0), respondents were asked if they wished to nominate a work-based observer to join the study to answer some questions about the participants' behaviors at work. If *yes* was selected, the first name and email address of the nominee were entered, and the observer was invited to complete brief surveys about their observations of their paired participant's behaviors at each of the study time points.

#### *Randomization, Blinding, and Consent*

An independent statistician (PO) randomized eligible participants into the 3 groups, stratified by whether they had an observer. Group allocations were sent to the TTC, who notified the participants of their training schedule and coordinated the seminars. It was not feasible to blind the TTC staff, study participants, or teacher to treatment [39]. All data were collected via the web using surveys administered using REDCap (Research Electronic Data Capture; Vanderbilt University) [40]. The CHERRIES (Checklist for Reporting Results of Internet e-Surveys) [37] study is included in Multimedia Appendix 1. Research personnel only interacted with randomized participants by email to administer the web-based surveys, and all analyses were conducted on deidentified data. Consent to participate in the research was given at the commencement of each survey, and no incentives were provided. The CONSORT (Consolidated Standards of Reporting Trials) checklist is included in Multimedia Appendix 1.

#### *Interventions*

Released to the market in 2014, the Smiling Mind Workplace Program aims to enable working adults to develop mindfulness skills and embed mindfulness practices into daily life. The established low-dose mindfulness program involves a series of 5 learning modules delivered in 4 interactive 1-hour face-to-face workshops. These are led by a Smiling Mind facilitator over 8 weeks and supported by the use of the Smiling Mind Workplace Program app. This app comprises 41 elements, including videos and audio lessons, guided meditations, and practical activities such as moving with awareness between meetings, breathing techniques, and listening exercises to help cultivate workplace mindfulness. Use of the app-based activities and meditations is supported by fortnightly emails relating to the content covered in the workshops and app-based lessons. The recommended minimum engagement with the Smiling Mind Workplace Program app is 10 to 20 minutes' mindfulness practice each weekday. Smiling Mind Workplace Program history and content are provided in Multimedia Appendix 1.

To maximize accessibility, Smiling Mind Workplace Program workshops were delivered in a seminar format in university venues located in the north, northwest, and south of the state. Classes ran twice, in the morning and afternoon, on the advertised dates. Participants were able to attend in person or via videoconferencing, and catch-up recordings were made available. All classes were led by the same mindfulness teacher with certification from the University of Massachusetts Center for Mindfulness and >10 years of teaching experience. No supplementary messaging, incentives, or other forms of contact from the study team were used to encourage intervention engagement.

The app+classes group participants were invited to download and use the Smiling Mind Workplace Program app and attend four 1-hour classes scheduled fortnightly during work time. These participants were sent fortnightly generic emails from the Smiling Mind team to support the use of the app-based materials.

The app-only group participants were invited to download and use the Smiling Mind Workplace Program app and received fortnightly emails but were not invited to attend the classes.

The WLC group participants received no information during T0 to T1. After data collection for T1 was complete, the WLC group was invited to a single 2-hour seminar and to download and use the Smiling Mind Workplace Program app self-guided, in conjunction with the fortnightly emails.

All groups retained access to the Smiling Mind Workplace Program app for 12 months.

## Measures

Demographic variables (age, sex, marital status, educational attainment, work role, and schedule) were collected from participants at T0, as were past or planned exposure to other mindfulness or stress management training and self-ratings of readiness for change (percent).

The 10-item Perceived Stress Scale (PSS; [41]) was used to assess the primary outcome at all time points. Response options were summed (range 0-40), with higher scores indicating higher perceived stress. The baseline PSS data showed good internal consistency (Cronbach α=.92).

The PHQ-9 [38] was used for eligibility screening. Established clinical cutoff points were followed for mild (5), moderate (10), moderately severe (15), and severe (20) depression. The baseline data indicated good internal consistency (Cronbach α=.86).

The Kessler 10-item measure [42] was used to assess psychological distress at all time points. Cutoff points from Australian norms signify a severe risk of a clinical mental health condition for people who score >30, high risk for people who score between 22 and 29, moderate risk for people who score between 16 and 21, and low risk for people who score <15 [43]. The baseline data indicated good internal consistency (Cronbach α=.91).

The 15-item Mindful Attention and Awareness Scale [44] was used to measure the mindfulness of respondents at all time points. Mean responses across the 15 items were computed, with higher mean scores (range 1-6) indicating higher trait mindfulness. Internal consistency was good at baseline (Cronbach α=.91).

The 35-item, 8-dimension Assessment of Quality of Life (AQoL) measure [45], which assesses quality of life related to physical health (independent living, pain, and senses) and psychosocial health (mental health, happiness, coping, relationships, and self-worth), was used at all time points. Scores were computed using the 8-dimension AQoL algorithm (range 0.09-1.00). A score of 0.00 equates to death, and 1.00 equates to full health.

Perceptions of job demand, control, and support were used to assess work-related psychosocial risk at all time points. Demand and control were assessed using 7 items drawn from the Household, Income, and Labour Dynamics in Australia survey [46]. Scores were summed for 4 demand items (range 0-24) and 3 control items (range 0-18). A higher risk of job-related stress is indicated when demand scores are higher, and control scores are lower. Job support was assessed using summed responses to 6 items drawn from the Swedish Demand, Control, and Support Survey [47]. Higher scores (range 4-24) indicated a lower psychosocial risk of job stress. Internal inconsistency was weaker for the demand scale (Cronbach α=.65) than for the control (Cronbach α=.80) and support (Cronbach α=.80) measures.

Effects on productivity were based on estimates of health-related LPT [48]. Participants were asked to think about their work attendance in the previous 4 weeks and report the number of days they stayed away from work because of ill health (absentee days) and the number of days they went to work but were unwell (presenteeism days). Absentee days were considered 100% lost (eg, 2 absentee days=2 lost days). If presenteeism days were reported, an estimate of productivity (percentage) on those days was recorded. The number of lost productive days was assessed as the product of the number of presenteeism days and lost productivity on those days. For example, 3 presenteeism days at 60% productivity were calculated as follows:

$$(3 \times [100-60]) = 1.2 \text{ lost days} \quad \textbf{(1)}$$

The total number of days lost through absenteeism and presenteeism is thus reported as health-related LPT.

The degree to which changes in participants' mindful behaviors (eg, attentiveness, awareness, and acceptance) were noticeable to work colleagues was assessed at all time points using a 9-item observed mindfulness measure (OMM; [49]). This instrument includes items such as "The person has difficulty staying focused on what is happening to/around them as it occurs (Attentiveness)," "When asked how he or she is feeling, the person can identify their emotions easily (Awareness)," and "The person seems to recover well from unpleasant or stressful experiences (Acceptance)." Response options indicated the frequency of observed behaviors (1=not at all and 5=all the time). Scores for 3 items (items 1, 4, and 7) were reversed before summing to obtain subscale scores for observed mindful acceptance, awareness, and attentiveness and the total score. The internal consistency of OMM data at baseline was good (Cronbach α=.88).

A 16-item Organizational Citizenship Behaviors observer report instrument [50] was used at all time points to assess noticeable participant behaviors at work. Response options indicated the frequency of observed behaviors, and higher summed scores indicated higher degrees of altruism (range 5-30) and compliant behaviors (range 4-20). Cronbach test showed some internal inconsistency at baseline (altruism Cronbach α=.72 and compliance Cronbach α=.62).

Intervention adherence was assessed using self-reported seminar attendance and app use data from the Smiling Mind Workplace Program server. Whether the participants downloaded and engaged with the app (yes or no) was recorded. Engagement was calculated as the proportion of time spent in the Smiling Mind Workplace Program app activities out of a potential maximum of 343 minutes for the entire program. Participants' perceptions of the acceptability of the intervention were assessed using qualitative data from 2 open questions in the T1 survey. Observers provided free-text responses at the end of each survey about their experience in the study and to share any additional information about their paired participants.

### Statistical Analysis

The required sample size was calculated using a pooled PSS estimate from a meta-analysis of 13 RCTs of workplace-based mindfulness programs (Cohen *d*=−0.54; mean difference −4.21, SE 0.14) [1]. A minimum of 198 participants was required to achieve a power of 0.8 and α=.025 (maintaining a family-wise error rate of 0.05) [51]. The recruitment target (n=261) allowed for 25% attrition.

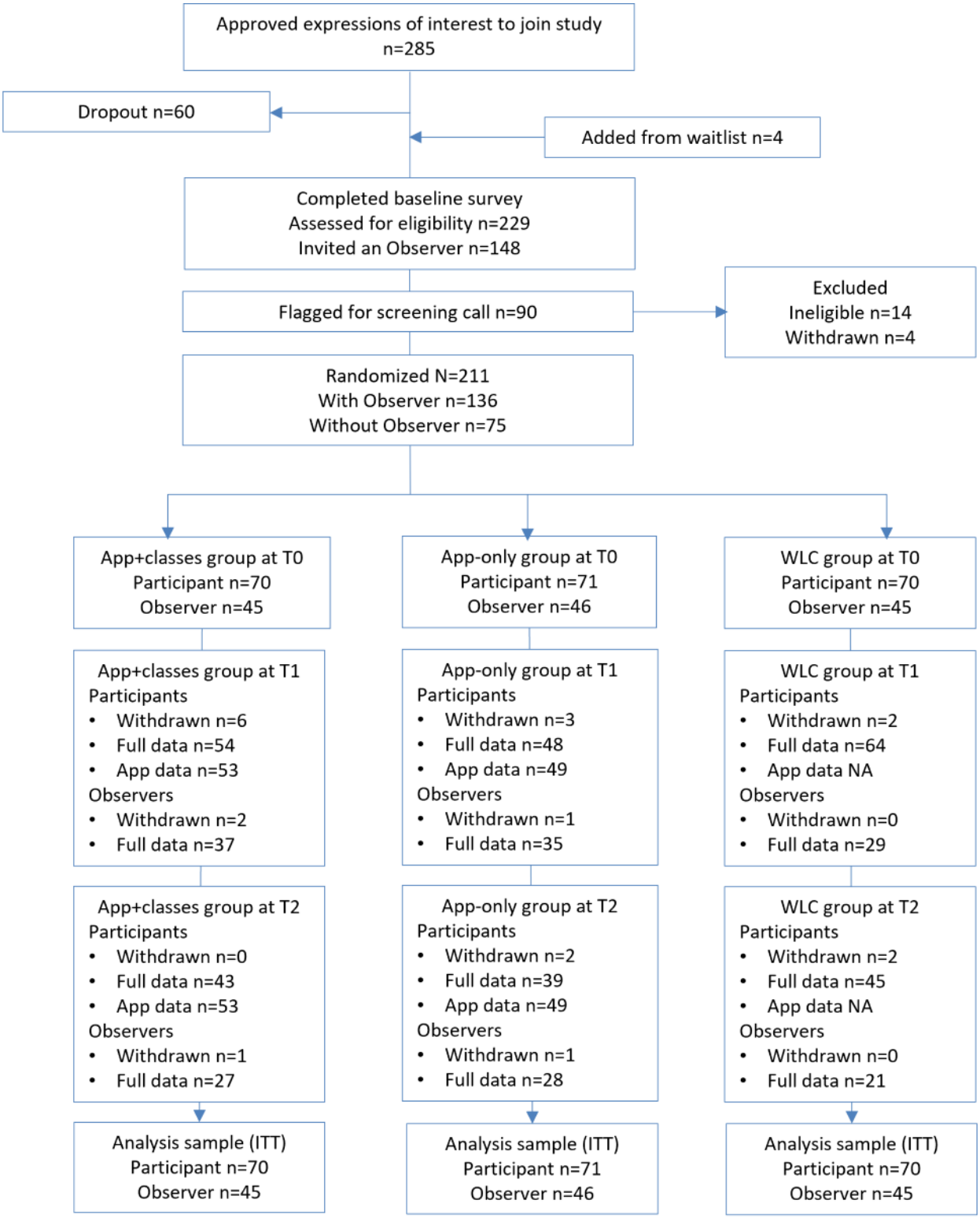
Intention-to-treat analyses were conducted using an original assigned group approach [52]. Significance tests (α=.05) were adjusted using the Tukey method for multiple comparisons when >2 groups were included in the model. Analyses were conducted in the R (version 3.4.3; R Foundation for Statistical Computing) [53] using the psych [54], lme4 [55], and lmerTest packages [56]. Repeated measures linear mixed models were used to assess changes in the app+classes and app-only groups compared with the WLC group from T0 to T1, with age, sex, prior mindfulness training, and main occupation included to inform missing data computations. Two-group comparisons were used to test the difference in effect retention between the app+classes and app-only groups beyond T1. Cohen *d* standardized mean difference effect estimates were computed using the Lakens [57] guidelines (0.2=weak, 0.5=moderate, and 0.8=strong). Agreement between participants and their observers was assessed using intraclass correlation coefficient (ICC) estimates in 2-way random effects models following the Koo and Li [58] guidelines (0.5=poor, 0.5-0.75=moderate, 0.75-0.9=good, and >0.9=excellent agreement). Spearman correlations were used to test the relationship between program adherence and study outcomes. Chi-square and Fisher exact tests were used to explore the differences in intervention engagement and health-related LPT. Qualitative data were read twice by 2 authors (AJM and LB), with frequent themes identified, coded, and assessed using a content analysis approach [59].

## Results

### Participant Enrollment and Attrition

The flow of participants and observers is illustrated in Figure 1. Of an approved pool of 285 TSS employees, baseline measures were completed by 229 (80.4%) employees. Of the 229 respondents, 90 (39.3%) were invited to a screening interview by the study psychologist, of whom 14 (16%) were deemed clinically ineligible, an additional 4 (4%) withdrew, and 2 (2%) were excluded because of nonresponse. The starting sample of 211 individuals included 136 (64.5%) participants with paired observers. Group assignments were app+classes (participants 70/211, 33.2%; observers 45/136, 33.1%), app-only (participants 71/211, 33.6%; observers 46/136, 33.8%), and WLC (participants 70/211, 33.2%; observers 45/136, 33.1%). Statistical power for the hypothesized moderate-sized PSS effect was achieved in the starting sample.

**Figure 1.** Participant flow diagram. Ineligible: did not meet inclusion criteria; Withdrawn: requested no further surveys, available data not withdrawn from analyses; ITT: intention to treat; T0: time point 0; T1: time point 1; T2: time point 2; WLC: waitlist control.




Of the 211 participants and 136 observers, 15 (7.1%) participants and 6 (4.4%) observers advised withdrawal during the study period. The participants' reasons for withdrawal were time pressures (4/15, 27%), changing job (4/15, 27%), difficulty accessing the app-based materials (1/15, 7%), extended leave (3/15, 20%), and no reason (3/15, 20%). Observers' reasons included no longer being in contact with their paired participant (3/6, 50%) or their participant had withdrawn (3/6, 50%). Of the 211 participants, complete survey data were provided by 167 (79.1%) participants at T1 and 129 (61.1%) participants at T2.

## Participant Characteristics

Participant characteristics were similar across the intervention groups (Table 1), except for full-time workers. Just under half

of the sample reported some prior exposure to mindfulness, and readiness to commence training was >80% across groups.

**Table 1.** Participant characteristics (n=211).

| Characteristics variables | WLC[a] (n=70) | App (n=71) | App+classes (n=70) | Difference (*P* value)[b] |
|---|---|---|---|---|
| **Age category (years), n (%)** | | | | |
| 18 to 34 | 7 (10) | 9 (13) | 9 (13) | .60 |
| 35 to 44 | 18 (26) | 20 (28) | 23 (33) | .60 |
| 45 to 55 | 20 (29) | 22 (31) | 24 (34) | .60 |
| 55 to 64 | 23 (33) | 17 (24) | 14 (20) | .60 |
| >65 | 2 (3) | 3 (4) | 0 (0) | .60 |
| Gender (female), n (%) | 53 (76) | 50 (70) | 50 (71) | .76 |
| **Educational attainment, n (%)** | | | | |
| High school | 2 (3) | 6 (9) | 6 (9) | .37 |
| College | 24 (34) | 16 (23) | 19 (27) | .37 |
| University | 44 (63) | 49 (69) | 45 (64) | .37 |
| Living as married, n (%) | 55 (79) | 56 (79) | 52 (74) | .77 |
| Prior mindfulness training, n (%) | 34 (49) | 35 (49) | 31 (44) | .81 |
| **Main occupation, n (%)** | | | | |
| Blue collar | 1 (1) | 1 (1) | 1 (1) | .21 |
| Clerical or admin | 15 (21) | 5 (7) | 12 (17) | .21 |
| Technical or services | 4 (6) | 9 (13) | 10 (14) | .21 |
| Professional | 38 (54) | 48 (68) | 35 (50) | .21 |
| Senior manager | 12 (17) | 8 (11) | 12 (17) | .21 |
| Works full time, n (%) | 49 (70) | 61 (86) | 56 (80) | .07 |
| **Work schedule, n (%)** | | | | |
| Regular daytime | 64 (91) | 61 (86) | 62 (89) | .85 |
| Regular evening or night | 2 (3) | 2 (3) | 2 (3) | .85 |
| Irregular or rotating | 4 (6) | 8 (11) | 6 (9) | .85 |
| Percentage readiness for training, mean (SD) | 86 (16) | 85 (18) | 82 (21) | .45 |

[a]WLC: waitlist control.

[b]Difference between-group *P* values computed using analysis of variance for continuous variables and chi-square tests of group equivalence for categorical variables.

## Aim 1: Intervention Effects for Perceived Stress

Postintervention RCT effect estimates are presented in Table 2. Although there was a downward trend in perceived stress, when compared with the WLC, there was no significant change for either the app+classes or app-only group. Prior exposure to mindfulness, readiness to commence training, or depression severity at baseline were not significant moderators.

**Table 2.** Postintervention randomized controlled trial effect estimates.

| Outcome variables | Time point 0, mean (SE)[a] | Time point 1, mean (SE) | Effect estimates | | |
|---|---|---|---|---|---|
| | | | β[a] (SE) | *P* value[a,b] | Cohen *d*[c] (95% CI) |
| **Perceived stress**[d] | | | | | |
| WLC[e,f] | 16.37 (0.75) | 15.32 (0.77) | —[g] | — | — |
| App-only[h] | 17.40 (0.74) | 14.91 (0.84) | −1.44 (1.01) | .16 | −0.06 (−0.39 to 0.27) |
| App+classes[i] | 17.15 (0.75) | 15.38 (0.81) | −0.73 (0.98) | .46 | 0.01 (−0.32 to 0.34) |
| **Mindfulness**[j] | | | | | |
| WLC | 3.83 (0.09) | 3.65 (0.10) | — | — | — |
| App-only | 3.83 (0.09) | 3.79 (0.10) | .15 (0.12) | .23 | 0.17 (−0.16 to 0.50) |
| App+classes | 3.69 (0.09) | 3.81 (0.10) | .31 (0.12) | .01 | 0.19 (−0.14 to 0.52) |
| **Psychological distress**[k] | | | | | |
| WLC | 18.68 (0.67) | 19.46 (0.68) | — | — | — |
| App-only | 19.08 (0.66) | 18.65 (0.73) | −1.21 (0.78) | .12 | −0.14 (−0.47 to 0.19) |
| App+classes | 19.21 (0.66) | 18.22 (0.71) | −1.77 (0.75) | .02 | −0.21 (−0.55 to 0.12) |
| **Job demands** | | | | | |
| WLC | 16.41 (0.43) | 15.64 (0.45) | — | — | — |
| App-only | 16.79 (0.43) | 15.90 (0.49) | −.13 (0.59) | .83 | 0.07 (−0.26 to 0.40) |
| App+classes | 16.93 (0.43) | 15.69 (0.47) | −.47 (0.57) | .41 | 0.01 (−0.32 to 0.34) |
| **Job control** | | | | | |
| WLC | 10.11 (0.47) | 10.45 (0.48) | — | — | — |
| App-only | 10.67 (0.47) | 11.25 (0.52) | .25 (0.55) | .65 | 0.19 (−0.14 to 0.52) |
| App+classes | 10.60 (0.47) | 11.03 (0.50) | .10 (0.53) | .86 | 0.14 (−0.19 to 0.47) |
| **Job support** | | | | | |
| WLC | 18.43 (0.39) | 18.40 (0.40) | — | — | — |
| App-only | 17.85 (0.39) | 18.70 (0.44) | .88 (0.50) | .08 | 0.09 (−0.24 to 0.42) |
| App+classes | 18.03 (0.39) | 18.08 (0.42) | .08 (0.48) | .87 | −0.09 (−0.42 to 0.24) |
| **QoL**[l,m]**: physical health** | | | | | |
| WLC | 0.75 (0.02) | 0.75 (0.02) | — | — | — |
| App-only | 0.76 (0.02) | 0.77 (0.02) | .00 (0.02) | .83 | 0.12 (−0.21 to 0.45) |
| App+classes | 0.75 (0.02) | 0.76 (0.02) | .01 (0.02) | .74 | 0.06 (−0.27 to 0.39) |
| **QoL: mental health** | | | | | |
| WLC | 0.37 (0.02) | 0.39 (0.02) | — | — | — |
| App-only | 0.37 (0.02) | 0.43 (0.02) | .03 (0.02) | .13 | 0.24 (−0.09 to 0.57) |
| App+classes | 0.35 (0.02) | 0.40 (0.02) | .02 (0.02) | .26 | 0.06 (−0.27 to 0.39) |
| **QoL: utility score** | | | | | |
| WLC | 0.71 (−0.02) | 0.73 (0.02) | — | — | — |
| App-only | 0.72 (−0.02) | 0.76 (0.02) | .02 (0.02) | .28 | 0.18 (−0.15 to 0.51) |
| App+classes | 0.69 (−0.02) | 0.73 (0.02) | .02 (0.02) | .33 | 0.00 (−0.33 to 0.33) |

[a]Estimated marginal means and effect estimates from maximum likelihood linear mixed models with age, sex, education, and prior mindfulness exposure as auxiliary variables; all analyses were based on intention-to-treat principles with all cases analyzed in their original assigned group.

[b]Significant with α=.05.

[c]Standardized mean difference effect estimate computed using time point 1 estimated marginal means and SE.

[d]Perceived Stress Scale (10 items).

[e]WLC: waitlist control group.

[f]n=70.

[g]WLC ceased to be comparator after time point 1; hence, data are not shown.

[h]Self-guided app group (n=71).

[i]Self-guided app use plus supporting classes (n=70).

[j]Mindful Awareness and Attention Scale.

[k]Kessler-10 scale.

[l]QoL: quality of life.

[m]Assessment of Quality of Life (8 dimension).

Among the 70 participants in the app+classes group, class attendance diminished over time, with 45 (64%) attendees in the first class, 36 (51%) in the second, 33 (53%) in the third, and 32 (46%) in the fourth class. Table 3 shows that the Smiling Mind Workplace Program app was downloaded by 70% (49/70) of the participants in the app+classes group and 49% (35/71) of participants in the app-only group. The app+classes group also had higher median engagement with the learning and practice elements within the app (45/343 total activity minutes) and with the meditation practices over the 8-week period (73 meditation minutes) than those in the app-only group (45/343 total activity minutes, with 27 meditation minutes). Perceived stress change was significantly correlated with intervention engagement in the app+classes group ($r$=−0.33) but not in the app-only group. Investigation of T0:T1 change in PSS scores by meditation time and program engagement suggests an inverse linear dose–response pattern in the app+classes group. This pattern was not evident in the app-only group (Figure 2).

**Table 3.** Smiling Mind Workplace Program app engagement indices for the app+classes and app-only groups between time point 0 and time point 1[a].

| Engagement variables | App-only[b] (n=71) | App+classes[c] (n=70) | Test of difference (*P* value) |
|---|---|---|---|
| App downloads, n (%) | 35 (49) | 49 (70) | —[d] |
| **App use, median (IQR)** | | | |
| Number lessons completed | 2 (0-14) | 4 (0-16) | .01 |
| Number activities completed | 0 (0-4) | 1 (0-7) | .09 |
| Total meditation minutes | 27 (0-296) | 73 (0-476) | .03 |
| Number meditations completed | 4 (0-44) | 11 (0-55) | .03 |
| Percentage of possible total engagement[e] | 13% (0%-126%) | 35% (1%-160%) | .05 |

[a]Tests of difference used 2-tailed *t* test using complete case data for normally distributed variables and Kruskal-Wallis rank sum test for nonnormally distributed variables.

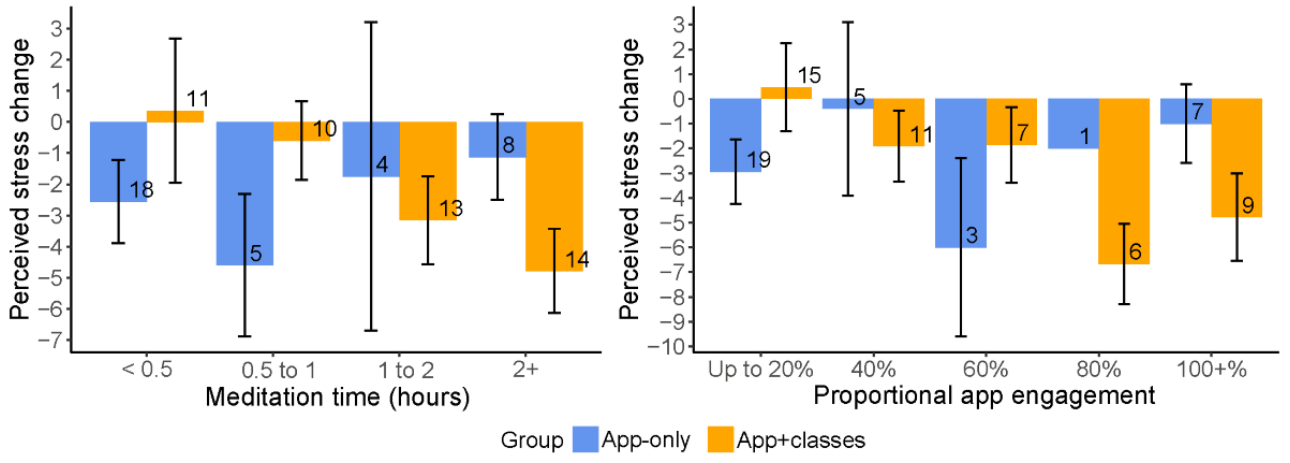
[b]Self-guided app use.

[c]Self-guided app use plus classes.

[d]Not conducted.

[e]Total time if all app-based activities were completed was 343 minutes.

**Figure 2.** Perceived stress change from baseline to after the intervention by meditation time and app engagement.



### Aim 2: Psychological Distress, Mindfulness, Work-Related Psychosocial Risks, and Quality of Life

The results (Table 2) show that compared with the WLC, the app+classes group reported small improvements in psychological distress (Cohen *d*=−0.21) and mindfulness (Cohen *d*=0.19). At T1, the Kessler-10 data showed that 15% (8/54) of respondents in the app+classes group transitioned into a lower category for risk of clinical mental health problems, whereas 2% (1/54) of participants shifted to a higher-risk category. No significant effects were found for either psychological distress or mindfulness in the app-only group, and an equal number reported beneficial (4/48, 8%) and detrimental changes in risk status (4/48, 8%). Of the 70 participants in the WLC, 14% (9/64) shifted to higher risk and 9% (6/64) to lower-risk categories during the initial intervention period.

No discernible trends in the quality of life data were evident for either the app+classes group or the app-only group when compared with the WLC group. Similarly, psychosocial risk factors did not change significantly in either active group at T1.

### Aim 3: Productivity and Workplace Incidents

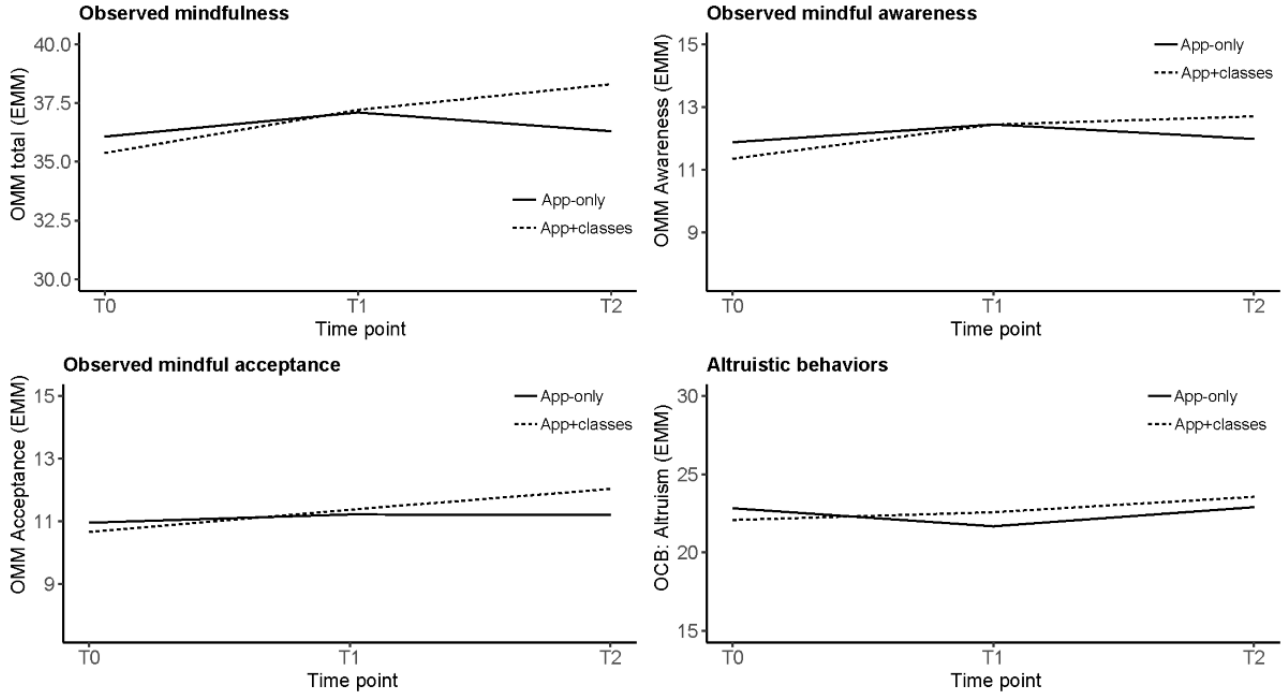
The raw productivity and workplace incident results are presented in Table S1 of Multimedia Appendix 1. Health-related LPT was categorized into four levels: no health-related LPT, up to 1 day, 1 to 3 days, and >3 days. The app+classes and app-only groups trended lower in health-related LPT than in the WLC group following training, but the difference was not significant. The number of app+classes participants who reported work success increased from 26% (18/70) at T0 to 39% (17/43) at T2. This change was stronger than that observed in the app-only (28/71, 39% to 17/39, 43%) and WLC (18/70, 26% to 13/45, 29%) groups. Work failures reduced from T0 to T2 for the active groups (app+classes: 6/70, 9% to 3/43, 7%; app-only: 10/71, 14% to 4/39, 10%), whereas failures increased in the same period for the WLC (4/70, 6% to 4/45, 9%). Workplace accidents were infrequent in all groups, with 1% (1/70) of participants in the app+classes group, 7% (5/71) in the app-only group, and 6% (4/70) in the WLC group endorsing this item at T0.

### Aim 4: Observer-Reported Mindfulness and Organizational Citizenship

Observer-reported outcomes are illustrated in Figure 3. The results are detailed in Table S2 of Multimedia Appendix 1. Changes in observer-reported mindful behaviors and self-reported mindfulness showed consistent agreement at each time point (T0: ICC=0.35, *P*=.01; T1: ICC=0.32, *P*=.03; T2: ICC=0.39, *P*=.03). At T1, observers reported a small but nonsignificant trend toward higher observed mindful behaviors in both active groups compared with the WLC. At the 6-month follow-up (T2), head-to-head comparison between the active groups showed that the app+classes participants displayed more noticeably mindful behaviors than the app-only participants (Cohen *d*=0.34, 95% CI −0.08 to 0.75).

The distribution of data in the organizational citizenship compliance subscale showed that responses were bounded at the top from baseline; thus, these data were excluded from the analyses. Although the results for altruism were not significant, plots (Figure 3) illustrate that the app+classes group trended higher on this measure at T1 and T2, whereas the app-only group initially trended toward lower altruism at T1, which was ameliorated at T2.

**Figure 3.** Change trends from baseline to 6 months: interactions between app-only and app+classes groups for observer-reported mindful and altruistic behaviours. Observed mindfulness measure (OMM) range 9 to 45; OMM awareness and acceptance range 3 to 15; and Organizational Citizenship Behavior Altruism subscale (OCB) range 5 to 30. EMM: estimated marginal mean; T0: time point 0; T1: time point 1; T2: time point 2.



### Aim 5: Effect Retention

Results comparing the app+classes and app-only groups at the 6-month follow-up (T2) are reported in Table 4. The effects observed for mindfulness and psychological distress developed further in both groups beyond intervention completion (T1) such that there was no significant difference between groups at T2. The app+classes group continued to trend lower than the app-only group in perceived job demands and higher in job control from T1 to T2; however, the social support results observed at T1 showed no further development at T2.

**Table 4.** Effect estimates for the app+classes group compared with the app-only group at 6-months follow-up for mindfulness, psychological distress, job demands, and job control.

| Outcome variable and group | T0[a], mean[b] (SE)[c] | T2[d], mean (SE) | Effect estimate T0:T2 | |
|---|---|---|---|---|
| | | | β[c] (SE) | *P* value[c] |
| **Mindfulness** | | | | |
| App-only | 3.82 (0.10) | 3.91 (0.11) | Reference | Reference |
| App+classes | 3.68 (0.10) | 3.94 (0.11) | .04 (0.16) | .82 |
| **Psychological distress** | | | | |
| App-only | 19.08 (0.70) | 18.21 (0.79) | Reference | Reference |
| App+classes | 19.16 (0.70) | 17.69 (0.78) | −.52 (1.11) | .64 |
| **Job demands** | | | | |
| App-only | 16.72 (0.44) | 16.46 (0.52) | Reference | Reference |
| App+classes | 16.90 (0.44) | 15.08 (0.51) | −1.38 (0.73) | .06 |
| **Job control** | | | | |
| App-only | 10.70 (0.45) | 10.65 (0.53) | Reference | Reference |
| App+classes | 10.64 (0.46) | 11.39 (0.52) | .73 (0.74) | .33 |

[a]T0: time point 1 (baseline).

[b]Estimated marginal means.

[c]β, SE, and *P* values from the 2-group comparison of effects in linear mixed models, with app-only group set as reference.

[d]T2: time point 2 (6-months from baseline).

### Intervention Acceptability

The frequency of themes derived from the qualitative data is reported in Table 5. Reports from the 2 active groups showed overall satisfaction with the mindfulness training. Responses to the free-text questions from the participants (57/141, 40.4%) indicated that they found the training useful, practical, helpful, or beneficial, more frequently among the app+classes (35/70, 50%) participants than app-only participants (22/71, 31%). Approximately 19% (13/70) of members of the app+classes group reported finding the program immediately beneficial, whereas this was volunteered by only 6% (4/71) of the app-only participants. The app was considered easy to use by 14.9% (21/141) of all participants. However, although 8.5% (12/141) of participants reported that they were incorporating the practice into daily life, 12.7% (18/141) of respondents found establishing a routine difficult, and 8.5% (12/141) of participants reported that it was not feasible to engage with the program while at work. Comments from 24% (17/70) of the app+classes group participants indicated that they found the seminars motivating. However, more app+classes group participants reported difficulties associated with time demands (5/70, 7%) and establishing a practice routine (12/70, 17%) than the app-only group participants (3/71, 4% and 6/71, 8%, respectively). A small number of participants reported technical problems with the app and seminars. One of the individuals in each group reported that they felt the research surveys were independently helpful in sensitizing them to their mental well-being. The in-app elements considered most useful by participants in both active groups were meditations, ranked highest by 57% (55/97) of respondents. Micropractices, which are brief mindful activities that can be used throughout the day, were rated very useful by 41% (40/97) of participants, in-app lessons by 32% (31/96) of participants, and body scan practices by 31% (30/97; data not shown).

**Table 5.** Frequency of themes derived from postintervention free-text responses regarding the usefulness of the program (N=141).

| Themes derived from qualitative data | All respondents, n (%) | App+classes group (n=70), n (%) | App-only group (n=71), n (%) |
|---|---|---|---|
| **Participant view of outcomes** | | | |
| Improved well-being | 7 (5) | 4 (6) | 3 (4) |
| Improved sleep | 4 (3) | 2 (3) | 2 (3) |
| Improved productivity | 3 (2) | 2 (3) | 1 (1) |
| Improved recovery | 2 (1) | 1 (1) | 1 (1) |
| Improved relationships | 1 (1) | 0 (0) | 1 (1) |
| **Acceptability** | | | |
| Useful, practical, helpful, and beneficial | 57 (40) | 35 (50) | 22 (31) |
| Immediate benefit and real-time application | 17 (12) | 13 (19) | 4 (6) |
| Variety, choices, and range of app elements | 11 (8) | 7 (10) | 4 (6) |
| Found app irritating and disruptive | 6 (4) | 2 (3) | 4 (6) |
| Would recommend | 4 (3) | 3 (4) | 1 (1) |
| **Feasibility** | | | |
| Easy to use, accessible, and flexible | 21 (15) | 9 (13) | 12 (17) |
| Establishing routine is difficult | 18 (13) | 12 (17) | 6 (8) |
| Seminars were motivating and beneficial | 17 (12) | 17 (24) | 0 (0) |
| Incorporating practices into daily life | 12 (9) | 6 (9) | 5 (7) |
| Not feasible at work | 12 (9) | 5 (7) | 7 (10) |
| Technical problems with app | 8 (6) | 5 (7) | 3 (4) |
| Time challenges or demands of training | 8 (6) | 5 (7) | 3 (4) |
| Self-guided program difficult | 7 (5) | 1 (1) | 6 (8) |
| Technical problems with seminars | 3 (2) | 3 (4) | 0 (0) |
| No benefit from seminar attendance | 3 (2) | 3 (4) | 0 (0) |
| **Contextual circumstances** | | | |
| Major life stresses during the study | 10 (7) | 5 (7) | 2 (3) |
| Life got in the way (did not do training) | 10 (7) | 8 (11) | 2 (3) |
| Did not use the app | 8 (6) | 0 (0) | 8 (11) |
| Surveys made difference on their own | 2 (1) | 1 (1) | 1 (1) |

### Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author (LB).

## Discussion

### Principal Findings

This RCT assessed the effects of participating in a low-dose, app-based workplace-based mindfulness program delivered both with and without supporting classes in a sample of public sector employees. The study hypothesis that using the Smiling Mind Workplace Program app, either self-guided or with supporting classes, would result in moderate-sized reductions in perceived stress was not supported. Although the app+classes group engaged more with the training, neither group achieved the recommended dose. Despite the low engagement, when compared with the inactive control group, the app+classes group reported significant increases in mindfulness and decreases in psychological distress. These benefits were retained at 6-month follow-up, at which point the app+classes group also reported significantly lower perceived job demands than the app-only group. No significant effects were observed for either intervention group for health-related quality of life or productivity. Although the Smiling Mind Workplace Program app was well-received by most participants in the active groups, those whose training protocol was entirely self-guided engaged less with training and reported no statistically significant changes in any of the study outcomes.

The null result for perceived stress was unexpected, given consistent positive findings from other workplace-based mindfulness programs [1] and the apparent efficacy of the current intervention for significant and lasting benefits for psychological distress. Although the 2 constructs are usually correlated, they are not the same. Perceived stress refers to the

perceived capacity to meet the demands of presenting stressors, whereas psychological distress refers to health risks associated with sustained or unrelieved stress [60]. It is plausible that participants in the app+classes group developed skills through their mindfulness training protocol to regulate their emotions, thereby attenuating distress, whereas their perception of the demands and frequency of stressors may have remained unchanged. The PSS results for all 3 groups, including the control, trended lower over the main intervention period (T0 to T1), which might suggest a sample-wide reduction in stressor exposure; however, this was not detected or reported in other data collected for this study.

The significant changes in mindfulness and distress were encouraging but lower than meta-analytic estimates from workplace-based mindfulness programs delivered via face-to-face classes or web-based learning platforms [61-63]. These findings support the likelihood of a dose–response relationship, where the degree of exposure to mindfulness training and practice is associated with the size of the effects [64]. Despite the lower effect sizes, the psychological distress scores at T1 suggest that the app+classes training protocol was sufficient to realize meaningful mental health risk reduction for 15% (8/54) of participants.

Higher engagement with the Smiling Mind Workplace Program app by app+classes participants appears to have been motivated by seminar attendance, a sentiment volunteered in free-text data by 24% (17/70) of app+classes participants. For example, one of the participants stated:

> *I was fortunate to be selected to attend sessions which I believe was VERY important. This helped tremendously with getting the motivation to work through the app sessions. Other colleagues from my work who were not selected to attend sessions have very low motivation and barely did any of the app sessions.*

The self-guided app-only group not only missed the class-based educational and discursive opportunities but also engaged less than the app+classes participants with the in-app educational videos, lessons, and practice resources. This poorer engagement may explain the pattern in PSS changes depicted in Figure 2, where the app+classes group reported a clearer and more consistent dose–response than the app-only group. It is feasible that in the absence of feedback and guidance by a teacher, or the opportunity to discuss experiences with other learners, the app-only participants were less able to apply mindful awareness and acceptance, as their experiences arise and pass away during meditation practices, and thus derived less benefit [65].

The absence of significant improvement in mindfulness or distress in the app-only group indicates that self-guided use of the Smiling Mind Workplace Program app was insufficient to realize consistent changes within the main intervention period (T0 to T1). This finding is in keeping with previous work that has shown that face-to-face classes in the training protocol are associated with stronger improvements in mindfulness [64]. The continued development of mindfulness and reduction in psychological distress in the app-only group beyond T1 suggests that although classes boost training engagement and augment the benefits of app use, self-guided mindfulness training may still be beneficial with ongoing engagement; however, benefits may take longer to manifest.

Compared with the WLC group, no change was observed immediately after the intervention for either intervention group for participants' perceptions of psychosocial risk factors, job demand, control, and support. However, at 6 months, the app+classes group reported a reduction in job demands that approached significance and a trend toward higher job control compared with the app-only group. Job demands and control are key factors associated with work-related stress in the theoretical job-demands-resources model, where it is the perception that demands outweigh available resources that leads to job strain. Job strain is understood to be responsible for a range of workplace health and performance problems [33]. Mindfulness training aims to cultivate adaptive coping skills and should thus be considered a secondary level strategy for workplace health and well-being [12]. However, in this study, it appears that higher mindfulness may also support changes in the way psychosocial stressors are perceived. Our findings for job demands (and the trends for job control) indicate that the Smiling Mind Workplace Program app, when supported with classes, might be protective against job strain by reducing perceptions of imbalance between work-related demands and improving personal resources and perceived control over work experiences [3]. The fact that these effects were evident only at the 6- month follow-up might mean that changed perceptions of work-related psychosocial risks emerge sequentially following the development of higher mindfulness.

An explanation for the sequential development of benefits following mindfulness training is provided in the Garland [66] Mindfulness to Meaning model. According to this model, the initial stages of learning mindfulness meditation can help reduce stress reactivity by developing attentional control; however, it is the sustained application of mindful awareness in meditation practice that cultivates acceptance and reappraisal skills. These skills, in turn, support regulatory and coping resources and are known to underpin positive affect and general well-being [5,15,67].

The null result for quality of life was unexpected, given that significant improvements were recorded on the briefer 4-dimension AQoL following the pilot face-to-face workplace-based mindfulness program in the same population [35]. Moreover, prior work has shown increased general well-being following workplace-based mindfulness programs [2], even when delivered via an app [27]. Findings from an RCT of the Wildflowers mindfulness app in a nonwork setting [32] reported that changes in mindful acceptance appear to take longer and require a greater amount of meditation practice than changes in stress and mood. It is feasible that the degree of engagement with the app+classes intervention in this study was sufficient for the acquisition of elementary mindfulness skills (attentional control and awareness) that support stress appraisals and that these changes underpinned the beneficial findings for distress and psychosocial risk factors (job demand and job control). However, the training dose appears to have been inadequate for developing skills associated with positive affect

and general well-being, which are key factors associated with quality of life [66].

Trends in productivity data indicate that all 3 groups had decreased the number of health-related presenteeism and absenteeism days at the 6-month follow-up. Changes in productivity may also be sequential to changes in stress and mindfulness; however, our results did not show a causal link between mindfulness training and increased productivity. We propose that health-related LPT is an informative measure for assessing productivity effects in future workplace-based mindfulness program research, as higher mindfulness has been shown to alleviate psychological distress, depression, and anxiety, and these conditions are strongly associated with absenteeism and presenteeism [1,68].

The use of observer data to supplement self-reported changes in mindfulness and related behaviors addresses a limitation noted in approximately half of the published mindfulness studies [25]. Although the magnitude of interrater agreement was low, the consistent correspondence between self-reported mindfulness (Mindful Attention and Awareness Scale) and observer-reported mindful behaviors (OMM) strengthens the results reported in this study [69,70]. The work-based observers reported noticing increased mindful behaviors and a trend toward higher altruism among participants in the app+classes group but not in the app-only group at 6 months. These results lend weight to the potential for workplace-based mindfulness programs to have prosocial benefits in the workplace [18,71].

### Limitations, Strengths, and Future Research

There were timing and contextual considerations within our study. Baseline data collection coincided with the end of the summer break, a period during which many public sector employees are returning from annual leave. In contrast, the postintervention surveys coincided with political elections and major flooding in and around the state's capital city, where many public sector employees are located. Thus, employee stress levels may have been lower than usual in the preintervention surveys and elevated after the intervention through these contextual factors.

The necessary lack of blinding and use of a waitlist rather than an active control means that nonspecific factors such as social desirability, expectancy, or experimenter effects cannot be ruled out as potential effect moderators. For example, our qualitative data appear to suggest that participants in the app-only group may have felt their lower dose training protocol to have a lower status than the app+classes protocol. Careful design of the WLC conditions in future research is recommended to help address this bias risk. Although an additional survey was conducted 14 months from baseline (time point 3), there was a very high degree of attrition, with only 15.2% (32/211) of the starting sample providing data. Follow-up analyses were therefore limited to the 6-month data. Raw data for productivity and workplace incidents are provided in Multimedia Appendix 1 to support future pooled analyses.

Strengths of this study include participant characteristics reflecting those of the broader TSS workforce, meaning the reported findings can be generalized to similar public sector workplaces with some confidence. Collecting objective app use data enabled us to overcome a reliance on self-report adherence to the training protocol; however, we did not record engagement with the Smiling Mind generic emails and were therefore not able to include exposure to this guiding material in our dose-exposed calculations. The use of observer reports was another strength of this study, although the ceiling effects in the organizational citizenship and observed mindfulness data prevented complete analyses. The use of multisource data increases confidence in self-reported study findings, and this study has shown that the collection and use of observer-reported data are both feasible and informative. We suggest that more studies collect observer reports to help build an evidence base around the effects of mindfulness training on workplace social and performance outcomes. More work is needed to understand the effects of mindfulness training on workplace productivity and health-related LPT.

### Conclusions

Despite the absence of effects for the primary study outcome, that is, perceived stress, the results for mindfulness, distress, and job demands support the Smiling Mind Workplace Program app as a workplace stress reduction intervention when supported by classes. Importantly, no evidence of adverse effects was observed from this low-dose mindfulness intervention. However, previous workplace mindfulness training research [1,2] indicates that workplace-based mindfulness programs with stronger engagement and higher training doses are likely to realize greater benefits, both for employees' stress-related health and well-being and for organizational outcomes such as productivity and performance.

## Acknowledgments

The project was conceived and progressed by LB as part of her PhD (Medical Sciences) at the University of Tasmania's Menzies Institute for Medical Research. Coauthors (ALN [primary], KS, AJM, and MK) were supervisors for the first author (LB). PO provided statistical expertise. LB led to the development of the manuscript, with contributions from all coauthors. LB has since commenced a postdoctoral appointment at the University of Tasmania's Wicking Dementia Research and Education Centre. The authors are very grateful to the study participants for their involvement in this research and for the partnership between the Tasmanian State Service Management Office, the Tasmanian Training Consortium, Smiling Mind, and the University of Tasmania, which underpinned the project. Valuable assistance was provided by Sue Cole (volunteer) and Tim Albion (survey development).

The reported research was undertaken by (LB) as part of a PhD project. (LB) was supported throughout her PhD by an Australian Research Training Program scholarship and the TasNetworks Elite Health and Wellbeing Scholarship. ALN was supported by a Select Foundation Senior Research Fellowship.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[PDF File (Adobe PDF File), 1321 KB - mhealth_v10i2e30272_app1.pdf ]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1175 KB - mhealth_v10i2e30272_app2.pdf ]

## References

1. Bartlett L, Martin A, Neil AL, Memish K, Otahal P, Kilpatrick M, et al. A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. J Occup Health Psychol 2019 Feb;24(1):108-126. [doi: 10.1037/ocp0000146] [Medline: 30714811]
2. Vonderlin R, Biermann M, Bohus M, Lyssenko L. Mindfulness-based programs in the workplace: a meta-analysis of randomized controlled trials. Mindfulness 2020 Mar 02;11(7):1579-1598. [doi: 10.1007/s12671-020-01328-3]
3. Grover SL, Teo ST, Pick D, Roche M. Mindfulness as a personal resource to reduce work stress in the job demands-resources model. Stress and Health 2017 Oct;33(4):426-436. [doi: 10.1002/smi.2726] [Medline: 27862960]
4. Roche M, Haar JM, Luthans F. The role of mindfulness and psychological capital on the well-being of leaders. J Occup Health Psychol 2014 Oct;19(4):476-489. [doi: 10.1037/a0037183] [Medline: 24933594]
5. Avey JB, Luthans F, Jensen SM. Psychological capital: a positive resource for combating employee stress and turnover. Hum Resource Manage 2009 Sep;48(5):677-693. [doi: 10.1002/hrm.20294]
6. Ganster DC, Rosen CC. Work stress and employee health. J Manag 2013 Feb 19;39(5):1085-1122. [doi: 10.1177/0149206313475815]
7. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
8. Johns G. Attendance dynamics at work: the antecedents and correlates of presenteeism, absenteeism, and productivity loss. J Occup Health Psychol 2011 Oct;16(4):483-500. [doi: 10.1037/a0025153] [Medline: 21875212]
9. Cocker F, Sanderson K, LaMontagne AD. Estimating the economic benefits of eliminating job strain as a risk factor for depression. J Occup Environ Med 2017 Jan;59(1):12-17. [doi: 10.1097/JOM.0000000000000908] [Medline: 28045792]
10. Demsky C. Workplace aggression: a multi-study examination of work and nonwork consequences. Portland State University. 2015. URL: https://pdxscholar.library.pdx.edu/open_access_etds/2338/ [accessed 2022-01-21]
11. Memish K, Martin A, Bartlett L, Dawkins S, Sanderson K. Workplace mental health: an international review of guidelines. Prev Med 2017 Aug;101:213-222. [doi: 10.1016/j.ypmed.2017.03.017] [Medline: 28347696]
12. LaMontagne AD, Martin A, Page KM, Reavley NJ, Noblet AJ, Milner AJ, et al. Workplace mental health: developing an integrated intervention approach. BMC Psychiatry 2014 May 09;14:131 [FREE Full text] [doi: 10.1186/1471-244X-14-131] [Medline: 24884425]
13. Bhui K, Dinos S, Galant-Miecznikowska M, de Jongh B, Stansfeld S. Perceptions of work stress causes and effective interventions in employees working in public, private and non-governmental organisations: a qualitative study. BJPsych Bull 2016 Dec;40(6):318-325 [FREE Full text] [doi: 10.1192/pb.bp.115.050823] [Medline: 28377811]
14. Creswell JD, Lindsay EK, Villalba DK, Chin B. Mindfulness training and physical health: mechanisms and outcomes. Psychosom Med 2019 Apr;81(3):224-232 [FREE Full text] [doi: 10.1097/PSY.0000000000000675] [Medline: 30806634]
15. Folkman S, Moskowitz JT. Positive affect and the other side of coping. Am Psychol 2000 Jun;55(6):647-654. [doi: 10.1037//0003-066x.55.6.647] [Medline: 10892207]
16. Garland EL, Hanley AW, Baker AK, Howard MO. Biobehavioral mechanisms of mindfulness as a treatment for chronic stress: an RDoC perspective. Chronic Stress (Thousand Oaks) 2017 Feb;1:2470547017711912 [FREE Full text] [doi: 10.1177/2470547017711912] [Medline: 28840198]
17. Tang Y, Hölzel BK, Posner MI. The neuroscience of mindfulness meditation. Nat Rev Neurosci 2015 Apr;16(4):213-225. [doi: 10.1038/nrn3916] [Medline: 25783612]
18. Donald JN, Sahdra BK, Van Zanden B, Duineveld JJ, Atkins PW, Marshall SL, et al. Does your mindfulness benefit others? A systematic review and meta-analysis of the link between mindfulness and prosocial behaviour. Br J Psychol 2019 Feb;110(1):101-125. [doi: 10.1111/bjop.12338] [Medline: 30094812]

19. Karremans JC, van Schie HT, van Dongen I, Kappen G, Mori G, van As S, et al. Is mindfulness associated with interpersonal forgiveness? Emotion 2020 Mar;20(2):296-310. [doi: 10.1037/emo0000552] [Medline: 30652883]
20. Gillions A, Cheang R, Duarte R. The effect of mindfulness practice on aggression and violence levels in adults: a systematic review. Aggression Violent Behav 2019 Sep;48:104-115. [doi: 10.1016/j.avb.2019.08.012]
21. Good DJ, Lyddy CJ, Glomb TM, Bono JE, Brown KW, Duffy MK, et al. Contemplating mindfulness at work. J Manag 2015 Nov 19;42(1):114-142. [doi: 10.1177/0149206315617003]
22. Reb J, Choi E. Mindfulness in organizations. In: Psychology of meditation. Hauppauge, NY: Nova Science Publishers; 2014.
23. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
24. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
25. Goldberg SB, Tucker RP, Greene PA, Simpson TL, Kearney DJ, Davidson RJ. Is mindfulness research methodology improving over time? A systematic review. PLoS One 2017;12(10):e0187298 [FREE Full text] [doi: 10.1371/journal.pone.0187298] [Medline: 29088283]
26. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
27. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
28. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
29. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
30. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
31. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
32. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 08;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]
33. Bakker AB, Demerouti E. Job demands-resources theory: taking stock and looking forward. J Occup Health Psychol 2017 Jul;22(3):273-285. [doi: 10.1037/ocp0000056] [Medline: 27732008]
34. Creswell JD, Lindsay EK. How does mindfulness training affect health? A mindfulness stress buffering account. Curr Dir Psychol Sci 2014 Dec 16;23(6):401-407. [doi: 10.1177/0963721414547415]
35. Bartlett L, Lovell P, Otahal P, Sanderson K. Acceptability, feasibility, and efficacy of a workplace mindfulness program for public sector employees: a pilot randomized controlled trial with informant reports. Mindfulness 2016 Dec 1;8(3):639-654. [doi: 10.1007/s12671-016-0643-4]
36. Smiling Mind workplace program. Smiling Mind. URL: https://info.smilingmind.com.au/smiling-mind-workplace-program-?hsCtaTracking=46dc6b21-4edd-4a73-924c-d94c95ccc6be%7Cddd84d27-d66b-42ea-b35f-fc8d404c2ee8 [accessed 2022-01-21]
37. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309 [FREE Full text] [doi: 10.7326/0003-4819-148-4-200802190-00008] [Medline: 18283207]
40. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
41. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

42. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
43. Slade T, Grove R, Burgess P. Kessler Psychological Distress Scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2011 Apr;45(4):308-316. [doi: 10.3109/00048674.2010.543653] [Medline: 21332432]
44. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
45. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient 2014;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]
46. Butterworth P, Leach LS, Rodgers B, Broom DH, Olesen SC, Strazdins L. Psychosocial job adversity and health in Australia: analysis of data from the HILDA Survey. Aust N Z J Public Health 2011 Dec;35(6):564-571. [doi: 10.1111/j.1753-6405.2011.00771.x] [Medline: 22151164]
47. Sanne B, Torp S, Mykletun A, Dahl AA. The Swedish Demand-Control-Support Questionnaire (DCSQ): factor structure, item analyses, and internal consistency in a large population. Scand J Public Health 2005;33(3):166-174. [doi: 10.1080/14034940410019217] [Medline: 16040456]
48. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? J Affect Disord 2007 Aug;101(1-3):65-74. [doi: 10.1016/j.jad.2006.10.024] [Medline: 17156851]
49. Bartlett L, Martin AJ, Bruno R, Kilpatrick M, Sanderson K, Neil AL. Is mindfulness a noticeable quality? Development and validation of the observed mindfulness measure. J Psychopathol Behav Assess 2021 Nov 17. [doi: 10.1007/s10862-021-09936-6]
50. Smith CA, Organ DW, Near JP. Organizational citizenship behavior: its nature and antecedents. J Applied Psychol 1983;68(4):653-663. [doi: 10.1037/0021-9010.68.4.653]
51. Cohen J. A power primer. Psychol Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
52. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
53. R: a language and environment for statistical computing. The R Foundation. URL: https://www.r-project.org/foundation/ [accessed 2022-01-21]
54. psych: procedures for personality and psychological research. Northwestern University. URL: https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research [accessed 2022-01-21]
55. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Softw 2015;67(1). [doi: 10.18637/jss.v067.i01]
56. Kuznetsova A, Brockhoff PB, Christensen RH. Package: tests in linear mixed effects models. J Stat Softw 2017;82(13). [doi: 10.18637/jss.v082.i13]
57. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
58. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
59. Frost N. Qualitative Research Methods in Psychology: Combining Core Approaches. New York, United States: McGraw-Hill Education; 2011.
60. Kessler RC, Sampson NA, Berglund P, Gruber MJ, Al-Hamzawi A, Andrade L, et al. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. Epidemiol Psychiatr Sci 2015 Jun;24(3):210-226 [FREE Full text] [doi: 10.1017/S2045796015000189] [Medline: 25720357]
61. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
62. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
63. Virgili M. Mindfulness-based interventions reduce psychological distress in working adults: a meta-analysis of intervention studies. Mindfulness 2013 Dec 13;6(2):326-337. [doi: 10.1007/s12671-013-0264-0]
64. Strohmaier S. The relationship between doses of mindfulness-based programs and depression, anxiety, stress, and mindfulness: a dose-response meta-regression of randomized controlled trials. Mindfulness 2020 Mar 02;11(6):1315-1335. [doi: 10.1007/s12671-020-01319-4]
65. Is mindfulness safe? Oxford Mindfulness Centre. URL: https://www.oxfordmindfulness.org/news/is-mindfulness-safe/ [accessed 2022-01-21]

66. Garland EL, Hanley AW, Goldin PR, Gross JJ. Testing the mindfulness-to-meaning theory: evidence for mindful positive emotion regulation from a reanalysis of longitudinal data. PLoS One 2017;12(12):e0187727 [FREE Full text] [doi: 10.1371/journal.pone.0187727] [Medline: 29211754]
67. Folkman S. Stress: appraisal and coping. In: Encyclopedia of Behavioral Medicine. New York: Springer; 2013.
68. Sanderson K, Andrews G. Common mental disorders in the workforce: recent findings from descriptive and social epidemiology. Can J Psychiatry 2006 Feb;51(2):63-75. [doi: 10.1177/070674370605100202] [Medline: 16989105]
69. Olino TM, Klein DN. Psychometric comparison of self- and informant-reports of personality. Assessment 2015 Dec;22(6):655-664 [FREE Full text] [doi: 10.1177/1073191114567942] [Medline: 25612626]
70. Vazire S, Mehl MR. Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. J Pers Soc Psychol 2008 Nov;95(5):1202-1216. [doi: 10.1037/a0013314] [Medline: 18954202]
71. Choi E, Tobias J. Mind the gap: the link between mindfulness and performance at work needs more attention. Industrial Organ Psychol 2015 Dec 17;8(4):629-633. [doi: 10.1017/iop.2015.90]

## Abbreviations

**AQoL:** Assessment of Quality of Life
**CHERRIES:** Checklist for Reporting Results of Internet e-Surveys
**CONSORT:** Consolidated Standards of Reporting Trials
**ICC:** intraclass correlation coefficient
**LPT:** lost productive time
**OMM:** observed mindfulness measure
**PHQ-9:** Patient Health Questionnaire-9
**PSS:** Perceived Stress Scale
**RCT:** randomized controlled trial
**REDCap:** Research Electronic Data Capture
**T0:** time point 0
**T1:** time point 1
**T2:** time point 2
**TSS:** Tasmanian State Service
**TTC:** Tasmanian Training Consortium
**WLC:** waitlist control

*Edited by L Buis; submitted 08.05.21; peer-reviewed by P Jimenez, S Taki; comments to author 08.09.21; revised version received 22.09.21; accepted 20.12.21; published 10.02.22.*

*Please cite as:*
*Bartlett L, Martin AJ, Kilpatrick M, Otahal P, Sanderson K, Neil AL*
*Effects of a Mindfulness App on Employee Stress in an Australian Public Sector Workforce: Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2022;10(2):e30272*
*URL: https://mhealth.jmir.org/2022/2/e30272*
*doi:10.2196/30272*
*PMID:35142630*

Original Paper

# Effects on Adherence to a Mobile App–Based Self-management Digital Therapeutics Among Patients With Coronary Heart Disease: Pilot Randomized Controlled Trial

Yuxi Li[1,2,3], MD; Yanjun Gong[1,2], MD; Bo Zheng[1,2,3], MD; Fangfang Fan[1,2,3], MD; Tieci Yi[1,2,3], MD; Yimei Zheng[1], BSc; Pengkang He[1,2], MD; Jin Fang[1], BSc; Jia Jia[1,2,3], MPH; Qin Zhu[4], MSc; Jie Jiang[1,2*], MD; Yong Huo[1,2*], MD

[1]Department of Cardiology, Peking University First Hospital, Beijing, China

[2]Institute of Cardiovascular Disease, Peking University First Hospital, Beijing, China

[3]Hypertension Precision Diagnosis and Treatment Research Center, Peking University First Hospital, Beijing, China

[4]Stragety & New Business Development of Philips Greater China, Shanghai, China

[*]these authors contributed equally

**Corresponding Author:**
Yong Huo, MD
Department of Cardiology
Peking University First Hospital
Xishiku No. 8, Xicheng District
Beijing, 100034
China
Phone: 86 83572283
Email: huoyong@263.net.cn

## Abstract

**Background:** The adherence to secondary prevention treatment in patients with coronary heart disease (CHD) is low. Digital therapeutics (DTx) refers to an emerging branch of medicine that delivers medical interventions directly to patients using evidence-based, clinically evaluated, technology-based software algorithms or apps to facilitate disease management, which may be an efficient tool to optimize adherence.

**Objective:** This paper aims to investigate the effect of mobile app–based self-management DTx on long-term use of secondary prevention medications in patients with CHD in China.

**Methods:** This pilot study was a parallel-designed, open-labeled, single-center, randomized controlled trial. Hospitalized patients with CHD admitted to Peking University First Hospital between April 2016 and June 2017 were randomized before discharge on a 1:1 ratio. The intervention group received regular follow-up combined with DTx, which is a self-management mobile app already installed on an Android 5 (Mi Pad 1, Xiaomi Corporation) tablet. Structured data from the hospital informatics system were integrated automatically, and medication, lifestyle intervention plan, follow-up protocol, and patient education materials were also provided according to the diagnosis. Participants could use DTx for self-management at home. The control group was under conventional hospital–based follow-up care. Patients were followed up for 1 year, and the primary end point was the percentage of all guideline-recommended medications at 12 months. The secondary end points included the percentage adhered to standard secondary prevention medications at 6 months, the control rate of lipid profile, and blood pressure at 6 months and 1 year.

**Results:** Among 300 randomized patients with CHD, 290 (96.7%) were included in the final analysis, including 49.3% (143/290) and 50.7% (147/290) of patients from the intervention and control groups, respectively. Baseline characteristics were similar between the 2 groups. There was a statistically significant improvement in the percentage of all guideline-recommended medications at 12 months in the intervention group compared with the control group (relative risk [RR] 1.34, 95% CI 1.12-1.61; *P*=.001), and there was no interaction with baseline characteristics. The intervention group had a significantly higher proportion of patients achieving blood pressure under control (systolic blood pressure <140 mm Hg and diastolic blood pressure <90 mm Hg) and low-density lipoprotein cholesterol <1.8 mmol/L (RR 1.45, 95% CI 1.22-1.72; *P*<.001 and RR 1.40, 95% CI 1.11-1.75; *P*=.004, respectively) at 12 months. Furthermore, on logistic regression, the intervention group had a lower risk of withdrawing from guideline-recommended medications (odds ratio 0.46, 95% CI 0.27-0.78; *P*=.004).

**Conclusions:** Among patients with CHD, using a mobile app–based self-management DTx in addition to traditional care resulted in a significant improvement in guideline-recommended medication adherence at 12 months. The results of the trial will be applicable to primary care centers, especially in rural areas with less medical resources.

**Trial Registration:** ClinicalTrials.gov NCT03565978; https://clinicaltrials.gov/ct2/show/NCT03565978

***(JMIR Mhealth Uhealth 2022;10(2):e32251)*** doi:10.2196/32251

**KEYWORDS**

coronary heart disease; secondary prevention; self-management; mobile app; adherence; digital therapeutics; mobile phone

## Introduction

Coronary heart disease (CHD) is one of the leading causes of death globally, especially in China; the trend of mortality due to CHD is still increasing and the recurrence rate of major cardiovascular events remains high [1]. The long-term use of secondary prevention medications is widely recommended by the national and international guidelines, and proven to improve the prognosis [2-4].

However, the adherence rate to long-term secondary prevention therapies only varies from 30% to 50%, and it is even worse in limited-income countries [5-8], which shows the big gap between the real-world practice and the recommended guidelines and, thus, being a key challenge limiting the overall benefits of these therapies. Poor adherence has been demonstrated to be associated with a 50% to 80% relative increase in mortality and increased health care cost [9,10].

The causes of nonadherence are complex and not due to a single reason only; however, a good medical system with primary care support and a long-term follow-up plan, including cardiac rehabilitation programs, for patients with CHD, especially those discharged from a hospital, is crucial. For overpopulated countries, such as China, the lack of specialized cardiac rehabilitation staff is one of the main reasons for poor adherence [5,11]. A novel and effective management model other than conventional, hospital-based follow-up interventions is needed to optimize the long-term treatment of CHD. Recent advances in digital therapeutics (DTx), which delivers medical interventions directly to patients using evidence-based, clinically evaluated, technology-based software algorithms or apps to facilitate disease management, such as smartphones and technology, have made DTx a promising solution for secondary prevention management of chronic diseases [12-14]. SMS text messaging is one of the mobile health (mHealth) approaches that have been demonstrated to significantly reduce the low-density lipoprotein cholesterol (LDL-C) level, systolic blood pressure (SBP), and BMI of patients with CHD [15,16]. Recently published systematic review articles have shown that mHealth significantly improved patients' cardiovascular risk factors rather than mortality for secondary prevention; however, among the trials included in the meta-analysis, only a very few studies were conducted in limited-income countries [17,18].

Furthermore, patient self-management is one of the key issues in the long-term management of secondary prevention among patients with CHD [7]. This study aims to explore, in a randomized controlled trial, the effect of a mobile app designed for self-management DTx at home on the long-term use of secondary prevention medications in patients with CHD in China.

## Methods

### Study Design

This BAMA (name of a famous *longevity village* in the northwest of Guangxi province, China) pilot study was a parallel-designed, open-labeled, single-center, randomized controlled trial conducted in Peking University First Hospital, China. Patients with CHD were randomized to receive conventional hospital-based care and management (control group) or in addition to use a mobile app for self-management DTx (intervention group). Patients were followed up for 1 year. The objective measures of cardiovascular risk factors, including LDL-C level, blood pressure, and the percentage of patients who were taking guideline-recommended medications at 6 months and 1 year postrandomization were compared between the 2 groups.

Patients provided written informed consent, and the study protocol, in compliance with the principles outlined in the Declaration of Helsinki, was approved by the Institutional Ethics Committee of Peking University First Hospital and registered at ClinicalTrials.gov (NCT03565978).

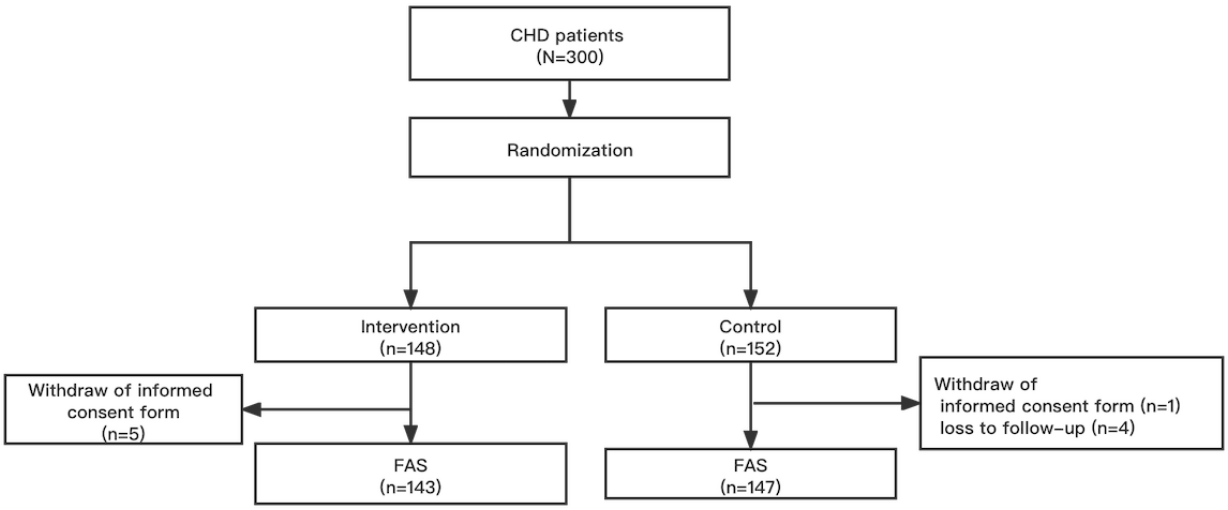
### Study Population

The patients were enrolled into the study only if they met all of the following criteria: (1) were male or female aged ≥18 years; (2) were diagnosed with CHD (defined as documented prior myocardial infarction, coronary artery bypass graft surgery, percutaneous coronary intervention, or ≥50% stenosis in at least one major epicardial vessel on coronary angiography); (3) were willing to undergo a self-management intervention and comply with the follow-up plan; (4) had basic reading skills in Chinese; and (5) voluntarily participated in the study and signed the informed consent form. Patients were excluded if they met one of the following criteria: (1) already enrolled in another interventional clinical trial, (2) refused to sign the informed consent or withdrew from the study for any specific reasons, (3) had cognitive disorder and were unable to communicate normally, and (4) had limited basic mobile technology skills to operate a mobile app after training.

Recruitment was performed between April 2016 and June 2017, and the follow-up continued until June 2018. The participant flowchart is shown in Figure 1. All participants were recruited before discharge from the hospital, after being hospitalized because of CHD. A comparison between recruited participants and hospitalized patients with CHD during the same period who

did not participate in this trial can be found in Table S1 in Multimedia Appendix 1. Enrollment of participants was continued until the planned sample size was reached.

**Figure 1.** Flowchart of the study participants. CHD: coronary heart disease; FAS: full analysis set.




## Group Allocation and Intervention

Randomization was conducted using a computerized web-based randomization program. The random allocation sequence was in a uniform 1:1 allocation ratio. Patients were randomized to the intervention group, which comprised participants who underwent regular follow-up combined with a mobile app–based self-management DTx, or the control group, which comprised participants who only received regular follow-up care and patient education. The control group was treated and followed up as usual. Printed patient education materials were given to patients at the beginning of the study and during each visit, and the same content was sent to the intervention group through the self-management mobile app.

The DTx for the intervention group is a self-management mobile app already installed on an Android tablet (provided by the study), and the app could also be installed on iOS or smartphones. The DTx system includes a physician portal, a health manager portal, and a patient portal and contains 3 modules. The first one is the discharge module, which was used before discharge. Structured data from the hospital informatics system were integrated automatically in the DTx system, and trial staff confirmed the medication and lifestyle intervention plan and arranged follow-up protocol in the DTx system and educated participants on how to use the patient portal. Patient education materials and instructions on medication were also provided in the app according to the diagnosis at discharge.

The second module is the home management module. An electronic blood pressure meter was given to all participants at the beginning of the study to encourage self-management. The intervention group had their blood pressure and heart rate data transferred through Bluetooth connection to the app, or they could also input the data on the app by themselves. In addition, an automatic alarm was set up in the patient portal to help manage their daily medical regiment. Participants in the intervention group could also record symptoms and notes in the DTx app, whereas the control group was asked to record their data manually on the patients' education book. All data recorded by the home management module are shared with trial staff physicians and nurses in the physician portal.

The third module is the follow-up module, which was used during the follow-up; the DTx had a dynamic design and dashboard overview displaying the latest discharge summaries, vital signs, symptoms, and medications. Physicians could update medication and lifestyle plans during each follow-up in the DTx, whereas participants in the control group used traditional electronic health records. An illustrated description slide is provided in Multimedia Appendix 2.

## Trial Procedures

In all participants, baseline demographic characteristics, experience in smartphone use, access to Wi-Fi at home, prior medical history, and tobacco use were assessed by nurses during their hospitalization. Blood pressure, heart rate, and BMI were measured according to standardized procedures. Blood pressure and heart rate were measured 3 times using an automatic Omron device (Omron Healthcare, Inc), and mean readings were used for the analyses. Plasma creatinine, LDL-C, and hemoglobin $A_{1c}$ levels were analyzed at local laboratories. The estimated glomerular filtration rate was calculated based on the Modification of Diet in Renal Disease equation [19]. The left ventricular ejection fraction was measured by certified cardiologists. The Teich method was used, unless the measurement based on the Simpson method was used if the patient had prior myocardial infarction or significant ventricle dilation.

An invasive diagnostic coronary angiography was performed during the hospitalization, and the secondary prevention treatment regimen was according to the discretion of the physician based on the patients' specific clinical condition. All patients were asked to visit the outpatient clinic every 3 months, which was also the routine follow-up plan for patients with

coronary diseases in this hospital. Blood pressure and heart rate were measured during each follow-up visit according to the aforementioned standard procedure. LDL-C levels were measured at the 3-, 6-, and 12-month follow-up visits at the local laboratories.

The primary end point of the study was the percentage of patients who used the standard guideline-recommended secondary prevention medications at 1 year after enrollment. This medication regimen included (1) aspirin, P2Y12 receptor inhibitor, or both in patients who had acute coronary syndrome or percutaneous coronary intervention during hospitalization; (2) statins; and (3) angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers and β-receptor blockers for patients with a history of myocardial infarction or heart failure; this regimen was provided to the patients if there was not any contraindication. During the follow-up visit at 6 and 12 months, the physicians obtained information on whether the patients were using each of the medications.

The secondary efficacy end points of the study included the percentage of patients who took standard secondary prevention medications at 6 months and the control rate of lipid profile (LDL-C <1.8 mmol/L) and blood pressure (SBP <140 mm Hg and diastolic blood pressure [DBP] <90 mm Hg) at 6 months and 1 year after enrollment.

### Statistical Analysis

According to previous data, the percentage of patients who used the standard guideline-recommended secondary prevention medications was approximately 60% when there was no intervention. The proportion in the intervention group was expected to increase to 75%; thus, 152 participants were needed for each group to have a power of 80% (2-tailed and at a 5% significance level) to detect the difference. All intervention evaluations were performed on the principle of per-protocol analysis. Participants were analyzed by the original assigned groups.

Description of continuous variables are summarized as means and SDs unless skewed and then presented as medians and IQRs. Categorical variables are presented as frequencies and percentages.

For the primary and categorical secondary end points, the proportion of patients with a positive primary end point was compared between the 2 randomized groups (intervention and control) using a log-binomial regression, and the result was presented as relative risk (RR) and 95% CI. Similarly, the mean differences in blood pressure, heart rate, and LDL-C levels were compared between the 2 randomized groups (intervention and control) using a log-binomial regression, and the result was presented as the mean difference and 95% CI. A logistic regression analysis was made to compare the rate of withdrawal from the guideline-recommended medications within the 12-month follow-up period between the 2 groups, and the result was presented as odds ratio and 95% CI. The analyses were otherwise unadjusted. Subgroup analyses were conducted by sex, age, BMI, current tobacco use, prior history of hypertension, dyslipidemia, left ventricular ejection fraction, and discharge diagnoses of stable CHD versus acute coronary syndrome using logistic regression.

Data were collected during each follow-up visit. Two individuals with experience in data entry independently entered all data into separate EpiData (version 3.1, EpiData Software) databases. The 2 databases were compared, and discrepancies were resolved by checking the original data. Data quality monitoring was conducted by an independent monitoring staff. If necessary, data queries were also made.

Analyses were conducted using R (version 3.6.2; R Foundation for Statistical Computing) [20] and RStudio (version 1.2.5033; RStudio, Inc) [21]. All statistical tests were 2-tailed, and a 5% significance threshold was maintained.

## Results

### Overview

Between April 8, 2016, and June 28, 2017, 300 patients who met the inclusion and exclusion criteria were enrolled in the study. Figure 1 illustrates the flow diagram of the participant recruitment, randomization, and waning throughout the trial. As mentioned above, given that the recruitment was based on the willingness of patients and their ability to handle smartphones, it was difficult to determine the number of potential patients who met the inclusion and exclusion criteria. A comparison between recruited patients and hospitalized patients with CHD during the same period who did not participate in this trial is shown in Table S1 in Multimedia Appendix 1, and the results showed no significant differences in the baseline characteristics between the 2 groups. Randomization yielded 49.3% (148/300) and 50.7% (152/300) patients in the intervention and control groups, respectively. However, 2% (6/300) of the patients withdrew consent after randomization, and 1.3% (4/300) of the patients were lost to follow-up during the study. Of the 300 patients, 290 (96.7%) patients were followed up for 12 months after randomization.

As shown in Table 1, the mean age 61.38 (SD 8.88) years versus 62.27 (SD 9.87) years ($P$=.42) and sex proportion (29/143, 20.3% vs 36/147, 24.5%, female sex; $P$=.47) between the 2 groups were similar. Other baseline characteristics, including prior medical history, cardiac risk factors, laboratory tests, CHD presentation, and angiographic features, were all similar between the 2 groups.

**Table 1.** Patients' baseline characteristics.

| | All (n=290) | Intervention (n=143) | Control (n=147) | *P* value |
|---|---|---|---|---|
| **Demographics** | | | | |
| Female sex, n (%) | 65 (22.4) | 29 (20.3) | 36 (24.5) | .47 |
| Age (years), mean (SD) | 61.83 (9.39) | 61.38 (8.88) | 62.27 (9.87) | .42 |
| Han, n (%) | 271 (94.1) | 137 (96.5) | 134 (91.8) | .11 |
| Never used a smartphone, n (%) | 25 (8.8) | 7 (5) | 18 (12.4) | .15 |
| Wi-Fi available at home, n (%) | 264 (92.6) | 131 (93.6) | 133 (91.7) | .71 |
| **Prior medical history, n (%)** | | | | |
| Myocardial infarction | 48 (16.6) | 21 (14.7) | 27 (18.4) | .49 |
| Heart failure | 12 (4.1) | 8 (5.6) | 4 (2.7) | .35 |
| Prior PCI[a] | 80 (27.6) | 40 (28) | 40 (27.2) | .99 |
| Prior CABG[b] | 4 (1.4) | 3 (2.1) | 1 (0.7) | .60 |
| Hypertension | 196 (67.6) | 99 (69.2) | 97 (66) | .64 |
| Diabetes | 131 (45.2) | 68 (47.6) | 63 (42.9) | .49 |
| Dyslipidemia | 176 (60.7) | 86 (60.1) | 90 (61.2) | .95 |
| Renal insufficiency | 26 (9) | 11 (7.7) | 15 (10.2) | .59 |
| Current dialysis | 4 (1.4) | 3 (2.1) | 1 (0.7) | .60 |
| Cerebral disease | 35 (12.1) | 13 (9.1) | 22 (15) | .18 |
| Peripheral artery disease | 5 (1.7) | 1 (0.7) | 4 (2.7) | .38 |
| Current tobacco use (<1 year) | 179 (61.7) | 89 (62.2) | 90 (61.2) | .61 |
| Premature CHD[c] family history | 36 (12.4) | 19 (13.3) | 17 (11.6) | .25 |
| **Physical examination** | | | | |
| BMI (kg/m$^2$), mean (SD) | 26.11 (3.27) | 25.93 (3.07) | 26.28 (3.46) | .37 |
| BMI>24 (kg/m$^2$), n (%) | 218 (76) | 109 (77.9) | 109 (74.2) | .55 |
| SBP[d] (mm Hg), mean (SD) | 135.70 (20.03) | 134.43 (18.48) | 136.93 (21.42) | .29 |
| DBP[e] (mm Hg), mean (SD) | 78.31 (12.43) | 77.74 (11.35) | 78.86 (13.42) | .44 |
| HR[f] (bpm[g]), mean (SD) | 70.84 (10.07) | 70.09 (9.39) | 71.58 (10.67) | .21 |
| **Laboratory results** | | | | |
| eGFR[h] (mL/min × 1.73 m²), mean (SD) | 76.77 (19.27) | 76.33 (18.09) | 77.19 (20.40) | .70 |
| eGFR≥60 (mL/min × 1.73 m²), n (%) | 245 (84.5) | 122 (85.3) | 123 (83.7) | .82 |
| LDL[i] (mmol/L), mean (SD) | 2.47 (0.86) | 2.41 (0.88) | 2.53 (0.85) | .25 |
| $HbA_{1c}$[j] (%), mean (SD) | 6.89 (1.55) | 6.96 (1.58) | 6.82 (1.52) | .48 |
| LVEF[k] (%), mean (SD) | 67.01 (10.48) | 66.91 (11.01) | 67.12 (9.97) | .87 |
| **Discharge diagnoses, n (%)** | | | | .80 |
| Stable angina | 54 (18.6) | 25 (17.5) | 29 (19.7) | |
| Unstable angina | 181 (62.4) | 92 (64.3) | 89 (60.5) | |
| Acute myocardial infarction | 55 (19) | 26 (18.2) | 29 (19.7) | |
| **Coronary artery lesions, n (%)** | | | | .20 |
| Single-vessel lesions | 100 (34.5) | 55 (38.5) | 45 (30.6) | |
| Double-vessel lesions | 90 (31) | 43 (30.1) | 47 (32) | |
| Triple-vessel lesions | 98 (33.8) | 43 (30.1) | 55 (37.4) | |

| | All (n=290) | Intervention (n=143) | Control (n=147) | *P* value |
|---|---|---|---|---|
| Left main lesion | 2 (0.7) | 2 (1.4) | 0 (0) | |
| Number of stents, median (quartile 1, quartile 3) | 2 (1, 2) | 2 (1, 2) | 2 (1, 2.5) | .50 |
| Full revascularization, n (%) | 192 (66.2) | 96 (67.1) | 96 (65.3) | .84 |

[a]PCI: percutaneous coronary intervention.
[b]CABG: coronary artery bypass grafting.
[c]CHD: coronary heart disease.
[d]SBP: systolic blood pressure.
[e]DBP: diastolic blood pressure.
[f]HR: heart rate.
[g]bpm: beat per minute.
[h]eGFR: estimated glomerular filtration rate.
[i]LDL: low-density lipoprotein.
[j]$HbA_{1c}$: hemoglobin $A_{1c}$.
[k]LVEF: left ventricular ejection fraction.

## Primary Outcome

As shown in Table 2, there was a statistically significant improvement in the percentage of patients using all guideline-recommended medications at 12 months in the intervention group compared with the control group (RR 1.34, 95% CI 1.12-1.61; *P*=.001). As shown in Table S2 in Multimedia Appendix 1, similar results were found for aspirin and P2Y12 receptor inhibitors. Regarding statin use, there was a marginal improvement observed in the intervention group compared with the control group; however, there were no differences in angiotensin-converting enzyme inhibitor or angiotensin-receptor blocker or β-receptor blocker use between the 2 groups. However, there was no interaction between baseline characteristics and intervention for the primary outcome in all subgroups (Multimedia Appendix 1, Figure S1).

The intervention group had a significantly higher proportion of patients with controlled blood pressure (SBP <140 mm Hg and DBP <90 mm Hg) and LDL-C level <1.8 mmol/L (RR 1.45, 95% CI 1.22-1.72; *P*<.001 and RR 1.40, 95% CI 1.11-1.75; *P*=.004, respectively) at 12 months compared with the control group. At 6 months, significant differences in blood pressure levels (RR 1.38, 95% CI 1.16-1.64; *P*<.001) were found between the 2 groups; however, no significant difference in LDL-C levels was found (RR 1.18, 95% CI 0.96-1.46; *P*=.12; Table 2).

**Table 2.** Secondary prevention medication and risk factor control between the 2 groups (N=290).

| | Intervention (n=143) | Control (n=147) | Relative risk (95% CI) | *P* value |
|---|---|---|---|---|
| **Primary end point** | | | | |
| All medications at 12 months, n (%) | 103 (72) | 79 (53.7) | 1.34 (1.12-1.61) | .001 |
| Adherence score at 12 months, mean (SD) | 0.72 (1.18) | 0.69 (1.18) | 0.71 (0.77-1.25) | .81 |
| **Secondary end point, n (%)** | | | | |
| All medications at 6 months | 121 (86.4) | 101 (69.2) | 1.25 (1.10-1.42) | <.001 |
| Adherence score at 6 months | 1.08 (1.45) | 1.01 (1.26) | 1.05 (0.58-1.36) | .69 |
| **Risk factors at 12 months, n (%)** | | | | |
| LDL[a]<1.8 mmol/L | 87 (60.8) | 64 (43.5) | 1.40 (1.11-1.75) | .004 |
| SBP[b]<140 mm Hg and DBP[c]<90 mm Hg | 99 (82.5) | 75 (56.8) | 1.45 (1.22-1.72) | <.001 |
| **Risk factors at 6 months, n (%)** | | | | |
| LDL<1.8 mmol/L | 85 (59.4) | 74 (50.3) | 1.18 (0.96-1.46) | .12 |
| SBP<140 mm Hg and DBP<90 mm Hg | 96 (67.1) | 77 (52.4) | 1.38 (1.16-1.64) | <.001 |

[a]LDL: low-density lipoprotein.
[b]SBP: systolic blood pressure.
[c]DBP: diastolic blood pressure.

### Change in Primary End Point From Baseline to the End of 12 Months

Figure 2 shows the changing trends of primary outcome measures at baseline and at 6 and 12 months between the 2 groups. Participants in both the intervention and control groups had poor adherence to key secondary prevention medications during the 1-year follow-up duration; however, the proportion of patients on standard medications was higher in the intervention group than in the control group. On logistic regression, participants in the intervention group had a lower risk of withdrawal from the guideline-recommended medications (odds ratio 0.46, 95% CI 0.27-0.78; *P*=.004).

As shown in Figure 3, the trends of changes in blood pressure, LDL-C levels, and heart rate were analyzed. During the study, SBP and DBP decreased during the first 3 months in both groups, but the decrease was slightly higher in the intervention group. However, with the decrease in medication adherence, the blood pressure in the control group began to increase, especially for SBP, which almost returned to the baseline values at 12 months. Meanwhile, both SBP and DBP in the intervention group remained well controlled. The same pattern was found in the patients' lipid profiles. The trend in the changes in heart rate between the 2 groups was similar.

Accordingly, in the linear regression analysis (Table 3), the changes in the values from baseline to the 12-month follow-up for SBP, DBP, and LDL-C levels were all significantly higher in the intervention group than in the control group (change in SBP −11 mm Hg, 95% CI −15 to −7; *P*<.001; change in DBP −3 mm Hg, 95% CI −6 to −4; *P*=.02; change in LDL-C level −0.22 mmol/L, 95% CI, −0.34 to −0.09; *P*=.001). However, the change in heart rate was not significantly different between the 2 groups (change in heart rate −1 bpm, 95% CI −4 to −1; *P*=.30).

**Figure 2.** Proportion of patients taking the standard secondary prevention medications. OR: odds ratio.

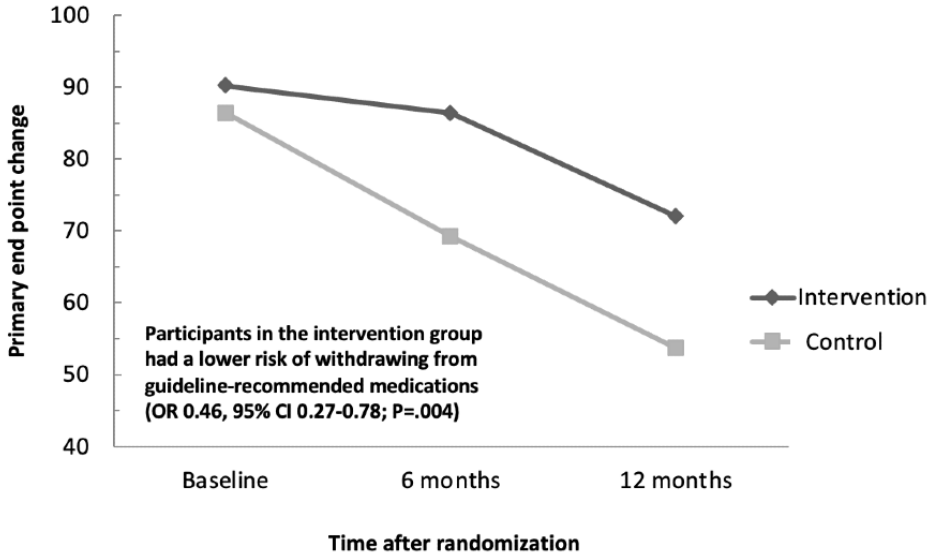


**Figure 3.** Blood pressure and low-density lipoprotein cholesterol (LDL-C) level changes between the 2 groups. bpm: beats per minute.

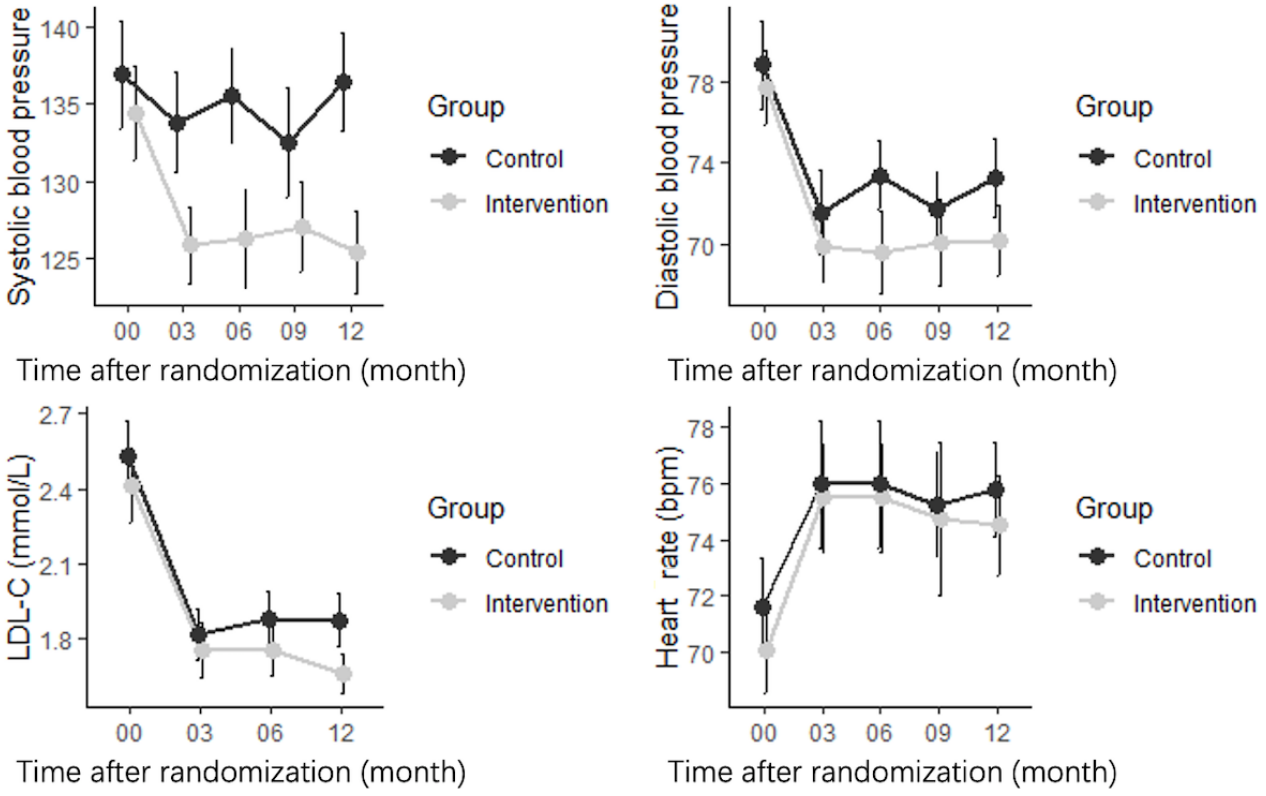



**Table 3.** The mean differences in blood pressure, heart rate, and low-density lipoprotein cholesterol (LDL-C) level between the 12-month follow-up and baseline values (N=290).

| Parameter | Value, mean (95% CI) | | Mean difference (95% CI) | *P* value |
|---|---|---|---|---|
| | Intervention (n=143) | Control (n=147) | | |
| SBP[a] (mm Hg) | 125 (122 to 127) | 136 (133 to 139) | −11 (−15 to −7) | <.001 |
| DBP[b] (mm Hg) | 70 (68 to 71) | 73 (71 to 75) | −3 (−6 to −4) | .02 |
| HR[c] (bpm[d]) | 74 (72 to 76) | 75 (74 to 77) | −1 (−4 to −1) | .30 |
| LDL-C (mmol/L) | 1.66 (1.58 to 1.74) | 1.87 (1.77 to 1.98) | −0.22 (−0.34 to −0.09) | .001 |

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HR: heart rate.

[d]bpm: beat per minute.

## Discussion

### Principal Findings

This study demonstrated that the utility of a mobile app–based self-management DTx intervention for patients with CHD led to a significant increase in adherence to guideline-recommended medications as well as an increase in the proportion of patients with controlled LDL-C and blood pressure during the 12-month follow-up duration. The RR of adherence to all guideline-recommended medications at 12 months between the 2 groups was 34%, which might be mainly attributed to the dual antiplatelet and statin therapy, which is the fundamental secondary prevention treatment for CHD. Furthermore, the effect did not show an interaction with other variables based on the subgroup analysis.

During the trial, the adherence to all guideline-recommended medications showed a decreasing trend with time, regardless of whether an intervention was provided or not. The adherence was lower at the 12-month follow-up than at the 6-month follow-up in both the intervention and control groups; however, the intervention group showed a slower decreasing trend of adherence to guideline-recommended medications at 12 months. The same trend can also be seen in the changes in the values of the risk factors, such as blood pressure and LDL-C, which showed a decreasing trend during the first 3 months in both groups. However, the blood pressure in the control group began to increase thereafter, and SBP almost returned to the baseline values at 12 months, whereas that of the intervention group remained well controlled.

Recently, DTx interventions have arisen as a potential means of modifying health behaviors. A topical study reports the findings of a pivotal trial investigating the efficacy and safety of a self-management DTx (a 12-week intervention followed by a 12-week follow-up) in Japanese patients with untreated essential hypertension (baseline office and ambulatory 24 hours BP ≥140/90 mm Hg and ≥130/80 mm Hg, respectively) [14]. Recent studies have evaluated the effectiveness of SMS text messaging services on the improvement in LDL-C level, blood pressure, BMI, and smoking status among patients with CHD [15]. However, a recent systemic review demonstrated insufficient evidence to draw conclusions on their effectiveness [22], and sufficiently powered, high-quality randomized trials are needed, particularly in developing countries. Although SMS text messaging is simple and less costly, the function is somehow limited, especially for important features, such as patient education, vital signs' monitoring, and medication or adverse event alert, which are commonly applied to smartphones. Studies on mobile app–based management on secondary prevention for cardiovascular diseases have been reported [18,23], but most of the studies had a short follow-up duration, and only 1 RCT included in a recent meta-analysis had an intervention or follow-up period of 12 months [18]. Given the chronic nature of CHD, the long-term effectiveness and clinical outcomes of these mobile app–based interventions remain to be determined because of the lack of long-term follow-up.

The findings of this pilot study are consistent with those of previous studies. However, our study greatly differed from the previous studies based on the following reasons: our study targeted multiple risk factors rather than individual risk factors; the sample size was larger, and the follow-up duration was longer. Another important difference is that we focused on the trend of the improvements in patients, which resulted from the app of the intervention during the 12-month period. Furthermore, a feasibility study was conducted before the formal trial to ensure that the perceptions and acceptance of mHealth are sufficient in patients with CHD [24]. Given that China is the largest developing country, which is the main battlefield of the future management of noncommunicable diseases including CHD, the positive result of this pilot trial will have more influence on the future health care system and delivery of management for chronic diseases.

The current trial had several limitations. First, this trial was conducted at a single large tertiary hospital in Beijing; thus, it is unclear whether the observed benefits are generalizable. However, a routine traditional follow-up program was already established for discharged patients with CHD at this center, indicating that this mobile app will bring greater benefits in rural areas and more remote communities, where traditional secondary prevention programs are more difficult to access. Second, this study was open labeled because of the difficulty of blind design, and the contact of the nursing staff during recruitment and randomization could influence the adherence and introduce bias. However, the control group was also under a routine traditional face-to-face follow-up program including physicians and nurses, and the frequency of contact between the 2 groups was the same, which could minimize the bias. Third, given that this work is a pilot study, medication adherence, rather than solid outcome, was used as the primary end point, which will be confirmed in future studies with a long-term follow-up duration. Third, further work is needed to evaluate the extent to which this mobile app–based self-management intervention may be useful for patients with CHD to improve their prognosis.

## Conclusions

Among patients with CHD, the use of a mobile app–based self-management intervention in addition to the traditional follow-up care resulted in a significant improvement in the guideline-recommended medication adherence at 12 months, compared with the regular care. The results of the trial will be applicable to primary care centers in China, especially in rural areas with less medical resources. The trial will provide new evidence of the efficacy of an internet-based service model in the secondary prevention management of CHD and may help improve risk factor control.

## Acknowledgments

The authors would like to thank Lifen Yu, Hongyu Chen, and Juan Liu for their efforts in data collection; Jialin Zhu, Fan Yang, and Sheng Xu for data management; and Xiaowei Sun and Xiaojin Zhu for their support in the electronic health record system of this trial.

## Authors' Contributions

J Jiang and YH contributed equally as co-corresponding authors on this manuscript.

## Conflicts of Interest

QZ is an employee of Philips Research.

Multimedia Appendix 1
Sensitivity analysis and subgroup analysis.
[DOCX File , 184 KB - mhealth_v10i2e32251_app1.docx ]

Multimedia Appendix 2

Illustrated description slides of the digital therapeutics.
[PPTX File , 1613 KB - mhealth_v10i2e32251_app2.pptx ]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 182 KB - mhealth_v10i2e32251_app3.pdf ]

## References


1. Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019 Sep;394(10204):1145-1158. [doi: 10.1016/s0140-6736(19)30427-1]
2. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, ESC Scientific Document Group. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 2020 Jan 14;41(3):407-477. [doi: 10.1093/eurheartj/ehz425] [Medline: 31504439]
3. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, American College of Cardiology Foundation. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. Circulation 2012 Dec 18;126(25):3097-3137. [doi: 10.1161/CIR.0b013e3182776f83] [Medline: 23166210]
4. Section of Interventional Cardiology of Chinese Society of Cardiology, Section of Atherosclerosis and Coronary Artery Disease of Chinese Society of Cardiology, Specialty Committee on Prevention and Treatment of Thrombosis of Chinese College of Cardiovascular Physicians. [Guideline on the diagnosis and treatment of stable coronary artery disease]. Zhonghua Xin Xue Guan Bing Za Zhi 2018 Sep 24;46(9):680-694. [doi: 10.3760/cma.j.issn.0253-3758.2018.09.004] [Medline: 30293374]
5. Zhao D, Liu J, Wang M, Zhang X, Zhou M. Epidemiology of cardiovascular disease in China: current features and implications. Nat Rev Cardiol 2019 Apr;16(4):203-212. [doi: 10.1038/s41569-018-0119-4] [Medline: 30467329]
6. Lv J, Yu C, Guo Y, Bian Z, Yang L, Chen Y, China Kadoorie Biobank Collaborative Group. Adherence to healthy lifestyle and cardiovascular diseases in the Chinese population. J Am Coll Cardiol 2017 Mar 07;69(9):1116-1125 [FREE Full text] [doi: 10.1016/j.jacc.2016.11.076] [Medline: 28254173]
7. Khatib R, Marshall K, Silcock J, Forrest C, Hall AS. Adherence to coronary artery disease secondary prevention medicines: exploring modifiable barriers. Open Heart 2019;6(2):e000997 [FREE Full text] [doi: 10.1136/openhrt-2018-000997] [Medline: 31354954]
8. Adherence to long-term therapies: evidence for action. World Health Organization. URL: https://who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2022-01-21]
9. Ho PM, Magid DJ, Shetterly SM, Olson KL, Maddox TM, Peterson PN, et al. Medication nonadherence is associated with a broad range of adverse outcomes in patients with coronary artery disease. Am Heart J 2008 Apr;155(4):772-779. [doi: 10.1016/j.ahj.2007.12.011] [Medline: 18371492]
10. Chisholm-Burns MA, Spivey CA. The 'cost' of medication nonadherence: consequences we cannot afford to accept. J Am Pharm Assoc (2003) 2012;52(6):823-826. [doi: 10.1331/JAPhA.2012.11088] [Medline: 23229971]
11. Report on Cardiovascular Diseases in China 2018. Beijing: Encyclopedia of China Publishing House; 2018.
12. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232]
13. Transforming global healthcare by advancing digital therapeutics. Digital Therapeutic Alliance. URL: https://dtxalliance.org [accessed 2022-01-21]
14. Kario K, Nomura A, Harada N, Okura A, Nakagawa K, Tanigawa T, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021 Oct 21;42(40):4111-4122 [FREE Full text] [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
15. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
16. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
17. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]

18. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 07;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
19. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 05;150(9):604-612 [FREE Full text] [doi: 10.7326/0003-4819-150-9-200905050-00006] [Medline: 19414839]
20. The R project for statistical computing. The R Foundation. URL: https://www.R-project.org/ [accessed 2022-01-21]
21. Deliver the insights that your stakeholders want with RStudio Connect. RStudio. URL: https://www.rstudio.com/ [accessed 2022-01-21]
22. Adler AJ, Martin N, Mariani J, Tajer CD, Owolabi OO, Free C, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4:CD011851 [FREE Full text] [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
23. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
24. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and acceptance of mHealth in patients with cardiovascular diseases: a cross-sectional study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]


## Abbreviations

**CHD:** coronary heart disease
**DBP:** diastolic blood pressure
**DTx:** digital therapeutics
**LDL-C:** low-density lipoprotein cholesterol
**mHealth:** mobile health
**RR:** relative risk
**SBP:** systolic blood pressure


*Edited by L Buis; submitted 19.07.21; peer-reviewed by Y Guo, A Hernández, A Salzwedel; comments to author 10.09.21; revised version received 11.10.21; accepted 10.12.21; published 15.02.22.*

*Please cite as:*
*Li Y, Gong Y, Zheng B, Fan F, Yi T, Zheng Y, He P, Fang J, Jia J, Zhu Q, Jiang J, Huo Y*
*Effects on Adherence to a Mobile App–Based Self-management Digital Therapeutics Among Patients With Coronary Heart Disease: Pilot Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2022;10(2):e32251*
*URL: https://mhealth.jmir.org/2022/2/e32251*
*doi:10.2196/32251*
*PMID:34906924*

Original Paper

# Effectiveness of a Step Counter Smartband and Midwife Counseling Intervention on Gestational Weight Gain and Physical Activity in Pregnant Women With Obesity (Pas and Pes Study): Randomized Controlled Trial

Elena Gonzalez-Plaza[1,2], MSc; Jordi Bellart[1,3], MSc, PhD; Ángela Arranz[1,2], MSc, PhD; Leila Luján-Barroso[2,4], MSc, PhD; Esther Crespo Mirasol[1,2], MSc, PhD; Gloria Seguranyes[2,5], MSc, PhD

[1]Maternal-Fetal Medicine Department at BCNatal, Clinic Hospital of Barcelona, Barcelona, Spain

[2]Department of Nursing, Public, Mental and Maternity and Child Health, School of Nursing, Faculty of Medicine and Health Science, University of Barcelona, Barcelona, Spain

[3]Department of Medicine, University of Barcelona, Barcelona, Spain

[4]Unit of Nutrition and Cancer, Cancer Epidemiology Research Programme, Bellvitge Biomedical Research Institute, Catalan Institute of Oncology, Barcelona, Spain

[5]Research Group on Sexual and Reproductive Health Care (GRASSIR), Barcelona, Spain

**Corresponding Author:**
Elena Gonzalez-Plaza, MSc
Maternal-Fetal Medicine Department at BCNatal
Clinic Hospital of Barcelona
Sabino de Arana, 1
Barcelona, 08028
Spain
Phone: 34 932275400 ext 7294
Email: eplaza@clinic.cat

## Abstract

**Background:** Women who are pregnant and have obesity and excessive gestational weight gain (GWG) present a higher risk of maternal and perinatal complications. The use of mobile apps and a wristband during pregnancy may contribute to promoting healthy lifestyles and, thus, improving maternal and neonatal health.

**Objective:** This study aims to evaluate the effectiveness of a complex digital health intervention, using a smartband and app with midwife counseling, on GWG and physical activity (PA) in women who are pregnant and have obesity and analyze its impact on maternal and perinatal outcomes. In addition, we aim to study the frequency of use, usability, and satisfaction with the mobile apps used by the women in the intervention group.

**Methods:** A parallel, 2-arm, randomized controlled trial was conducted. A total of 150 women who were pregnant and had obesity were included. The intervention group received a complex combined digital intervention. The intervention was delivered with a smartband (Mi Band 2) linked to the app Mi Fit to measure PA and the Hangouts app with the midwife to provide personal health information. The control group received usual care. The validated Spanish versions of the International Physical Activity Questionnaire–Short Form and the System Usability Scale were used. Satisfaction was measured on a 1- to 5-point Likert scale.

**Results:** We analyzed 120 women, of whom 30 (25%) were withdrawn because of the COVID-19 pandemic. The median GWG in the intervention group was 7.0 (IQR 4-11) kg versus 9.3 (IQR 5.9-13.3) kg in the control group ($P$=.04). The adjusted mean GWG per week was 0.5 (95% CI 0.4-0.6) kg per week in the control group and 0.3 (95% CI 0.3-0.4) kg per week in the intervention group ($df$=0.1, 95% CI −0.2 to 0.03; $P$=.008). During the 35 and 37 gestational weeks, women in the intervention group had higher mean PA than women in the control group (1980 metabolic equivalents of tasks–minutes per week vs 1386 metabolic equivalents of tasks–minutes per week, respectively; $P$=.01). No differences were observed between the study groups in the incidence of maternal and perinatal outcomes. In the intervention group, 61% (36/59) of the women who were pregnant used the smartband daily, and 75% (44/59) evaluated the usability of the Mi Fit app as excellent. All women in the intervention group used the Hangouts app at least once a week. The mean of the satisfaction scale with the health counseling app and midwife support was 4.8/5 (SD 0.6) points.

**Conclusions:** The use of a complex mobile health intervention was associated with adequate GWG, which was lower in the intervention group than in the control group. In addition, we observed that the intervention group had increases in PA. No differences were observed in maternal perinatal complications.

**Trial Registration:** ClinicalTrials.gov NCT03706872; https://www.clinicaltrials.gov/ct2/show/NCT03706872

***(JMIR Mhealth Uhealth 2022;10(2):e28886)*** doi:10.2196/28886

**KEYWORDS**

obesity; maternal; pregnancy; mHealth; mobile apps; telemedicine; telenursing; physical activity; gestational weight gain; lifestyle; mobile phone

## Introduction

### Background

Obesity during pregnancy is an increasingly prevalent public health problem in society. Prepregnancy obesity in Europe was estimated to be 7.8% to 25.6% [1]. It involves a greater risk of maternal and neonatal complications such as gestational diabetes, preeclampsia a pregnancy-induced hypertension disorders, a high rate of cesarean sections, fetal prematurity, macrosomy, and newborns large for gestational age (LGA) [2,3]. Women with excessive gestational weight gain (GWG) have a higher probability of presenting complications [4], which increases according to the class of obesity [5]. The *Institute of Medicine* (IOM) recommends a weight gain during pregnancy between 5 kg and 9 kg in women with obesity to minimize complications [6].

### Prior Work

Interventions promoting physical activity and healthy food habits in women who are pregnant have been effective in limiting GWG and have been associated with a reduction in diabetes, cesarean sections, and macrosomy [7,8]. However, these interventions have not demonstrated a reduction in maternal and neonatal complications in women who are pregnant and overweight and have prepregnancy obesity [9]. Furthermore, several studies in women who are pregnant and have obesity have described low adherence to the intervention [10]; thus, promoting healthy lifestyles using new information and communication technologies (ICTs) may be useful for health care professionals and are accessible to a larger population [11].

ICTs enhance self-control, self-evaluation, self-reinforcement, and personalized feedback through monitoring devices. Thus, ICT interventions based on social cognitive theory could be useful in promoting healthy habits in women who are pregnant [12].

Mobile health allows access to and receipt of health information, which may contribute to the promotion of healthy lifestyles and improvement of maternal and neonatal health [13].

In Spain, 97.7% of Spanish women aged between 16 and 54 years seek information on the internet on topics related to health, and 99.9% use their mobile telephone to do so [14]. According to a meta-analysis by Lau et al [15], 70% of women who are pregnant and overweight or have obesity consulted a webpage or used a mobile app to obtain information on adequate GWG. Furthermore, this meta-analysis reported a limiting GWG and self-reported increase in moderate physical activity during the postpartum period in the electronic-based lifestyle intervention group in women who are overweight or have obesity.

In addition, in the past years, wearable devices such as wristbands have emerged. In 2019, a total of 62.9 million wristband units were sold, with a trend toward a rise in sales being foreseen [16]. Wristbands help monitor different aspects of health habits, including monitoring of physical activity. One of the most commonly used wristbands is the smartband. These devices incorporate gamification functions; for example, challenges and prizes that increase commitment to digital health interventions [13].

There is emerging evidence that pedometer interventions may be successful in increasing activity levels in women who are pregnant and have obesity.

A recent intervention study on the feasibility of using a pedometer in women who are pregnant and have prepregnancy obesity reported promising results in GWG and increased physical activity [17]. Despite the growing number of women who are pregnant consulting the internet or using smartbands during pregnancy, few studies have analyzed the impact of their use in women who are pregnant and have prepregnancy obesity [15].

### Objective

The aim of the study is to evaluate the effectiveness of a complex digital health intervention, using a smartband and app with midwife counseling, on GWG and physical activity in women who are pregnant and have obesity. The secondary objectives are to assess the impact of these interventions on maternal and perinatal outcomes and identify the frequency of use, usability, and satisfaction with the mobile apps used by the women in the intervention group.

## Methods

### Study Design

This randomized parallel controlled trial (Pas and Pes; from Catalan language, *weight* and *step*) with 2 arms in a 1:1 (intervention and control group) ratio was conducted at the maternal–fetal department of the Hospital Clinic of Barcelona from June 2018 to October 2020. The trial was registered on the Clinical Trial Register of the National Library of Medicine of the United States (NCT03706872).

## Recruitment

Eligible participants were women who were pregnant and had prepregnancy obesity (BMI ≥30 kg/m$^2$ based on World Health Organization classification [18]) who attended hospital obstetric clinics during prenatal care.

Women who were pregnant and had prepregnancy BMI ≥30 kg/m$^2$ at 12 to 18 weeks of pregnancy, singleton pregnancy, aged ≥18 years, users of an Android smartphone or iPhone (iOS) with an internet connection, and who agreed to participate were included in the study.

The exclusion criteria were women who were pregnant who had already used an app for monitoring physical activity and weight. Women with a previous diagnosis of psychiatric disorders, endocrine–metabolic disorders, or chronic hypertension; pregnant women with a contraindication for performing exercise or mobility problems that do not allow moderate walking; and women with language difficulties in understanding Spanish were also excluded.

All women were recruited by midwives.

Women who were pregnant and attending hospital obstetric clinics who fulfilled the inclusion criteria were consecutively included in the study.

All participants provided written informed consent before being fully enrolled in the study.

## Sample Size

The sample size calculation was based on the variable of weight gain to detect a difference ≥3.4 (SD 7.1) kg [19]. An α risk of .05 and a β risk of .2 were accepted in the bilateral contrast. It was calculated that 81 women were needed in the intervention group, and 81 women were needed in the control group. A loss to follow-up of 20% was estimated.

## Randomization

Randomization was computer based. Two random number lists were created by the University of Barcelona, and opaque numbered envelopes were prepared to mask the group assignment.

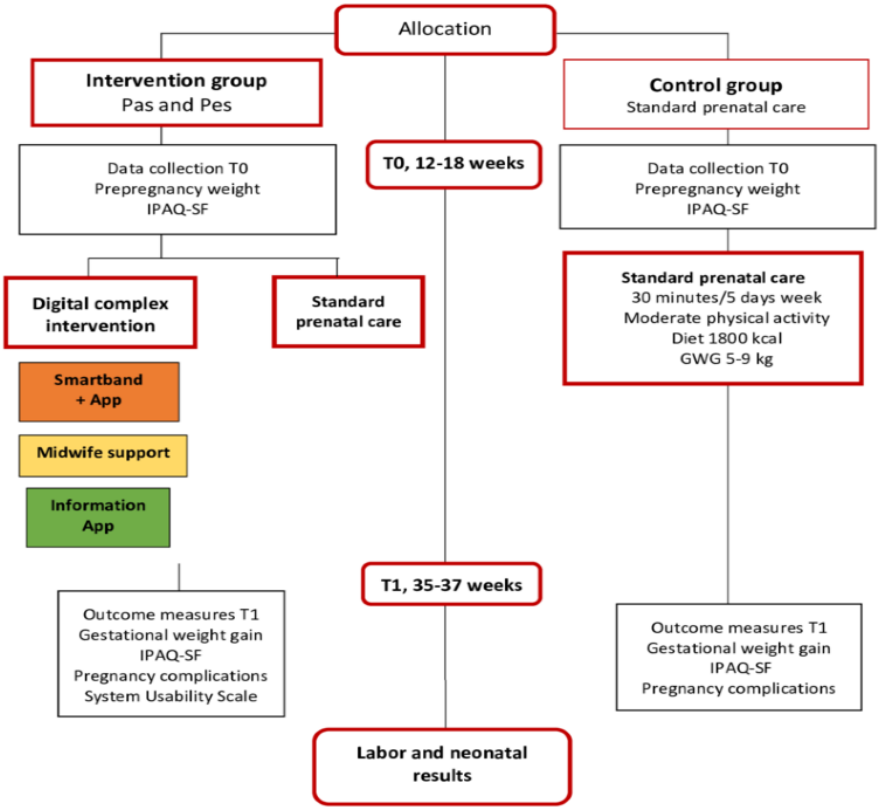
After the study participant had been informed about the study, and they accepted and signed the informed consent, the midwife opened the opaque and sealed envelope, and the woman who was pregnant was assigned to either the intervention or control group.

## Intervention

### *Usual Prenatal Care in the Control and Intervention Groups*

All the study participants received the standard prenatal care by midwives and obstetricians according to the *Pregnancy Monitoring Protocol* in Catalonia [20], which also includes health education in relation to physical activity GWG and food habits (Figure 1).

**Figure 1.** Flowchart of the Pas and Pes study. GWG: gestational weight gain; IPAQ-SF: International Physical Activity Questionnaire–Short Form; T0: time 0; T1: time 1.



### *Characteristics of the Intervention in the Intervention Group*

A complex digital intervention, based on social cognitive theory [12], was performed as a behavior-changing strategy of self-control, self-efficacy, and improvement of outcome expectations and to address barriers to the use of a smartband and an app for receiving information and support from a midwife.

### *Smartband (Mi Band 2) and Mi Fit App*

After participants were assigned to the intervention group, the midwife gave the participants a smartband (Mi Band 2) and explained that it should be worn during the day. Women who were pregnant were recommended to take 10,000 steps a day, equivalent to 30 minutes per day of moderate physical activity [21], over the week (≥5 days) according to the recommendations of the American College of Obstetricians and Gynecologists [22]. The smartband was linked to the Mi Fit app, which was free and available for Android and iOS systems. The midwife instructed the intervention group's participants on how to set up the step and weight goals through notifications of goals and activated alerts in the Mi Fit app. The smartband would vibrate during prolonged periods of inactivity or send prizes when goals were achieved. Women verified objective fulfillment by alerts and notifications from the Mi Fit app and the smartband (Mi Band 2).

### *Hangouts App*

The app for receiving health counseling and support from a midwife was Hangouts (Google LLC).

If necessary, the midwife instructed the women on how to download the app, which was free and available for Android and iOS systems (Hangouts), so that pregnant women could receive personalized information through SMS text messages or videos sent by the research team twice a week. One message corresponded to information regarding the physiological changes in the mother and fetus, and another was related to healthy eating habits, weight gain, physical activity, and information related to pregnancy, labor, and postpartum. The messages were personalized according to the gestational week and could contain videos (Multimedia Appendix 1). The source of information of the sent messages was extracted from a specialized webpage and the Inatal app. This specialized webpage is a social web designed by gynecologists and midwives from the Hospital Clinic, Barcelona. Permission was obtained for using its content. Video links for promoting physical activity and healthy eating habits were used from the webpage of the *Health Department of Catalonia*. Finally, we used videos and informative material from the *Catalan Midwives Association* available on their website. The women were to use the Hangouts app at least once a week.

In addition, the midwife asked the women who were pregnant about their current weight and motivated or reinforced their progress monthly (one by one woman) through the Hangouts app. Furthermore, women who were pregnant could ask questions to the midwife that were solved with an immediate response (<1 hour). No information regarding the data or results in the clinical history of the woman was delivered through Hangouts app (Figure 1).

### *Characteristic of the Intervention in Control and Intervention Group*

Women who were pregnant in the control group received oral information and written support material. With respect to physical activity, it was recommended to perform 30 min/day of moderate physical activity over the week (≥5 days) according to the recommendations of the American College of Obstetricians and Gynecologists [22]. Midwives gave instructions to gradually achieve the goal in those who were inactive or sedentary. Furthermore, midwives recommended a GWG between 5 kg and 9 kg to women who were pregnant, according to the IOM [6], and a balanced (Mediterranean) diet of 1800 kcal.

## Outcomes and Data Collection

### *Main Outcome Variables*

The main outcome variables were GWG and total physical activity. GWG was obtained by the difference between the weight of the woman who was pregnant between weeks 35 and 37 of pregnancy or time 1 (T1) and self-reported prepregnancy weight at the time of recruitment or time 0 (T0) and the mean GWG adjusted by the week of pregnancy in the study. The midwife weighed the dressed and shoeless woman who was pregnant in the midwife consultation using a Seca 704 scale. GWG was categorized according to the IOM recommendations as below (<5 kg), within (5-9 kg), and above the guidelines (>9 kg) [6].

Total physical activity was calculated using the global score of the *International Physical Activity Questionnaire–Short Form* [23,24], which participants self-reported at T0 and T1. The volume of physical activity was determined using metabolic equivalent of task (MET) units [25] and was calculated as METs-minutes per week. In addition, information on the types of physical activity was obtained: vigorous, moderate, and walking. Total physical activity was obtained by category (category 1 or low [≤600 METs-minutes per week], category 2 or moderate [600-3000 METs-minutes per week], and category 3 or high [>3000 METs-minutes per week]) and sitting time (minutes per week).

### *Secondary Outcome Variables*

Secondary outcomes variables were as follows:

- The incidence of maternal complications was a miscarriage at 22 weeks, gestational diabetes according to the diagnostic criteria of the International Association of Diabetes and Pregnancy Study Groups [26], preeclampsia or pregnancy-induced hypertensive disorder [27], and prematurity. The variable pregnancy composite morbidity (yes or no) was created, where it was coded as yes if the woman who was pregnant presented at least one of the four adverse results found during pregnancy.
- Incidence of birth induction, type of delivery, and unplanned cesarean section.
- Incidence of perinatal complications was macrosomy (weight >4000 g), low birth weight (weight <2500 g), small

for gestational age (percentile <10) and LGA (percentile >90) adjusted for newborn sex, postterm newborn, and neonatal death. The variable perinatal composite morbidity (yes or no) was created, which included the 7 adverse perinatal results mentioned above together with the variable of prematurity. Admissions to the neonatal intensive care unit were monitored.

Although the same woman or newborn might have had ≥1 adverse outcome during the process, the pregnancy composite morbidity and perinatal composite morbidity variables only counted as one event. Data on pregnancy complications were obtained by the midwives at T1. Data on delivery and newborns were retrospectively obtained by the research team through electronic clinical history.

Secondary specific outcome measures for the intervention group were frequency of smartband use and grade of usability of the Mi Fit app according to the total score and by categories (excellent, good, poor, and awful) of the System Usability Scale [28]. The satisfaction of women who received health counseling and midwife support through the app was evaluated with 6 questions answered using a 1- to 5-point Likert scale in which 1=not at all satisfied and 5=very satisfied. Self-reported questionnaires were answered anonymously by the women who were pregnant.

### Statistical Analysis

The analyses were performed on an intention-to-treat basis according to the treatment group allocated at randomization.

Descriptive data were presented as numbers and percentages, means and SDs, and medians and IQR. Bivariate analysis was performed between sociodemographic variables and prepregnancy BMI. For comparison of categorical variables, the nonparametric test of chi-square or Fisher exact (in case of small sample size of compared groups and expected frequency <5) and McNemar test were used. To compare quantitative variables, parametric Student *t* test (2-tailed) and nonparametric Mann–Whitney *U* or Wilcoxon tests were performed, depending on the normality distribution of compared groups.

Multinomial logistic regression was used to analyze the association between total physical activity (low, moderate, and high) at the end of the study, age and BMI at recruitment, previous births (yes or no), and test group (control and intervention). Adjusted odds ratios and 95% CIs were calculated for each model.

To evaluate the effect of the intervention on GWG per week (kg per week) of the participants at the end of the study, a linear regression model was used, which was adjusted for age and BMI at recruitment, previous births (yes or no), and total physical activity (low, moderate, and high) at the end of the study. The adjusted GWG per week was derived from this model.

All statistical tests were 2-sided and evaluated at an α level of .05. Analyses were performed using SPSS (version 25) and SAS (version 9.4).

### Ethical Aspects

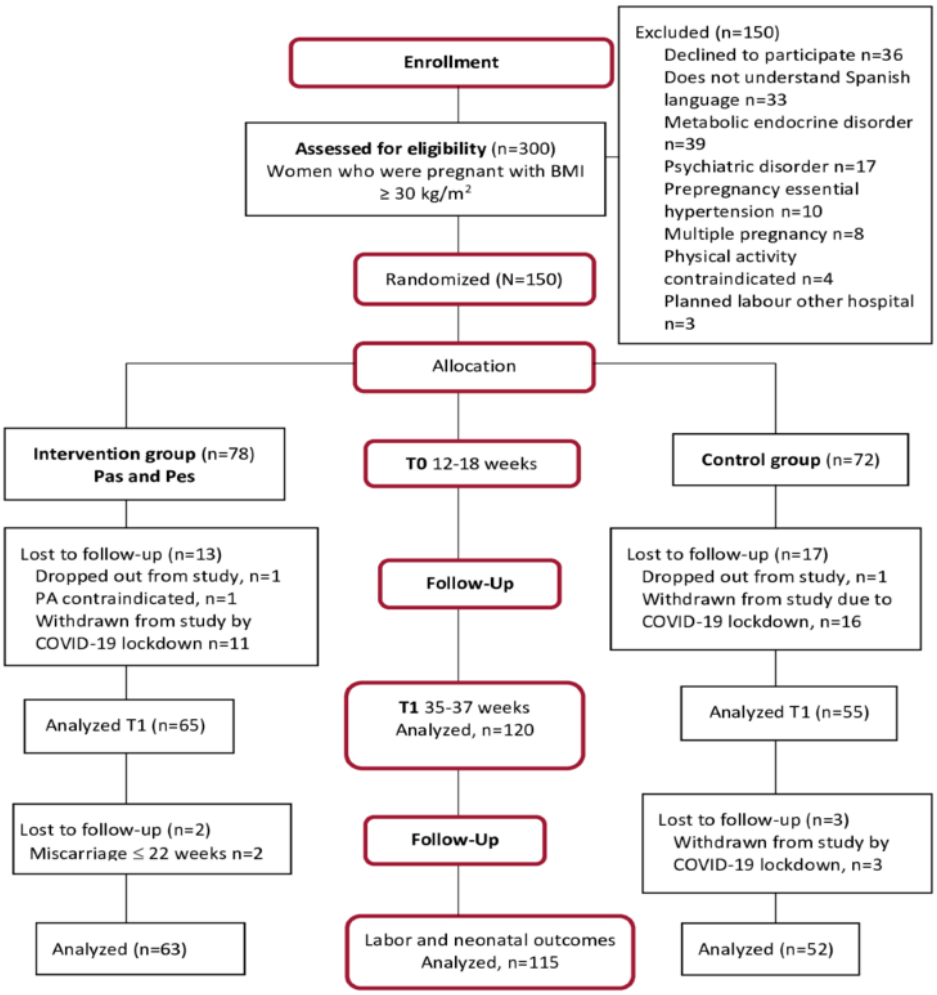
The study was approved by the ethics and clinical research committee of the Hospital Clinic of Barcelona (code HCB2017-0756). The anonymity and confidentiality of the data were always preserved in accordance with the Spanish Organic Law 3/2018, of December 5, on the Protection of Personal Data and guarantee of digital rights. Informed consent was obtained from all the participants.

## Results

### Participants

Figure 2 shows the flow diagram of the recruitment of study participants according to the recommendations of the CONSORT (Consolidated Standards of Reporting Trials) statement. Of the 300 women evaluated for recruitment, 150 (50%) fulfilling the inclusion criteria were randomized: 52% (78/150) in the intervention group and 48% (72/150) in the control group.

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) flowchart of participants in the Pas and Pes study. PA: physical activity; T0: time 0; T1: time 1.




The COVID-19 pandemic in Spain led to strict home confinement, which interfered with the physical activity of women. As prenatal care was delivered only by telematic means on April 1, 2020, up to 20% (30/150) of women who had not reached 35 weeks of pregnancy were withdrawn from the study. At T1, of the 150 women, 120 (80%) were analyzed, and variables related to delivery and the neonates of 115 (76.7%) women were analyzed (Figure 2).

The baseline characteristics of the study participants are shown in Table 1. There were no statistically significant differences in age, country of origin, number of previous births, or prepregnancy BMI between the 2 groups. The mean follow-up was 21.5 (SD 3.2) weeks in the intervention group and 21.1 (SD 2.4) weeks in the control group ($P$=.48).

**Table 1.** Baseline demographics and clinical characteristics by treatment group (N=150).

| Variables | Intervention group (n=78) | Control group (n=72) | *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 32.4 (5.4) | 33.4 (4.7) | .36[a] |
| **Country of origin, n (%)** | | | .48[b] |
| Spanish | 40 (51) | 41 (57) | |
| Foreign | 38 (49) | 31 (43) | |
| **Educational level, n (%)** | | | .83[b] |
| Primary | 8 (10) | 7 (10) | |
| Secondary | 37 (47) | 31 (43) | |
| Higher | 33 (42) | 34 (47) | |
| **Employed, n (%)** | | | .82[b] |
| Yes | 65 (83) | 59 (82) | |
| No | 13 (17) | 13 (18) | |
| **Cohabiting partner, n (%)** | | | .09[b] |
| Yes | 66 (85) | 53 (74) | |
| No | 12 (15) | 19 (26) | |
| **Prepregnancy weight (kg)** | | | .34[a] |
| Values, mean (SD) | 86.1 (10.4) | 84.3 (9.9) | |
| Values, median (IQR) | 84 (79.7-92.3) | 84 (77-90) | |
| **Prepregnancy BMI (kg/m$^2$)** | | | .06[a] |
| Values, mean (SD) | 33.1 (2.9) | 32.7 (3.3) | |
| Values, median (IQR) | 32.6 (31.1-34.2) | 31.3 (30.4-33.6) | |
| **Obesity class, n (%)** | | | .55[b] |
| Class I (30-34.9 kg/m$^2$) | 63 (81) | 61 (85) | |
| Class II (35-39.9 kg/m$^2$) | 13 (17) | 8 (11) | |
| Class III (≥40 kg/m$^2$) | 2 (2) | 3 (4) | |
| **Previous births, n (%)** | | | .06[b] |
| Yes | 36 (46) | 44 (61) | |
| No | 42 (54) | 28 (39) | |
| **Smoking, n (%)** | | | .29[b] |
| Yes | 8 (10) | 4 (6) | |
| No | 70 (90) | 68 (94) | |

[a]Mann–Whitney *U* test.

[b]Chi-square test.

## Main Outcomes

### *GWG Outcome*

The intervention group median of GWG 7.0 (IQR 4-11) kg was statistically significantly lower than the control group median of 9.3 (IQR 5.9-13.3 kg; *P*=.04). At T1, the median GWG per week was 0.3 in the intervention group versus 0.4 in the control group (*P*=.01).

An inverse association was observed between the GWG (kg per week) at the end of the study in the intervention group compared with the control group (β=−.1, 95% CI −0.2 to −0.03) at the same levels of age, BMI at recruitment, physical activity, and previous births (Multimedia Appendix 2). Derived from the model, we obtained an adjusted mean weight gain per week that was 0.5 (95% CI 0.4-0.6) kg per week for the control group and 0.3 (95% CI 0.3-0.4) kg per week for the intervention group; (*df*=0.1, 95% CI −0.2 to 0.03; *P*=.008; Table 2).

**Table 2.** Gestational weight gain by study group (N=113)[a].

| Gestational weight gain | Intervention group (n=60) | Control group (n=53) | Mean difference (95% CI) | β | *P* value |
|---|---|---|---|---|---|
| **Continuous (kg)** | | | | | |
| Values, mean (SD) | 7.6 (5.5) | 10.1 (6.4) | 2.5 (0.2 to 4.7) | N/A[b] | .02[c] |
| Values, median (IQR) | 7.0 (4.0 to 11.0) | 9.3 (5.9 to 13.3) | N/A | N/A | .04[d] |
| **Weekly weight gain (kg)** | | | | | |
| Values, mean (SD) | 0.3 (0.3) | 0.4 (0.3) | 0.1 (0.03 to 0.2) | N/A | .01[e] |
| Values, median (IQR) | 0.3 (0.2 to 0.5) | 0.4 (0.3 to 0.6) | N/A | N/A | .01[d] |
| Adjusted mean | 0.3 | 0.5 | 0.1 (−0.2 to −0.03) | −.1 | .008[e] |
| **Categorical based on IOM[f] guidelines, n (%)** | | | N/A | N/A | .08[g] |
| Below guidelines | 18 (30) | 9 (17) | | | |
| Within guidelines | 21 (35) | 15 (28) | | | |
| Above guidelines | 21 (35) | 29 (55) | | | |

[a]N=113; missing data of 2 miscarriages and 5 premature deliveries at ≤35 weeks.

[b]N/A: not applicable.

[c]Student *t* test (2-tailed).

[d]Mann–Whitney *U* test.

[e]*P* value adjusted for age (years), BMI at time 0, and previous births (yes or no).

[f]IOM: Institute of Medicine.

[g]Chi-square test.

The proportion of women with adequate GWG according to IOM recommendations was 35% (21/60) in the intervention group versus 28% (15/53) in the control group; GWG below guidelines was 30% (18/60) in the intervention group versus 17% (9/53) in the control group; and GWG above guidelines was 35% (21/60) in the intervention group versus 55% (29/53) in the control group (*P*=.08; Table 2).

### *Physical Activity*

Regarding total physical activity, in intragroup comparison, women in the intervention group performed greater total physical activity at T1 than at T0 (1980 vs 990 METs-minutes per week; *P*=.001), whereas women in the control group did not modify their METs-minutes per week at T1 compared with T0 (*P*=.69). When we compared the 2 groups at T1 (intervention group vs control group), women in the intervention group had higher mean total physical activity than women in the control group (1980 METs-minutes per week vs 1386 METs-minutes per week; *P*=.01; Figure 3 and Table 3). Regarding sitting time, women in the intervention group obtained a lower mean of 1260 minutes per week than 2100 minutes per week in the control group (*P*=.02; Table 3).

**Figure 3.** Physical activity by study group at time 0 and time 1 (time 0=12-18 weeks; time 1=35-37 weeks). CG: control group; IG: intervention group; MET: metabolic equivalent of task; PA: physical activity.

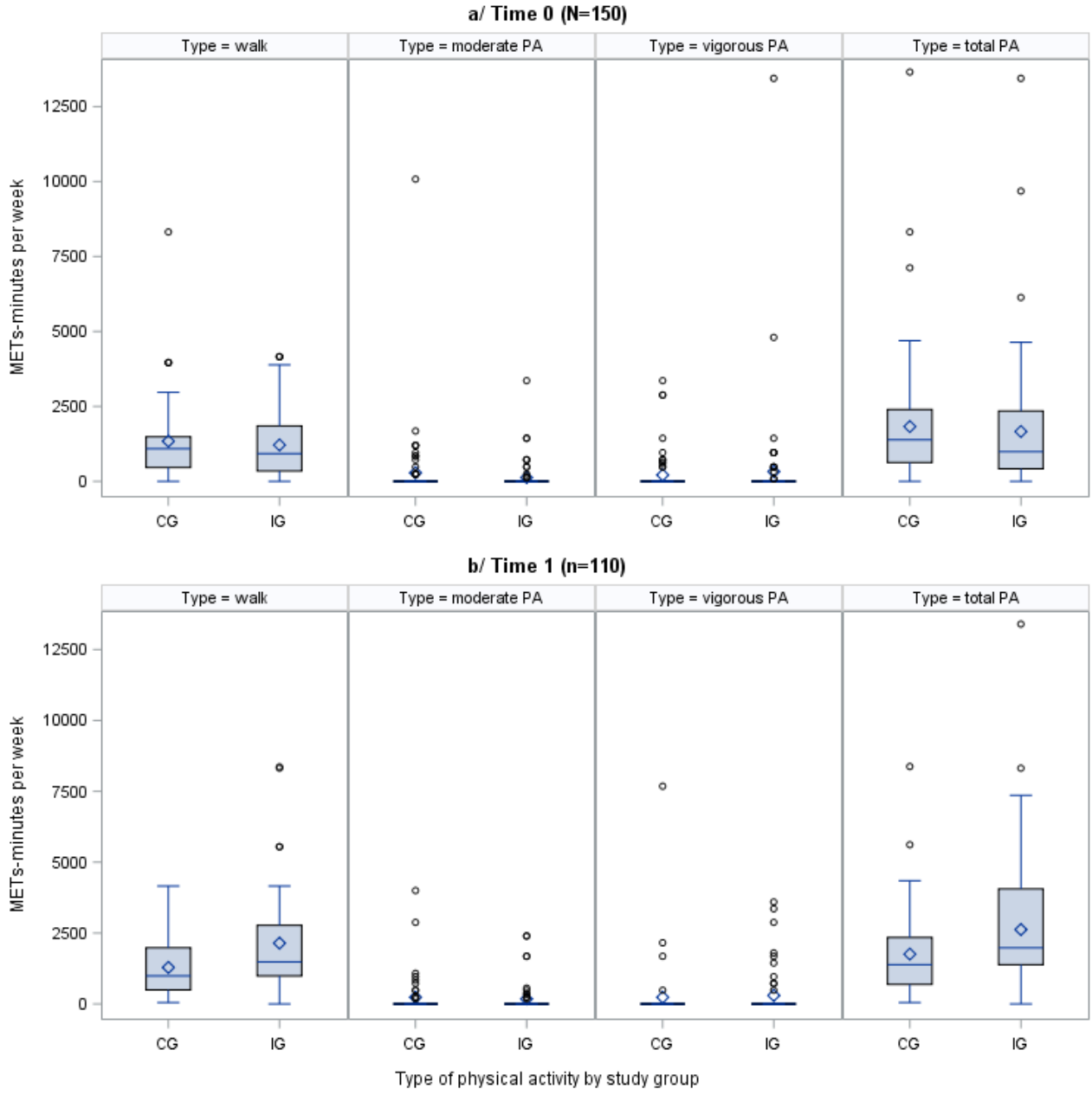


**Table 3.** Intragroup physical activity outcomes by period time 0 (T0) and time 1 (T1) and physical activity outcomes in T1 by study group (N=110)[a].

| Physical activity | Intervention group (n=59) | | Control group (n=51) | | *P* value |
|---|---|---|---|---|---|
| | T0 (12-18 weeks) | T1 (35-37 weeks) | T0 (12-18 weeks) | T1 (35-37 weeks) | |
| **Total** | | | | | |
| Values (METs[b]-minutes per week), median (IQR) | 990 (396-2376) | 1980 (1386-4060) | 1386 (495-2685) | 1386 (693-2346) | .01[c] |
| Values (METs-minutes per week), minimum-maximum | 0-13,400 | 0-13,400 | 0-8316 | 50-8373 | .01[c] |
| *P* value[d] (intragroup) | N/A[e] | <.001 | N/A | .69 | N/A |
| **Vigorous** | | | | | |
| Values (METs-minutes per week), median (IQR) | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-0) | .67[c] |
| Values (METs-minutes per week), minimum-maximum | 0-13,440 | 0-3600 | 0-3360 | 0-7680 | .67[c] |
| 0 METS-minute per week, n (%) | 47 (80) | 49 (83) | 42 (82) | 47 (92) | N/A |
| *P* value[d] (intragroup) | N/A | .83 | N/A | .23 | N/A |
| **Moderate** | | | | | |
| Values (METs-minutes per week), median (IQR) | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-0) | .16[c] |
| Values (METs-minutes per week), minimum-maximum | 0-3360 | 0-2400 | 0-1680 | 0-4000 | .16[c] |
| 0 METS-minutes per week, n (%) | 48 (81) | 40 (78) | 39 (76) | 48 (81) | N/A |
| *P* value[d] (intragroup) | N/A | .93 | N/A | .97 | N/A |
| **Walking** | | | | | |
| Values (METs-minutes per week), median (IQR) | 693 (330-1782) | 1485 (990-2772) | 693 (346.5-1980) | 990 (495-1980) | .003[c] |
| Values (METs-minutes per week), minimum-maximum | 0-4158 | 0-8360 | 0-8316 | 50-4158 | .003[c] |
| 0 METS-minutes per week, n (%) | 4 (7) | 1 (2) | 1 (2) | 0 (0) | N/A |
| *P* value[d] (intragroup) | N/A | <.001 | N/A | .55 | N/A |
| **Physical activity by category** | | | | | |
| Category I: low, n (%) | 22 (37) | 6 (10) | 15 (29) | 12 (3) | .10[f] |
| Category II: moderate, n (%) | 30 (51) | 36 (61) | 29 (57) | 30 (59) | .10[f] |
| Category III: high, n (%) | 7 (12) | 17 (29) | 7 (14) | 9 (18) | .10[f] |
| *P* value[g] (intragroup) | N/A | <.001 | N/A | .83 | N/A |
| **Sitting time (minutes per week)[h]** | | | | | |
| Values (METs-minutes per week), median (IQR) | 1680 (840-2940) | 1260 (420-2100) | 1680 (840-2940) | 2100 (1260-2520) | .02[c] |
| Values (METs-minutes per week), minimum-maximum | 0-5880 | 55-5460 | 20-5040 | 55-7560 | .02[c] |
| *P* value[g] (intragroup) | N/A | .16 | N/A | .81 | N/A |

[a]N=110; missing data of 2 miscarriages, 5 premature deliveries at ≤35 weeks, and 3 non–International Physical Activity Questionnaire–Short Form data.

[b]MET: metabolic equivalent of task.

[c]Wilcoxon test.

[d]Mann–Whitney *U* test of the period time 1 data intervention group versus time 1 control group.

[e]N/A: not applicable.

[f]Chi-square test.

[g]McNemar test of the period time 1 data intervention group versus time 1 control group.

[h]In the intervention group, n=52 at time 0 and n=54 at time 1 and in the control group, n=47 at time 0 and n=45 at time 0.

Multimedia Appendix 3 shows results of the multinomial logistic regression for categorical physical activity. The probability of high versus low physical activity, with the other variables in the model remaining constant, was 3.9-fold higher in the intervention group (95% CI 1.1-14.3) than in the control group.

## Secondary Outcomes

### *Maternal and Perinatal Complications During Pregnancy and Delivery*

Pregnancy, labor, and perinatal complications by study group are detailed in Table 4.

**Table 4.** Pregnancy, labor, and perinatal complications by study group (N=120).

| Complications | Intervention group (n=65) | Control group (n=55) | *P* value |
|---|---|---|---|
| **Gestational complications, n (%)** | | | |
| Composite pregnancy morbidity | 22 (34) | 20 (36) | .77[a] |
| Miscarriage ≤22 weeks | 2 (3) | 0 (0) | .49[b] |
| Gestational diabetes[c] | 10 (15) | 12 (22) | .36[a] |
| Preeclampsia or gestational hypertension[c] | 6 (9) | 9 (16) | .23[a] |
| Preterm labor ≤37 weeks[d] | 7 (11) | 5 (9) | .17[a] |
| **Labor complications, n (%)[e,f]** | | | |
| **Type of labor onset** | | | .15[a] |
| Spontaneous | 23 (38) | 22 (44) | |
| Induction | 26 (43) | 25 (50) | |
| Planned cesarean | 11 (18) | 3 (6) | |
| **Type of labor** | | | |
| Vaginal | 37 (62) | 39 (78) | .06[a] |
| **Cesarean** | 23 (38) | 11 (22) | .06[a] |
| Planned | 11 (48) | 3 (27) | .29[b] |
| Unplanned | 12 (52) | 8 (73) | .29[b] |
| **Perinatal complications, n (%)[d,g]** | | | |
| Composite perinatal morbidity | 24 (38) | 24 (46) | .38[a] |
| Birthweight ≥4000 g | 4 (6) | 6 (12) | .34[b] |
| Birthweight ≤2500 g | 5 (8) | 2 (4) | .45[b] |
| **Large for gestational age centiles** | | | .87[b] |
| ≤5th | 6 (10) | 3 (6) | |
| 5-10th | 2 (3) | 1 (2) | |
| 10-90th | 43 (68) | 37 (71) | |
| ≥90th | 12 (19) | 11 (21) | |
| Postterm | 7 (11) | 8 (15) | .49[a] |
| **Perinatal death** | 1 (2) | 2 (4) | .58[b] |
| Early neonatal death | 1 (100) | 1 (50) | |
| Antepartum stillbirth | 0 (0) | 1 (50) | |
| Admission to NICU[h,i] | 4 (6) | 5 (10) | .72[b] |

[a]Chi-square test.

[b]Fisher exact test.

[c]N=118; missing data for 2 miscarriages.

[d]N=115; missing data for 2 miscarriages and 3 COVID-19 lockdowns in delivery.

[e]N=110, missing data of 2 miscarriages, 5 premature deliveries at ≤35 weeks, and 3 COVID-19 lockdowns in delivery.

[f]n=60 for the intervention group and n=50 for the control group.

[g]n=63 for the intervention group and n=52 for the control group.

[h]NICU: neonatal intensive care unit.

[i]N=114; missing data for 2 miscarriages, 3 COVID-19 lockdowns in delivery, and 1 antepartum stillbirth.

### *Frequency of Using Mi Band 2 and Hangouts App, Grade of Usability Mi Fit App, and Grade of Satisfaction in Women in the Intervention Group (Hangouts) App*

Information was obtained from 91% (59/65) of the women in the intervention group at T1. None of the women showed adverse effects with the use of the smartband (Mi Band 2). The smartband was used daily by 61% (36/59) of the women. The mean System Usability Scale score of the app linked to the smartband (Mi Fit) was 89.7 (SD 14.9) points, and 75% (44/59) evaluated its use as excellent. All 59 women reported having consulted the information provided in the app (Hangouts). All these women used the Hangouts app at least once a week, and they received midwives' feedback once a month and every time they formulated questions. The mean grade of overall satisfaction with receiving messages related to pregnancy and health counseling and midwife support through the app was 4.8/5 (SD 0.6) points (Table 5).

**Table 5.** Usability score of the Mi Fit app, frequency of Mi Band 2 and Hangouts app use, and grade of satisfaction with Hangouts app in the intervention group (N=59)[a].

| Measures | Values |
|---|---|
| Mi Fit app: usability score (SUS[b]), mean (SD) | 89.7 (14.9) |
| **Mi Fit app: usability score (SUS)—categorical, n (%)** | |
| Excellent (≥80.3) | 44 (75) |
| Good (68 to 80.3) | 8 (14) |
| Poor (51-67) | 6 (10) |
| Awful (≤51) | 1 (2) |
| **Frequency of smartband use (Mi Band 2), n (%)** | |
| Daily | 36 (61) |
| 3-4 times per week | 11 (19) |
| 2 times per week | 6 (10) |
| 1 time per week | 5 (10) |
| Never | 0 (0) |
| **Satisfaction with app information (Hangouts), mean (SD)** | |
| Utility of pregnancy advice | 4.6 (0.6) |
| Utility of healthy lifestyles advice | 4.6 (0.6) |
| **Satisfaction with midwife support by (Hangouts) app, mean (SD)** | |
| Midwife accessibility | 4.7 (0.7) |
| Ease of use of the chat | 4.7 (0.8) |
| Be able to take advice without having to scroll | 4.7 (0.6) |
| Global satisfaction, mean (SD) | 4.8 (0.6) |

[a]The grade of satisfaction was analyzed with a Likert scale in which the minimum grade of satisfaction was 1 point and maximum 5 points; N=59; missing data of 2 miscarriages at 22 weeks, 3 premature deliveries at 35 weeks, and 1 no data of System Usability Scale and of the satisfaction questionnaire scale.

[b]SUS: System Usability Scale.

## Discussion

### Principal Findings

Our findings show that the use of a complex digital intervention was associated with lower GWG and an increase in physical activity during pregnancy in women who were pregnant and had obesity. No differences in the incidence of maternal and perinatal complications between the 2 study groups were found. All women in the intervention group used the smartband and health counseling app at least once a week. In addition, the usability of the app linked to the smartband was evaluated as excellent, and the grade of overall satisfaction with the health counseling app and support by the midwife was very high.

### Relation to Prior Literature

Recent research has suggested that interventions promoting healthy lifestyles and self-control using social networks of mobile apps in women who are pregnant have a moderate or low effect on maternal weight control [11,13]. Moreover, interventions accompanied by the use of self-monitoring devices [11] or those combined with professional reinforcement [13] are more effective for weight management.

In relation to GWG, our findings showed a mean difference in weight gain of 2.5 kg between the 2 groups, being lower in the intervention group. This GWG was lower than that reported in previous randomized controlled trials (RCTs) performed in women who were pregnant and had prepregnancy obesity. The

intervention in those studies was using new technologies independently or combined with professional reinforcement through the sending of SMS text messages [29], social networks [30], telephone reinforcement [31], or with pedometers and telephone calls [32-34].

Pollak et al [35], based on n=34 (22 women in the chat group and 12 women in the SMS text messaging group) women who were pregnant and had obesity, provided health counseling for the management of GWG through SMS text messaging and a chat with professionals and observed a difference of 2.7 kg, which was similar to the GWG observed in our study. However, a recent RCT that included n=30 (10 per group for control, app, and app-coach) women who were pregnant and had obesity achieved a difference of 5.3 kg between the women who were pregnant and used a smartwatch linked to an app and the women who were pregnant and underwent an in-person coaching intervention or the group that used a smartwatch [17].

Similar to other studies, we observed that the proportion of excessive GWG of the women who were pregnant in the intervention group was lower than that of the control group [30,33]. In addition, we found a higher proportion of women with GWG <5 kg according to IOM [31,33].

Regarding physical activity, we showed that women who were pregnant and used the smartband and the app were more active, similar to the studies of Renault et al [33] and Poston et al [32]. Furthermore, our study and the study by Simmons et al [31] observed that women in the intervention group also spent less time sitting than women in the control group. In addition, we found that women in the intervention group increased their physical activity at T1 compared with T0, which was derived from the increase in physical activity by walking, as described by Darvall et al [17]. We observed 4 women with vigorous or high moderate physical activity, 3 (75%) of whom were derived from occupational physical activity and low walking physical activity at T0. At T1, those women decreased their occupational physical activity (probably because of increased onset of the usual symptoms of pregnancy between 35 and 37 weeks) and increased their walking physical activity.

Our results are in line with those based on a systematic review by Hussain et al [36], where an intervention combining several technological resources, such as the smartband with a reinforcement app with information and support from a midwife, was associated with better results in weight gain and physical activity during pregnancy.

With respect to maternal and perinatal complications, no differences were observed between the 2 groups, as in the previously mentioned RCT, although there was a trend toward presenting a lower incidence of gestational complications. As in other studies, there was a lower incidence of gestational diabetes [33,37], preeclampsia [33,37], macrosomy [33] and LGA [29,32], postterm newborns, and lower admission to the neonatal intensive care unit [30] in the intervention group than in the control group. However, there was a higher incidence of prematurity [29,30], small for gestational age or restricted intrauterine growth [29,32], and low newborn birth weight, in contrast to what was described by Poston et al [32].

Contrary to other studies [30,32,33], we observed a greater proportion of cesarean sections in the intervention group, similar to the findings of Okesene-Gafa et al [29]. However, there was a lower incidence of unplanned cesarean sections in the intervention group than in the control group, probably because of the higher incidence of planned cesarean sections in the intervention group.

In our study, the frequency of the use of the smartband and the linked app was very high, as all women who were pregnant in the intervention group used them at least once a week, similar to the study by Baruth et al [38]. This finding contrasts with the low adherence reported in the UK Pregnancies Better Eating and Activity Trial [32,39] or the RCT of Ainscough et al [40] and Szmeja et al [41] in women who were pregnant and overweight and had obesity.

The usability of the Mi Fit app linked to the smartband in our study was evaluated as excellent, as in the RCT Fit4two [42]. We observed that satisfaction with the messages and midwife support through the app was very high, and the acceptability of the intervention agreed with other RCTs in women who were pregnant and overweight and had obesity, such as SMARTMOMS [43], txt4two [44], or studies with women who were pregnant and had any BMI, such as RCT Interact [45] and the RCTs of Choi et al [46] and Coughlin et al [47].

Taking all of this into account, the use of a smartband and providing information and the support by a midwife through an app could be recommended to promote physical activity and adequate weight gain in the prenatal control of women who are pregnant and have obesity. It would also be useful to provide evidence-based information and solve doubts from a distance as health professionals have described difficulties in the management of GWG in women who are pregnant and have obesity and a lack of time in the consultation [48]. In addition, telematics access provides the opportunity for professionals to gain access to a greater population, even to women who are pregnant and who less frequently attend health care centers. Finally, providing information through apps increases quality and safety in the care of women during pregnancy and contributes to reducing the heterogeneity of information regarding health and pregnancy that women who are pregnant see on the internet [49].

## Strengths and Limitations

The strengths of this study include the ability to evaluate the effectiveness of a complex digital intervention by the use of a wearable device and apps in a clinical study, taking into account the increasing use of these devices worldwide in women with prepregnancy obesity. Randomized assignment to the intervention reduced the probability of selection bias and ensured that the study groups were homogeneous. Furthermore, this study provides information related to the usability of the app linked to the smartband using a validated questionnaire widely used by the scientific community. Similarly, we describe information on the frequency of use and satisfaction with the app with which the women who were pregnant received information and could consult midwives regarding doubts.

The main limitation of this study is that the estimated sample size could not be achieved because of the COVID-19 pandemic. Approximately 20% (30/150) of women who were pregnant included in our study were confined at home during the first wave of the pandemic (from March 14, 2020) and had to be withdrawn from the analysis of the study as we considered that this could influence the results, as the power of the analysis reduced to 63%. Nonetheless, these women continued in the study, and the results obtained are pending publication. The reduced sample size may have contributed to the lack of statistically significant differences in the trend of presenting less gestational diabetes, preeclampsia, and macrosomy observed in the intervention group, and, in turn, the remaining observed findings could not demonstrate a size effect because of limited statistical power. However, multinominal models were performed to adjust the effect of the intervention on the weight gain variables and physical activity by categories, showing that GWG was lower in the intervention group than in the control group and that there was a 4-fold higher probability of the intervention group performing physical activity than the control group.

The data collected by the app linked to the smartband in relation to the number of steps or physical activity performed by the women who were pregnant in the intervention group was not monitored as the objective of the study was to compare the physical activity between the 2 groups at T0 and T1. We used the validated self-reported International Physical Activity Questionnaire–Short Form questionnaire, which may have induced a memory bias with underestimated or overestimated reporting by the women [50,51]. Nonetheless, this questionnaire has been used in multiple studies evaluating physical activity in the population [24] and in the pregnant population who are overweight and have obesity [32,52].

Regarding to the Hangouts app questions that pregnant women asked the midwife through the app, we have not performed qualitative analyses.

Finally, we did not measure body composition in pregnancy with fat percentage and total body water using bioelectrical impedance analysis. The clinical utility of body composition measurements in pregnancy is an ongoing future area of research.

## Conclusions

Our results suggest that the use of a complex mobile health intervention was associated with adequate GWG, which was lower in the intervention group than in the control group. In addition, we observed that the intervention group increased their physical walking activity, although it did not reduce maternal and perinatal complications compared with the control group. Furthermore, our findings provide some support for the effectiveness and safety of the use of a smartband and an app for providing health counseling and support from a midwife during pregnancy in women who are pregnant and have obesity, which could be applied to promote healthy lifestyles in prenatal control. The frequency of use; satisfaction with the smartband, health counseling app, and midwife support; and usability of the app linked to the smartband were satisfactorily evaluated.

The findings were obtained with a reduced sample size, and thus, the size effect should be interpreted with caution. Furthermore, clinical studies in larger samples of women who are pregnant and have prepregnancy obesity are necessary to evaluate the effectiveness and feasibility, if any, of the use of new technologies during pregnancy and their influence on maternal and perinatal health.

## Acknowledgments

This study was funded in part by the *Nurse and Society Foundation* as part of the Nurse Research Projects grants (PR 389/2019), Barcelona, Spain.

EGP received a research grant from *La Pedrera Foundation* (Nurse Intensification grant), Barcelona, Spain.

Permission was obtained from the pertinent companies to use the Mi Band 2 smartband and the app linked to the wristband, Mi Fit, and the Hangouts app to provide health advice and establish 24-hour communication with the midwife.

The authors are grateful to Mrs Mariuxi Burgos for reading this manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of SMS text messages and videos delivered to women who were pregnant through the app in the intervention group.
[DOCX File , 23 KB - mhealth_v10i2e28886_app1.docx ]

Multimedia Appendix 2
Linear model of gestational weight gain variables with the adjustment variables of age, BMI, physical activity, and study group.
[DOCX File , 24 KB - mhealth_v10i2e28886_app2.docx ]

Multimedia Appendix 3

Multinomial model of physical activity variable with the adjustment variables of age, BMI, physical activity, and study group.
[DOCX File , 23 KB - mhealth_v10i2e28886_app3.docx ]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 412 KB - mhealth_v10i2e28886_app4.pdf ]

## References


1. Euro-Peristat Project. European perinatal health report: core indicators of the health and care of pregnant women and babies in Europe in 2015 (Updated November 2018). Euro-Peristat. 2018. URL: http://www.europeristat.com [accessed 2020-12-31]
2. Dolin CD, Kominiarek MA. Pregnancy in women with obesity. Obstet Gynecol Clin North Am 2018;45(2):217-232. [doi: 10.1016/j.ogc.2018.01.005] [Medline: 29747727]
3. Catalano PM, Shankar K. Obesity and pregnancy: mechanisms of short term and long term adverse consequences for mother and child. BMJ 2017;356:j1. [doi: 10.1136/bmj.j1] [Medline: 28179267]
4. Goldstein RF, Abell SK, Ranasinha S, Misso ML, Boyle JA, Harrison CL, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med 2018;16(1):153 [FREE Full text] [doi: 10.1186/s12916-018-1128-1] [Medline: 30165842]
5. Thompson AM, Thompson JA. An evaluation of whether a gestational weight gain of 5 to 9 kg for obese women optimizes maternal and neonatal health risks. BMC Pregnancy Childbirth 2019;19(1):126 [FREE Full text] [doi: 10.1186/s12884-019-2273-z] [Medline: 30975086]
6. Institute of Medicine (US), National Research Council (US) Committee to Reexamine IOM Pregnancy Weight Guidelines. Composition and components of gestational weight gain: physiology and metabolism. In: Rasmussen KM, Yaktine AL, editors. Weight gain during pregnancy: reexamining the guidelines. Washington D.C: The National Academies Press (US); 2009:71-110.
7. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. Cochrane Database Syst Rev 2015;15(6):CD007145. [doi: 10.1002/14651858.CD007145.pub3] [Medline: 26068707]
8. International Weight Management in Pregnancy (i-WIP) Collaborative Group. Effect of diet and physical activity based interventions in pregnancy on gestational weight gain and pregnancy outcomes: meta-analysis of individual participant data from randomised trials. BMJ 2017;358:j3119 [FREE Full text] [doi: 10.1136/bmj.j3119] [Medline: 28724518]
9. Yeo S, Walker JS, Caughey MC, Ferraro AM, Asafu-Adjei JK. What characteristics of nutrition and physical activity interventions are key to effectively reducing weight gain in obese or overweight pregnant women? A systematic review and meta-analysis. Obes Rev 2017;18(4):385-399. [doi: 10.1111/obr.12511] [Medline: 28177566]
10. O'Brien OA, McCarthy M, Gibney ER, McAuliffe FM. Technology-supported dietary and lifestyle interventions in healthy pregnant women: a systematic review. Eur J Clin Nutr 2014;68(7):760-766. [doi: 10.1038/ejcn.2014.59] [Medline: 24781682]
11. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
12. Bandura A. Social foundations of thought and action: a social cognitive theory. 2nd edition. Engelwood Cliffs: Prentice-Hall; 1986.
13. Rhodes A, Smith AD, Chadwick P, Croker H, Llewellyn CH. Exclusively digital health interventions targeting diet, physical activity, and weight gain in pregnant women: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(7):e18255 [FREE Full text] [doi: 10.2196/18255] [Medline: 32673251]
14. European Health Survey in Spain (EHSS), 2014. National Statistics Institute. 2015. URL: https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176784&menu=resultados&idp=1254735573175 [accessed 2021-01-31]
15. Lau Y, Klainin-Yobas P, Htun TP, Wong SN, Tan KL, Ho-Lim ST, et al. Electronic-based lifestyle interventions in overweight or obese perinatal women: a systematic review and meta-analysis. Obes Rev 2017;18(9):1071-1087. [doi: 10.1111/obr.12557] [Medline: 28544551]
16. Vailshery LS. Forecast unit shipments of wrist-worn wearables worldwide from 2019 to 2024 (in million units). Statista. 2021. URL: https://www.statista.com [accessed 2021-01-31]
17. Darvall JN, Wang A, Nazeem MN, Harrison CL, Clarke L, Mendoza C, et al. A pedometer-guided physical activity intervention for obese pregnant women (the Fit MUM Study): randomized feasibility study. JMIR Mhealth Uhealth 2020;8(5):e15112 [FREE Full text] [doi: 10.2196/15112] [Medline: 32348280]
18. Obesity and overweight. World Health Organization. 2018. URL: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-01-31]
19. Bogaerts AF, Devlieger R, Nuyts E, Witters I, Gyselaers W, Van den Bergh BR. Effects of lifestyle intervention in obese pregnant women on gestational weight gain and mental health: a randomized controlled trial. Int J Obes (Lond) 2013;37(6):814-821. [doi: 10.1038/ijo.2012.162] [Medline: 23032404]

20. Pregnancy monitoring protocol in Catalonia. Catalan Public Health Agency. Barcelona (Spain): Catalan Government; 2018. URL: https://salutpublica.gencat.cat/ca/ambits/promocio_salut/Embaras-part-i-puerperi/Protocol-de-seguiment-de-lembaras/index.html#googtrans(ca|en) [accessed 2021-01-31]
21. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
22. American College of Obstetricians and Gynecologists. Committee opinion no. 650 summary: physical activity and exercise during pregnancy and the postpartum period. Obstet Gynecol 2015;126(6):1326-1327. [doi: 10.1097/AOG.0000000000001209] [Medline: 26595580]
23. Hagströmer M, Oja P, Sjöström M. The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr 2006;9(6):755-762. [doi: 10.1079/phn2005898] [Medline: 16925881]
24. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
25. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett Jr DR, Tudor-Locke C, et al. 2011 Compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
26. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010;33(3):676-682 [FREE Full text] [doi: 10.2337/dc09-1848] [Medline: 20190296]
27. Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet 2005;365(9461):785-799. [doi: 10.1016/S0140-6736(05)17987-2] [Medline: 15733721]
28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Okesene-Gafa KA, Li M, McKinlay CJ, Taylor RS, Rush EC, Wall CR, et al. Effect of antenatal dietary interventions in maternal obesity on pregnancy weight-gain and birthweight: healthy mums and babies (HUMBA) randomized trial. Am J Obstet Gynecol 2019;221(2):152.e1-152.13. [doi: 10.1016/j.ajog.2019.03.003] [Medline: 30878323]
30. Daly N, Farren M, McKeating A, O Kelly R, Stapleton M, Turner MJ. A medically supervised pregnancy exercise intervention in obese women: a randomized controlled trial. Obstet Gynecol 2017;130(5):1001-1010. [doi: 10.1097/AOG.0000000000002267] [Medline: 29016485]
31. Simmons D, Devlieger R, van Assche A, Jans G, Galjaard S, Corcoy R, et al. Effect of physical activity and/or healthy eating on GDM risk: the DALI lifestyle study. J Clin Endocrinol Metab 2017;102(3):903-913 [FREE Full text] [doi: 10.1210/jc.2016-3455] [Medline: 27935767]
32. Poston L, Bell R, Croker H, Flynn AC, Godfrey KM, Goff L, UPBEAT Trial Consortium. Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): a multicentre, randomised controlled trial. Lancet Diabetes Endocrinol 2015;3(10):767-777 [FREE Full text] [doi: 10.1016/S2213-8587(15)00227-2] [Medline: 26165396]
33. Renault KM, Nørgaard K, Nilas L, Carlsen EM, Cortes D, Pryds O, et al. The treatment of obese pregnant women (TOP) study: a randomized controlled trial of the effect of physical activity intervention assessed by pedometer with or without dietary intervention in obese pregnant women. Am J Obstet Gynecol 2014;210(2):134.e1-134.e9. [doi: 10.1016/j.ajog.2013.09.029] [Medline: 24060449]
34. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021;9(3):e26091 [FREE Full text] [doi: 10.2196/26091] [Medline: 33704075]
35. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
36. Hussain T, Smith P, Yee LM. Mobile phone-based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries: systematic review. JMIR Mhealth Uhealth 2020;8(5):e15111 [FREE Full text] [doi: 10.2196/15111] [Medline: 32463373]
37. Garnæs KK, Mørkved S, Salvesen Ø, Moholdt T. Exercise training and weight gain in obese pregnant women: a randomized controlled trial (ETIP trial). PLoS Med 2016;13(7):e1002079. [doi: 10.1371/journal.pmed.1002079] [Medline: 27459375]
38. Baruth M, Schlaff RA, Deere S, Walker JL, Dressler BL, Wagner SF, et al. The feasibility and efficacy of a behavioral intervention to promote appropriate gestational weight gain. Matern Child Health J 2019;23(12):1604-1612. [doi: 10.1007/s10995-019-02812-6] [Medline: 31541375]
39. Flynn AC, Seed PT, Patel N, Barr S, Bell R, Briley AL, UPBEAT consortium. Dietary patterns in obese pregnant women; influence of a behavioral intervention of diet and physical activity in the UPBEAT randomized controlled trial. Int J Behav Nutr Phys Act 2016;13(1):124 [FREE Full text] [doi: 10.1186/s12966-016-0450-2] [Medline: 27894316]
40. Ainscough KM, O'Brien EC, Lindsay KL, Kennelly MA, O'Sullivan EJ, O'Brien OA, et al. Nutrition, behavior change and physical activity outcomes from the PEARS RCT-An mHealth-supported, lifestyle intervention among pregnant women

with overweight and obesity. Front Endocrinol (Lausanne) 2019;10:938 [FREE Full text] [doi: 10.3389/fendo.2019.00938] [Medline: 32117047]
41. Szmeja MA, Cramp C, Grivell RM, Deussen AR, Yelland LN, Dodd JM. Use of a DVD to provide dietary and lifestyle information to pregnant women who are overweight or obese: a nested randomised trial. BMC Pregnancy Childbirth 2014;14:409 [FREE Full text] [doi: 10.1186/s12884-014-0409-8] [Medline: 25495459]
42. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women - the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]
43. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a novel eHealth intervention for management of gestational weight gain: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]
44. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
45. Bogaerts A, Bijlholt M, Mertens L, Braeken M, Jacobs B, Vandenberghe B, et al. Development and field evaluation of the INTER-ACT app, a pregnancy and interpregnancy coaching app to reduce maternal overweight and obesity: mixed methods design. JMIR Form Res 2020;4(2):e16090. [doi: 10.2196/16090] [Medline: 32130109]
46. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
47. Coughlin JW, Martin LM, Henderson J, Dalcin AT, Fountain J, Wang NY, et al. Feasibility and acceptability of a remotely-delivered behavioural health coaching intervention to limit gestational weight gain. Obes Sci Pract 2020;6(5):484-493. [doi: 10.1002/osp4.438] [Medline: 33082990]
48. Kominiarek MA, Gay F, Peacock N. Obesity in pregnancy: a qualitative approach to inform an intervention for patients and providers. Matern Child Health J 2015;19(8):1698-1712 [FREE Full text] [doi: 10.1007/s10995-015-1684-3] [Medline: 25652058]
49. Al Wattar BH, Pidgeon C, Learner H, Zamora J, Thangaratinam S. Online health information on obesity in pregnancy: a systematic review. Eur J Obstet Gynecol Reprod Biol 2016;206:147-152. [doi: 10.1016/j.ejogrb.2016.09.016] [Medline: 27693936]
50. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
51. Sanda B, Vistad I, Haakstad LA, Berntsen S, Sagedal LR, Lohne-Seiler H, et al. Reliability and concurrent validity of the international physical activity questionnaire short form among pregnant women. BMC Sports Sci Med Rehabil 2017;9:7 [FREE Full text] [doi: 10.1186/s13102-017-0070-4] [Medline: 28316789]
52. Sanda B, Vistad I, Sagedal LR, Haakstad LA, Lohne-Seiler H, Torstveit MK. Effect of a prenatal lifestyle intervention on physical activity level in late pregnancy and the first year postpartum. PLoS One 2017;12(11):e0188102. [doi: 10.1371/journal.pone.0188102] [Medline: 29176762]

## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials
**GWG:** gestational weight gain
**ICT:** information and communication technology
**IOM:** Institute of Medicine
**LGA:** large for gestational age
**MET:** metabolic equivalent of task
**RCT:** randomized controlled trial
**T0:** time 0
**T1:** time 1

*Edited by L Buis; submitted 15.04.21; peer-reviewed by J Alvarez Pitti, J Nicklas, S Lippke, E Hohman; comments to author 22.07.21; revised version received 30.09.21; accepted 10.12.21; published 15.02.22.*

*Please cite as:*
*Gonzalez-Plaza E, Bellart J, Arranz Á, Luján-Barroso L, Crespo Mirasol E, Seguranyes G*
*Effectiveness of a Step Counter Smartband and Midwife Counseling Intervention on Gestational Weight Gain and Physical Activity in Pregnant Women With Obesity (Pas and Pes Study): Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2022;10(2):e28886*
*URL: https://mhealth.jmir.org/2022/2/e28886*
*doi:10.2196/28886*
*PMID:35166684*

Original Paper

# Characterizing and Modeling Smoking Behavior Using Automatic Smoking Event Detection and Mobile Surveys in Naturalistic Environments: Observational Study

DongHui Zhai[1,2], PhD; Ruud van Stiphout[3], PhD; Giuseppina Schiavone[3], PhD; Walter De Raedt[2], MSc; Chris Van Hoof[1,2,3], PhD

[1]Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium

[2]imec, Leuven, Belgium

[3]imec at OnePlanet Research Center, Wageningen, Netherlands

**Corresponding Author:**
Ruud van Stiphout, PhD
imec at OnePlanet Research Center
Bronland 10
Wageningen, 6708WH
Netherlands
Phone: 31 317 745 801
Email: ruud.vanstiphout@imec.nl

## Abstract

**Background:** There are 1.1 billion smokers worldwide, and each year, more than 8 million die prematurely because of cigarette smoking. More than half of current smokers make a serious quit every year. Nonetheless, 90% of unaided quitters relapse within the first 4 weeks of quitting due to the lack of limited access to cost-effective and efficient smoking cessation tools in their daily lives.

**Objective:** This study aims to enable quantified monitoring of ambulatory smoking behavior 24/7 in real life by using continuous and automatic measurement techniques and identifying and characterizing smoking patterns using longitudinal contextual signals. This work also intends to provide guidance and insights into the design and deployment of technology-enabled smoking cessation applications in naturalistic environments.

**Methods:** A 4-week observational study consisting of 46 smokers was conducted in both working and personal life environments. An electric lighter and a smartphone with an experimental app were used to track smoking events and acquire concurrent contextual signals. In addition, the app was used to prompt smoking-contingent ecological momentary assessment (EMA) surveys. The smoking rate was assessed based on the timestamps of smoking and linked statistically to demographics, time, and EMA surveys. A Poisson mixed-effects model to predict smoking rate in 1-hour windows was developed to assess the contribution of each predictor.

**Results:** In total, 8639 cigarettes and 1839 EMA surveys were tracked over 902 participant days. Most smokers were found to have an inaccurate and often biased estimate of their daily smoking rate compared with the measured smoking rate. Specifically, 74% (34/46) of the smokers made more than one (mean 4.7, SD 4.2 cigarettes per day) wrong estimate, and 70% (32/46) of the smokers overestimated it. On the basis of the timestamp of the tracked smoking events, smoking rates were visualized at different hours and were found to gradually increase and peak at 6 PM in the day. In addition, a 1- to 2-hour shift in smoking patterns was observed between weekdays and weekends. When moderate and heavy smokers were compared with light smokers, their ages ($P$<.05), Fagerström Test of Nicotine Dependence ($P$=.01), craving level ($P$<.001), enjoyment of cigarettes ($P$<.001), difficulty resisting smoking ($P$<.001), emotional valence ($P$<.001), and arousal ($P$<.001) were all found to be significantly different. In the Poisson mixed-effects model, the number of cigarettes smoked in a 1-hour time window was highly dependent on the smoking status of an individual ($P$<.001) and was explained by hour ($P$=.02) and age ($P$=.005).

**Conclusions:** This study reported the high potential and challenges of using an electronic lighter for smoking annotation and smoking-triggered EMAs in an ambulant environment. These results also validate the techniques for smoking behavior monitoring and pave the way for the design and deployment of technology-enabled smoking cessation applications.

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2018-028284

*(JMIR Mhealth Uhealth 2022;10(2):e28159)* doi:10.2196/28159

**KEYWORDS**

smoking behavior modeling; ambulatory study; wearable sensors; temporal patterns of smoking; Poisson mixed-effects model; mobile phone

## Introduction

### Background

After more than 100 years of popularity, cigarette smoking remains the single largest cause of preventable disease and death even in the 21st century [1,2]. Globally, there are 1.1 billion current smokers, and every year more than 8 million die prematurely because of smoking. In addition, smoking induces many other health and economic costs on the society [2]. Although more than half of current smokers make a serious attempt to quit each year in the United States [3], 90% of unaided quitters relapse within the first 4 weeks due to the lack of limited access to cost-effective and efficacious smoking cessation tools [4], and only around 2% can quit for good. If no better solutions are developed to help increase the success ratio of smoking abstinence, the prevalence of smoking will decline very slowly and can be expected to remain at high levels for decades into the future [5]. Nonetheless, for designing readily accessible and effective smoking cessation applications, there are generally 2 obstacles ahead.

The first challenge is the lack of appropriate tools for smoking prevention and monitoring in daily use. Primary care plays a central role in smoking cessation, but its high-quality services are costly and often constrained by physical factors such as distance or time [6]. In fact, only 8% of the smokers go to smoking cessation clinics or physicians for counseling when they try to quit smoking [7]. In addition, smoker-initiated retrospective reports or diaries are the main methods used in previous studies on smoking research, but there are challenges with synchronizing events with digitalized measurements and recall of annotations by participants. To overcome this barrier, researchers have been designing and using many smart gadgets for smoking behavior monitoring and modeling. For example, radio frequency sensors and inertial sensors to measure respiration and arm movements have been used for smoking detection [8,9]. Acoustic sensors and breath carbon monoxide sensors were used for monitoring smoking, in combination with electric lighters and wrist-worn sensors [10,11]. Finally, the most recent use of e-cigarettes makes it easier to track and model this behavior [12].

The second challenge is how to transform theoretical models on smoking into actionable guidance tools in the dynamic context of daily use [13]. Many existing smoking cessation apps only use simplistic tools such as calculators, educational text [14], photoaging images [15], and self-trackers [16] and fall short of providing features such as smart tracking, learning, and tailored feedback, which are mostly demanded by end users. To enable adaptive smoking interventions, a prerequisite is to collect multimodal data concurrent with smoking and then use them to analyze the temporal and contextual windows associated with smoking behavior. In the literature, a few groups have reported progress in this direction. For example, Saleheen et al [9] designed a multi-sensor approach (electrocardiography, 3-axis accelerometer, and respiration sensors) and used it to collect smoking-related data from 45 smokers. Later, they conducted a study with 55 participants to test their app (MyQuitPal) designed for smokers during their initial cessation process [17,18]. However, it was only used among hospitalized smokers for 4 days and mainly to test various visualization techniques of their prototype system with no evaluation of its efficacy.

Accurate monitoring and modeling of smoking behavior in real-life settings are crucial for designing and delivering appropriate smoking cessation interventions. To fulfill this goal, mobile health technology, combining the measurement of multimodal sensors with the computation power of ubiquitous mobile phones, could enable a quantified observation of ambulatory smoking behavior 24/7 in real life. Because this technique can capture diverse information relevant to the behavior of interest, it can not only support accurate analysis and modeling of smoking behavior but also deliver customized interventions.

### Objectives

The aim of this study is to acquire a better understanding of smoking behavior by analyzing data from a longitudinal study, in which smoking events were automatically tracked and smoking-contingent context and mood states were assessed using mobile technology.

## Methods

### Study

This was an observational study of smoking behavior in a real-life setting, following the protocol reported in [19]. Smoker volunteers were recruited from the Flanders area of Belgium to participate in a 4-week experiment. Inclusion criteria were adults aged between 18 and 65 years, current smokers, office workers, and with no psychological, cardiac, or respiratory problems. An intake questionnaire and informed consent form were filled out when the participants passed the screening phase and were registered for this study. The intake questions were about personal background information such as age, gender, BMI, and the 6-item Fagerström Test of Nicotine Dependence (FTND). The FTND is a validated standardized smoking assessment that can be converted to a final score ranging from 0 to 10 and is used to indicate the nicotine dependence of smokers [20]. The FTND score measures physiological dependence (ie, tolerance and withdrawal). However, it does not capture the behavioral and psychosocial dimensions of nicotine dependence [21].

When the experiment began, the participants downloaded and installed an experimental app called ASSIST [19] on their smartphone. Next, they were given 2 wearable sensors, 1

electrical lighter, and instructions on the use of these sensors and the app. They were also informed to solely use the assigned lighter to light cigarettes when they smoke, and they were asked not to share it with other smokers. The lighter was also connected to the app on their phone via Bluetooth and was used to trigger surveys.

### Ecological Momentary Assessment Surveys

Ecological momentary assessment (EMA) surveys have been used in many experiments to study smoking behavior to capture, for example, environmental factors and affect, which are common reasons for smoking relapse [22]. EMAs aim to capture more reliable experience sampling because of their more relevant timing around the event of smoking, as reported by Serre et al [23].

In the current design of this study, participants were prompted to make annotations about their emotional state such as affect and arousal, dependence symptoms such as craving, enjoyment of cigarettes, and difficulty resisting smoking and other contexts related to smoking (social, activity, etc). These prompts were primarily triggered by the smoking events captured by the electric lighter. To prevent smokers, especially heavy smokers, from overburden, EMA surveys could only pop up at least 45 minutes apart. In addition, when the Bluetooth connection was down, the triggering fell back on a predefined randomization mechanism. In such cases, users received at most 5 randomized surveys per day.

### Statistical Analysis

The EMA score correlations were assessed using the Spearman correlation coefficient. Comparison of the smoker groups (light smokers: ≤10 cigarettes per day vs moderate to heavy smokers: >10 cigarettes per day) was performed using the Mann-Whitney *U* test for continuous variables, including EMA variables (assuming a sufficient range of discrete values). A generalized Poisson regression model from the GLMMadaptive package in R (R Foundation for Statistical Computing) was used to model smoking rates in 1-hour windows [24]. This model was selected because hourly smoking rates followed a Poisson distribution with 1 cigarette being the most common value and a rapid decline for a higher number of cigarettes smoked.

## Results

### Data Set and User Statistics

In total, 52 adult smokers volunteered to participate in the study, but, of these, 6 (12%) decided to quit the study and were excluded from the data set. Table 1 lists the characteristics of the data set. Of the 46 participants, 28 (61%) were men and 18 (39%) were women, with a mean age of 36 (SD 9.9) years. In all, 26% (12/46) of participants did not report their BMI, and the rest (34/46, 74%) had a mean BMI of 25 (SD 4.8) kg/m$^2$.

Regarding FTND, no smokers were assessed with high nicotine dependence in the study; all the smokers belonged to the first 3 groups. During the experiment, 8639 cigarettes were tracked by the lighter over 902 participant days. Specifically, 67% (31/46) of participants smoked ≤10 cigarettes a day on average and were labeled as light smokers. A total of 28% (13/46) of moderate smokers consumed 10-20 cigarettes on average daily. Only 4% (2/46) of heavy smokers had smoked >20 cigarettes a day. In contrast, there were 52% (24/46) of light smokers and 44% (20/46) of moderate smokers, according to the self-reported average daily cigarette consumption.

Figure 1 shows the average and SD of cigarettes smoked daily by smokers in this study. The participants were ranked by average daily consumption of cigarettes. We observed that most smokers had a day-to-day variation >1 cigarette per day (CPD), except for the first light smoker M1. Specifically, 72% (33/46) of the smokers had a moderate variation of 2 to 5 CPD, 20% (9/46) with a variation of 5 to 7 CPD, and 7% (3/46) with a variation >7 CPD based on the measured smoking records [11,25,26]. Figure 2 compares the self-reported number of cigarettes and objectively measured ones with the electric lighter. It shows that 74% (34/46) of the smokers made estimations that deviated more than one cigarette (mean 4.7 per day, SD 4.2 per day), and 70% (32/46) overestimated it compared with the lighter measurements.

**Table 1.** Characteristics of the study population (N=46).

| Characteristics and categories | Participants, n (%) |
|---|---|
| **Gender** | |
| Male | 28 (61) |
| Female | 18 (39) |
| **Age (years)** | |
| <30 | 13 (28) |
| 30-55 | 31(68) |
| >55 | 2 (4) |
| **BMI (kg/m$^2$)** | |
| Underweight (<18.5) | 1 (2) |
| Normal (18.5-24.9) | 20 (44) |
| Overweight (25-29.9) | 5 (11) |
| Obese (>30) | 6 (13) |
| Unknown (not reported) | 14 (30) |
| **Fagerström Test of Nicotine Dependence** | |
| Very low (0-2) | 18 (39) |
| Low (3-4) | 18 (39) |
| Moderate (5-7) | 10 (22) |
| High (8-10) | 0 (0) |
| **Measured average daily cigarette consumption** | |
| Light (≤10 cigarettes per day) | 31 (68) |
| Moderate (10-20 cigarettes per day) | 13 (28) |
| Heavy (>20 cigarettes per day) | 2 (4) |
| **Self-reported average daily cigarette consumption** | |
| Light (≤10 cigarettes per day) | 24 (52) |
| Moderate (10-20 cigarettes per day) | 20 (44) |
| Heavy (>20 cigarettes per day) | 2 (4) |

**Figure 1.** The average number of cigarettes smoked daily by each participant during the experiment. The colored bars and black lines represent the mean and SD, respectively. The participants on the x-axis are sorted by the mean in ascending order. F: female; M: male.

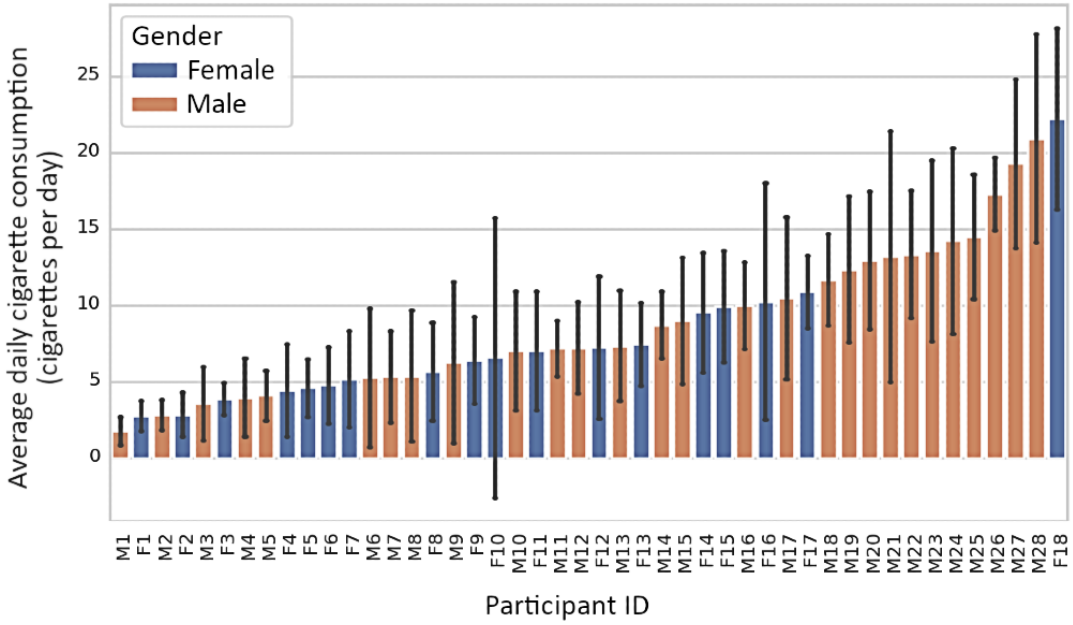
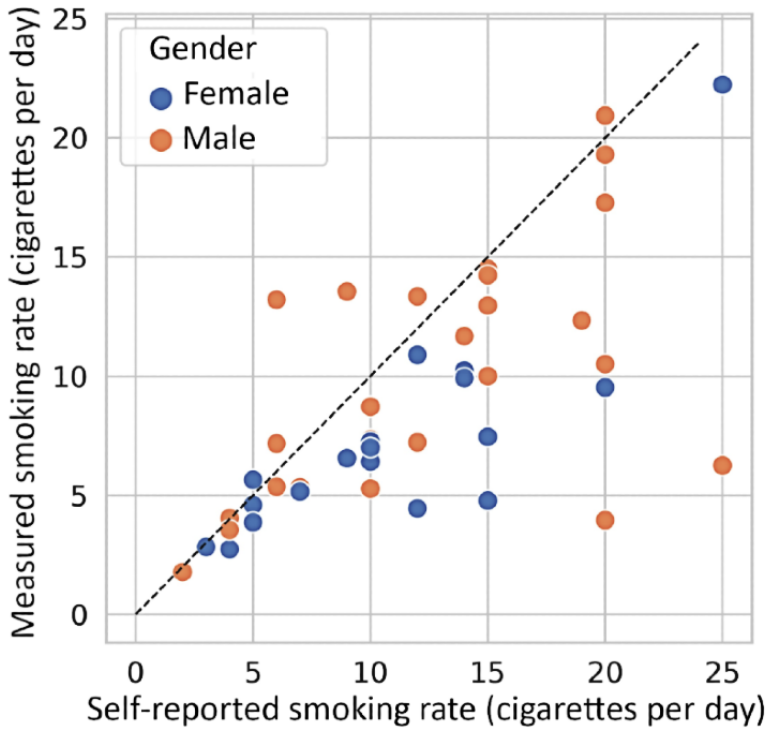


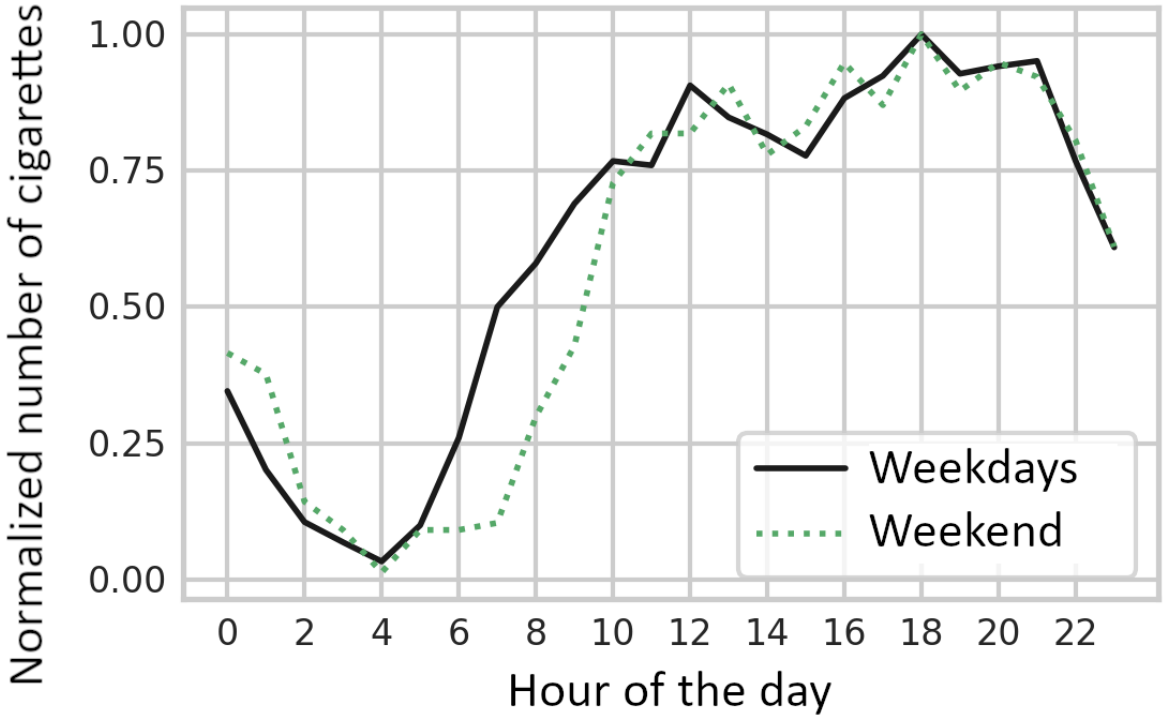

**Figure 2.** Comparison of the self-reported average number of cigarettes smoked daily with the measured consumption by the electric lighter. The dashed line is the diagonal of equal values. Gender is specified by colored markers as depicted in legend.




## Characterization of Smoking Patterns

Nation-wide surveys have shown differences in cigarette consumption between nonwork days and workdays [27]. As the smokers in this study were office workers, we assessed the differences between the days at work and at home, so the smoking records from all smokers during the study period of 4 weeks were aggregated and rescaled by the maximums for the weekdays and weekends. Figure 3 shows the distribution of cigarette consumption over 24 hours on weekdays and on weekends. The main finding is a 1- to 2-hour shift in the hourly cigarette consumption curves; on weekends, people smoke later, which is in line with a shift in sleeping times. For the weekdays, another peak is seen at 12 PM, which is usually the lunch time, and a valley at approximately 3 PM. For the weekends, however, 2 peaks in the afternoon are observed at around 1 PM and 4 PM, respectively. In addition, it can be observed from both curves that the number of cigarettes smoked generally increases later in the day, and most cigarettes are smoked around 6 PM.

**Figure 3.** Average number of cigarettes smoked per hour of the day, maximum normalized, and split between weekdays and weekends.




## EMA Reports of Smoking

To complement the FTND score on the psychosocial dimensions of nicotine dependence, this study used smoking-contingent EMA surveys to assess 5 smoking-relevant feelings. According to the dimensional models, emotion consists of at least two distinct dimensions; that is, valence and arousal [28]. In the EMA survey, these 2 emotions were assessed on a 9-point rating scale (numerically from −4 to 4). Emotional valence describes the extent to which an emotion is positive or negative, whereas arousal refers to its intensity; that is, the strength of the associated emotional state ranging from extremely calm to extremely excited [29]. There were another 3 questions used to assess the strength of craving, enjoyment of cigarettes, and difficulty of resisting smoking, respectively, on a scale from 0 to 4. The lower the score, the less intense the feeling. In total, 1839 EMA annotations were logged into the database.

The answer distributions of the 5 EMA questions are shown in Figure 4. The distributions of craving, enjoyment, and difficulty of resisting smoking were very similar and were mostly >2, which is the middle of the scale. A Spearman correlation test was used to verify the associations among them. Table 2 lists the mean and SD, as well as the coefficients of correlation and significance level. Their craving for cigarettes was 2.4 (SD 0.7), their average enjoyment was 2.5 (SD 0.7), and the average difficulty of resisting smoking was 2.4 (SD 0.7). In addition, craving is strongly correlated with enjoyment with a tested coefficient of 0.73 and $P<.001$, and it is also correlated with the difficulty of resisting smoking. In addition, enjoyment and difficulty of resisting smoking are correlated with a coefficient of 0.51 with high confidence. This result shows that increasing craving levels for cigarettes results in more enjoyment of smoking and more difficulty in resisting cigarettes.

Regarding the 2 dimensions of emotion, smoking was generally reported to be associated with more positive feelings for office workers during their daily lives. The emotional valence was 1.6 (SD 1.8). Most of the time, smokers were in a nonexcited state with a mean value of −0.6, but similar to emotional valence, a large variation exists. More detailed distributions of EMA answer data can be found in Multimedia Appendix 1 Figures S1-S4. Emotional valence was also negatively correlated with emotional arousal ($P<.001$).

**Figure 4.** The distribution of self-assessed levels of 5 different ecological momentary assessment (EMA) questions related to smoking. Median is depicted as horizontal line inside each box, the box itself shows the IQR, and the whiskers end 1.5 times IQR away from the IQR. Outliers are depicted by black diamonds.

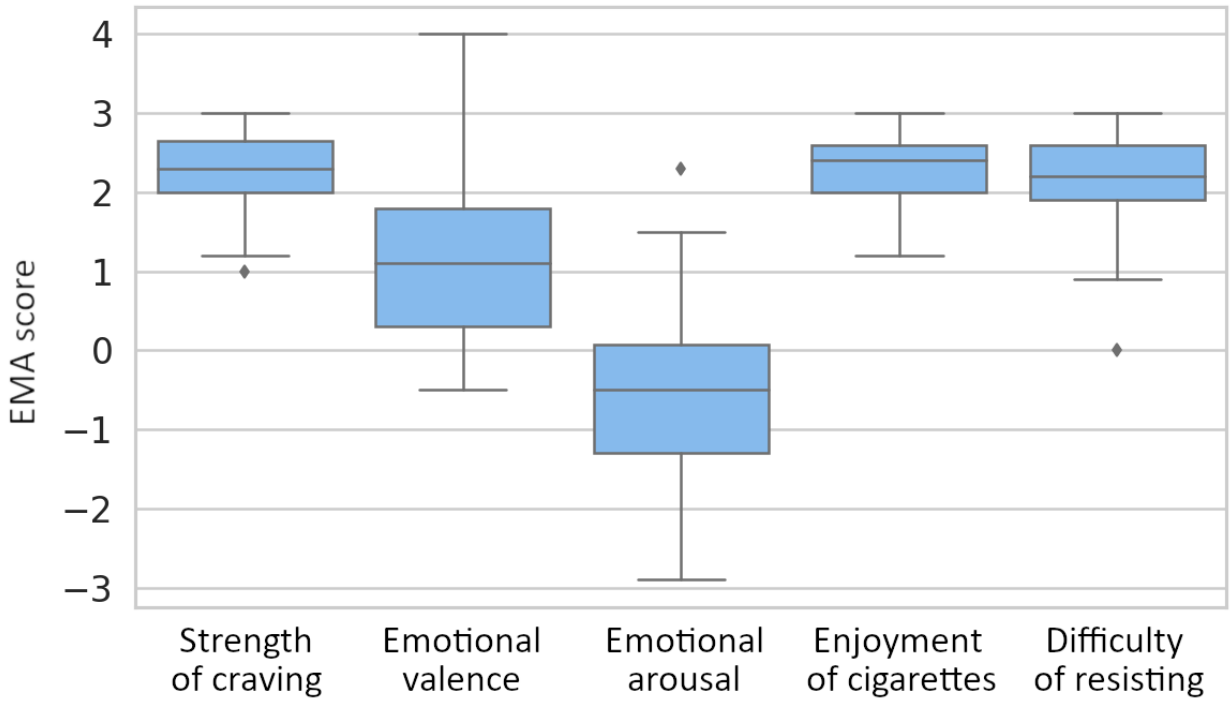



**Table 2.** Mean value, SD, and Spearman correlation coefficient, *P*, of 5 smoking-related feelings.

| Parameters | Strength of craving | Emotional valence | Emotional arousal | Enjoyment of cigarettes | Difficulty of resisting |
|---|---|---|---|---|---|
| Value, mean (SD) | 2.4 (0.7) | 1.6 (1.8) | −0.6 (2) | 2.5 (0.7) | 2.4 (0.7) |
| **Correlation coefficients (*P* value)** | | | | | |
| Strength of craving | N/A[a] | 0.07 | −0.08 | 0.73 | 0.65 |
| *P* value | N/A | .64 | .61 | <.001 | <.001 |
| Emotional valence | N/A | N/A | −0.50 | 0.35 | 0.16 |
| *P* value | N/A | N/A | <.001 | .02 | .33 |
| Emotional arousal | N/A | N/A | N/A | −0.22 | −0.00 |
| *P* value | N/A | N/A | N/A | .17 | .99 |
| Enjoyment of cigarettes | N/A | N/A | N/A | N/A | 0.51 |
| *P* value | N/A | N/A | N/A | N/A | <.001 |

[a]N/A: not applicable (duplication).

## Characterizing and Comparing the 2 Smoker Groups

Table 3 compares the characteristics of the 2 defined smoker type groups statistically. The 2 smoker groups were clustered based on objectively measured smoking rates, where moderate and heavy smokers were combined into 1 group (Table 1). Overall, 52% (16/31) of light smokers were men, whereas 80% (12/15) were men, in the moderate and heavy smoker groups. Although a lower percentage of male smokers were found in the light group, the difference was not significant (*P*=.10) when examined by the Fisher exact test. Regarding age, moderate and heavy smokers were 6 years older on average than light smokers in this study. In addition, the moderate and heavy groups tended to have longer smoking years, but this difference was not as significant as age. In addition, BMI and age at smoking initiation were not significantly different. Smokers, therefore, mostly initiate smoking in adolescence and are more likely to develop into moderate and heavy smokers as they smoke longer. Furthermore, the average FTND scores were significantly different between these 2 groups, with light smokers having 1.3 points lower mean scores. From the EMA answer comparison, it can be seen that craving, enjoyment, and difficulty in resisting smoking are all significantly stronger among moderate and heavy smokers than among light smokers. More positive and

calm feelings were reported among moderate and heavy smokers.

**Table 3.** Statistics of the features across the 2 smoker groups. *P* values are calculated by the Mann-Whitney U test.

| Type and feature (range) | Light smokers, mean (SD) | Moderate and heavy smokers, mean (SD) | *P* value |
|---|---|---|---|
| **Demographics** | | | |
| Age (years) | 33.7 (9.2) | 39.6 (10.3) | .02 |
| Smoking initiation age (years) | 17.5 (4.0) | 18.4 (4.2) | .23 |
| Smoking years | 16.2 (9.6) | 21.2 (11.4) | .09 |
| Fagerström Test of Nicotine Dependence (0 to 10) | 2.7 (1.9) | 4.0 (1.6) | .01 |
| **Ecological momentary assessment** | | | |
| Strength of craving (0 to 4) | 2.3 (0.8) | 2.6 (0.7) | <.001 |
| Emotional valence (−4 to 4) | 1.2 (1.8) | 1.8 (1.8) | <.001 |
| Emotional arousal (−4 to 4) | −0.2 (1.9) | −1.0 (2.1) | <.001 |
| Enjoyment of cigarettes (0 to 4) | 2.3 (0.7) | 2.6 (0.6) | .001 |
| Difficulty of resisting smoking (0 to 4) | 2.3 (0.7) | 2.5 (0.7) | .002 |

## Modeling the Count of Cigarettes in a 1-Hour Window

Cigarette smoking is a typical example of a recurrent event. The pattern of recurrent smoking events may depend on time-varying covariates. Meanwhile, demographics and background information such as age, gender, and nicotine dependence, which are time invariant over the time span of the experiment, also affect smoking patterns.

Modeling of the number of cigarettes in a 1-hour time window was performed with demographics (age, gender, and FTND) and timing of smoking (day of week and time of the day) as inputs. In total, there were 6654 such 1-hour windows during which 8631 cigarettes were smoked. The cigarette count in a 1-hour time window was affiliated timewise to the start of the time window. To decide on the selection of random effects, 3 Poisson mixed-effects models were compared with a baseline model, which assumes that the count of smoking events in a time window is constant (Table 4). When including the participants as the random intercept factor, the first mixed-effects model significantly improved with a *P* value <.001 in the analysis of variance test, which confirms that considerable between-participant variability exists in the data. Comparing the models where time of day (hour) was introduced as a random slope or fixed effect, resulted in hour being selected as a fixed factor because of its analysis of variance results and the smallest value for the Akaike information criterion. This model allows each participant to have a random intercept and has an hour of smoking as a fixed effect and was extended by age, gender, FTND, and day of the week as fixed factors.

> Count of cigarettes = approximately hour + age + gender + Fagerström Test of Nicotine Dependence + day of week + (1|participant_ID) **(1)**

In Table 5, the coefficients and corresponding SEs of the fixed factors from the final modeling results are included together with their *P* values. From the results, we can see only hour of smoking (*P*=.02) and age (*P*=.005) turned out to be informative for the number of cigarettes smoked in a 1-hour time window. FTND, gender, and day of the week were not informative for repeated smoking behavior. The model was also extended by each EMA variable, but these variables did not significantly improve the model (Multimedia Appendix 1 Table S1).

**Table 4.** Analysis of variance test results of 3 Poisson mixed-effects models relative to a baseline model for the selection of the random intercept or slope effects.

| Models | Akaike information criterion | *P* value (analysis of variance) |
|---|---|---|
| Count of cigarettes (approximately 1) | 15853.0 | N/A[a] |
| Count of cigarettes (approximately 1 + [1\|participant_ID]) | 15754.1 | <.001 |
| Count of cigarettes (approximately 1 + [hour\|participant_ID]) | 15754.5 | .17 |
| Count of cigarettes (approximately hour + [1\|participant_ID]) | 15750.7 | .02 |

[a]N/A: not applicable (reference model).

**Table 5.** Model coefficients, SEs on the coefficients, and *P* values for the dependent variables in the derived Poisson mixed-effects model.

| Variable | Coefficient (SE) | *P* value |
|---|---|---|
| Intercept | −0.075 (0.11) | .48 |
| Hour | 0.005 (0.002) | .02 |
| Age (years) | 0.007 (0.003) | .005 |
| Gender (male) | 0.033 (0.052) | .53 |
| Fagerström Test of Nicotine Dependence | −0.012 (0.015) | .43 |
| Day of week | 0.002 (0.006) | .67 |

# Discussion

## Principal Findings

The strength of this study is that it reports on smoking behavior by using an electric lighter that provides objective annotations of smoking and EMAs triggered by these events. This combination aims to provide more accurate annotations on both the timing and context of smoking when compared with retrospective self-reporting. This approach, when validated sufficiently, can potentially help smokers quit smoking by recommending interventions such as nicotine replacement therapy at the right time and context. The challenge, however, is to measure sufficient internal drivers and environmental factors related to smoking behavior to accurately model the known inter- and intravariability of smoking behavior. Ideally, models tailored to individuals must be developed, but these require extensive longitudinal data sets, which are not easy to obtain. This paper provides insights on smoking behavior with respect to timing, demographics, and relevantly timed EMAs and highlights potential variables and technologies for future studies.

In our data, we reported overestimation of self-reported smoking rates compared with measured smoking rates with the lighter. Reduced retrospective recall and lack of awareness of smoking behavior may have caused this [11,26]. However, our approach also relied on the compliance of participants to use the lighter for every smoking event. Therefore, the overestimation may be a result of suboptimal compliance. A recent study with this type of lighter showed that 92.2% of the smoking events were tracked by the lighter during 14 days of study among 22 participants [30]. They also found lower measured smoking rates and increased smoking rate variability compared with that in retrospective reporting. Therefore, we argue that using an electronic lighter provides a basis for annotation accuracy improvement.

The high variability of measured smoking rates within participants indicates the complexity of smoking behavior; for example, that habit is not the only driver of behavior. Similar variations in daily cigarette consumption were also reported by Hughes et al [25] using self-reported data. Time factors that we found to be important for modeling behavior are a delayed smoking pattern on the weekend and increasing smoking rates as the day progresses; that is, a sinusoidal pattern that has a maximum around dinner time. Of these 2, only hour of day was significant in the Poisson mixed-effects model, indicating that the found increasing smoking rates along the day are also characterized by increased repeated smoking in short time windows.

We compared the characteristics of light smokers and moderate to heavy smokers and found that nicotine dependence (FTND) and age were higher for heavier smokers but not for smoking initiation age and number of smoking years. Age was found to be a robust factor to describe smoking behavior, which was confirmed by the finding that it is the only significant static predictor remaining in the Poisson model. Gender on the other hand, did not show differences in smoking rate or smoking frequency. This is in line with the literature, although women were found to perceive more stress and nicotine withdrawal symptoms in a smoking cessation context, so gender may be important when modeling the risk of relapse [31,32].

Previous studies have suggested that people smoke cigarettes to regulate emotions and relieve negative emotions as reviewed by Kassel et al [33]. In our study, EMA annotations of craving, emotional valence, arousal, smoking enjoyment, and difficulty of resisting smoking were significantly different between light and moderate to heavy smokers but were not significant in the Poisson model to predict the number of cigarettes smoked in 1 hour. Higher craving has been linked to higher smoking incidence [34,35], and the difficulty of resisting smoking was found to be different among smoker types [36]. Emotional valence and arousal have been studied widely with respect to smoking behavior, and negative affect (NA) has been recognized as a nicotine withdrawal symptom and is correlated in some studies to increased smoking but is considered not a reliable antecedent of smoking, given that for example, stress influences NA as well [34]. Furthermore, NA and arousal seem to have a quadratic relationship with smoking probability, implying that linear models, such as the Poisson model in this paper, may perform suboptimally [37]. In addition, the effect of NA is diminished by other contextual factors such as other substance use including alcohol, which indicates that extensive experience sampling remains crucial [38]. The idea in this study was that every smoking event was annotated by an EMA. However, the ratio of EMA to smoking events was 21.29% (1839/8639), and more than half of the EMAs were not answered within a 1-hour window of smoking. This caused very few EMA-annotated smoking windows to be used in the Poisson models, resulting in low statistical power to find significance. A challenge is, therefore, to increase compliance and engagement in smokers when using EMAs, for example, with gamification.

To make effective smoking cessation tools, improved and extended data capture and modeling are needed. Our Poisson

model predicted whether 1 to 4 cigarettes were smoked in a 1-hour window. For nicotine replacement therapy strategies, prediction models for risk of a single smoking event in time (no smoking vs smoking) may be equally relevant, but that requires EMA data also to be available for time windows without smoking, which was not the case in this study owing to its design. These risk of relapse models should be tailored to each smoker because we also found the *participant* to be a significant random effect in the Poisson model. Continuous sensing of human physiology with wearables has the potential for capturing nicotine withdrawal stress responses as precursor to smoking. However, the challenges of high intra- and interparticipant variability and privacy concerns should be tackled before wearables can be used as a validated tool [39]. EMA reactivity; that is, the phenomenon of triggering a smoking event by answering a smoking-related EMA, has been shown to be an issue in smoking cessation contexts [40]. In our study, the most of the EMAs were triggered by smoking and only in rare cases EMAs were triggered randomly. EMA reactivity was therefore considered not an issue and was treated as another smoking cue modulated by the studied factors. However, when the focus shifts toward cessation tools, finding proxies for states derived from EMA that can be measured continuously and nonintrusively may become important. Examples of these are mobile health measures such as smartphone use and web-based activity that have the potential as a proxy for mood state. Like many other studies, this study may suffer from selection bias toward motivated and tech-savvy participants. Future model development and validation should be performed in larger trials, in which smoker population characteristics are matched. The resulting increased variety of smokers would also facilitate to learn which subpopulations benefit most from the current modeling approach.

## Conclusions

This study reported on the high potential and challenges of using an electronic lighter for smoking annotation and smoking-triggered EMAs in an ambulant environment. It is expected that to develop effective intervention strategies for smoking cessation, research needs to shift from population-based data sets based on self-reporting to richer data sets with objective environmental, physiological, and behavioral sensing so that individualized prediction models for relapse can be developed. We contributed to this by characterizing smoker types and by modeling smoking frequency using demographic, timing, and EMA data.

## Acknowledgments

We acknowledge the support of imec for hosting the program and research. Furthermore, we thank the volunteers who participated in this study. We would also like to thank the live-laboratory group of imec for their contribution to user recruitment and study management and the Enabling Digital Transformation (EDiT) department of imec for their work on the app and the website used in the study.

## Authors' Contributions

DHZ, GS, WDR, and CVH were involved in the trial and experiment design; DHZ and GS performed the data curation; DHZ, GS, and RvS performed the formal analysis and investigation; RvS, WDR, and CVH were involved in the project administration; CVH provided the resources; CVH supervised the study; DHZ was involved in writing and preparation of the original draft; DHZ and RvS were involved in writing, reviewing, and editing the manuscript; and CVH was the guarantor.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 645 KB - mhealth_v10i2e28159_app1.docx ]

## References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec;380(9859):2224-2260. [doi: 10.1016/S0140-6736(12)61766-8]
2. World Health Organization. Who Report on the Global Tobacco Epidemic 2019: Offer Help to Quit Tobacco Use. Geneva: World Health Organization; 2019.
3. Smoking cessation: a report of the Surgeon General (Executive Summary). U.S. Department of Health and Human Services. 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-executive-summary.pdf [accessed 2022-01-21]
4. Shahab L, West R, McIlvar M, Bauld L, Sutherland G. Why use CO-verified 4-week quit rates as the primary measure of stop smoking service success? National Centre for Smoking Cessation and Training (NCSCT). 2014. URL: https://www.ncsct.co.uk/usr/pub/NCSCT%20briefing%20on%204%20week%20quit%20rates.pdf [accessed 2020-06-23]

5. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2014.
6. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
7. Fiore MC, Novotny TE, Pierce JP, Giovino GA, Hatziandreu EJ, Newcomb PA, et al. Methods used to quit smoking in the United States. Do cessation programs help? J Am Med Assoc 1990;263(20):2760-2765. [Medline: 2271019]
8. Imtiaz M, Ramos-Garcia R, Senyurek V, Tiffany S, Sazonov E. Development of a multisensory wearable system for monitoring cigarette smoking behavior in free-living conditions. Electronics (Basel) 2017 Dec 28;6(4):104 [FREE Full text] [doi: 10.3390/electronics6040104] [Medline: 29607211]
9. Saleheen N, Ali AA, Hossain SM, Sarker H, Chatterjee S, Marlin B, et al. puffMarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. 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; September 7 - 11, 2015; Osaka Japan p. 999-1010. [doi: 10.1145/2750858.2806897]
10. Akyazi O, Batmaz S, Kosucu B, Arnrich B. SmokeWatch: a smartwatch smoking cessation assistant. In: Proceedings of the 25th Signal Processing and Communications Applications Conference (SIU). 2017 Presented at: 25th Signal Processing and Communications Applications Conference (SIU); May 15-18, 2017; Antalya, Turkey. [doi: 10.1109/siu.2017.7960536]
11. Scholl PM, Kücükyildiz N, Van Laerhoven K. When do you light a fire?: capturing tobacco use with situated wearable sensors. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. 2013 Presented at: UbiComp '13: The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 8 - 12, 2013; Zurich Switzerland p. 1309-1310. [doi: 10.1145/2494091.2499905]
12. Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. Circulation 2014 May 13;129(19):1972-1986 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.007667] [Medline: 24821826]
13. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
14. Finkelstein J, Cha EM. Using a mobile app to promote smoking cessation in hospitalized patients. JMIR Mhealth Uhealth 2016 May 06;4(2):e59 [FREE Full text] [doi: 10.2196/mhealth.5149] [Medline: 27154792]
15. Brinker TJ, Seeger W, Buslaff F. Photoaging mobile apps in school-based tobacco prevention: the mirroring approach. J Med Internet Res 2016 Jun 28;18(6):e183 [FREE Full text] [doi: 10.2196/jmir.6016] [Medline: 27352819]
16. Maramis C, Mylonopoulou V, Stibe A, Isomursu M, Chouvarda I. Developing a smartphone application to support smoking behavior change through social comparison. In: Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8856672]
17. Sharmin M, Weber T, Sarker H, Saleheen N, Kumar S, Ahmed S, et al. Opportunities and challenges in designing participant-centric smoking cessation system. In: Proceedings of the IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); July 4-8, 2017; Turin, Italy. [doi: 10.1109/compsac.2017.184]
18. Puchko A, Henning M, Jang C, Sharmin M. LapseAnalyzer: a web-based visual analytics tool for aiding smoking cessation. Western Washington University, Bellingham, WA, USA. URL: https://facultyweb.cs.wwu.edu/~sharmim/Publications/LapseAnalyzer-GHC2018.pdf [accessed 2022-01-21]
19. Zhai D, Schiavone G, Van Diest I, Vrieze E, DeRaedt W, Van Hoof C. Ambulatory smoking habits investigation based on physiology and context (ASSIST) using wearable sensors and mobile phones: protocol for an observational study. BMJ Open 2019 Sep 05;9(9):e028284 [FREE Full text] [doi: 10.1136/bmjopen-2018-028284] [Medline: 31492781]
20. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for nicotine dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
21. Dijkstra A, Tromp D. Is the FTND a measure of physical as well as psychological tobacco dependence? J Subst Abuse Treat 2002 Dec;23(4):367-374. [doi: 10.1016/s0740-5472(02)00300-8] [Medline: 12495799]
22. Vinci C, Haslam A, Lam CY, Kumar S, Wetter DW. The use of ambulatory assessment in smoking cessation. Addict Behav 2018 Aug;83:18-24 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.018] [Medline: 29398067]
23. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015 Mar 01;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
24. Rizopoulos D. GLMMadaptive: generalized linear mixed models using adaptive gaussian quadrature. GitHub. URL: https://drizopoulos.github.io/GLMMadaptive/index.html [accessed 2021-02-18]

25. Hughes JR, Shiffman S, Naud S, Peters EN. Day-to-day variability in self-reported cigarettes per day. Nicotine Tob Res 2017 Sep 01;19(9):1107-1111 [FREE Full text] [doi: 10.1093/ntr/ntx057] [Medline: 28339973]
26. Klesges RC, Debon M, Ray JW. Are self-reports of smoking rate biased? Evidence from the Second National Health and Nutrition Examination Survey. J Clin Epidemiol 1995 Oct;48(10):1225-1233. [doi: 10.1016/0895-4356(95)00020-5] [Medline: 7561984]
27. Cooper J, Borland R, Yong H, Hyland A, Cummings KM. Variations in daily cigarette consumption on work days compared with nonwork days and associations with quitting: findings from the international tobacco control four-country survey. Nicotine Tob Res 2013 Jan 30;15(1):192-198 [FREE Full text] [doi: 10.1093/ntr/nts110] [Medline: 22547777]
28. Barrett LF, Russell JA. The structure of current affect. Curr Dir Psychol Sci 2016 Jun 22;8(1):10-14. [doi: 10.1111/1467-8721.00003]
29. Leventhal AM, Zvolensky MJ. Anxiety, depression, and cigarette smoking: a transdiagnostic vulnerability framework to understanding emotion-smoking comorbidity. Psychol Bull 2015 Jan;141(1):176-212 [FREE Full text] [doi: 10.1037/bul0000003] [Medline: 25365764]
30. Tomko RL, McClure EA, Cato PA, Wang JB, Carpenter MJ, Karelitz JL, et al. An electronic, smart lighter to measure cigarette smoking: a pilot study to assess feasibility and initial validity. Addict Behav 2019 Nov;98:106052 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106052] [Medline: 31415971]
31. Lawless MH, Harrison KA, Grandits GA, Eberly LE, Allen SS. Perceived stress and smoking-related behaviors and symptomatology in male and female smokers. Addict Behav 2015 Dec;51:80-83 [FREE Full text] [doi: 10.1016/j.addbeh.2015.07.011] [Medline: 26240941]
32. Macy JT, Seo D, Chassin L, Presson CC, Sherman SJ. Prospective predictors of long-term abstinence versus relapse among smokers who quit as young adults. Am J Public Health 2007 Aug;97(8):1470-1475. [doi: 10.2105/AJPH.2006.101386] [Medline: 17600248]
33. Kassel JD, Stroud LR, Paronis CA. Smoking, stress, and negative affect: correlation, causation, and context across stages of smoking. Psychol Bull 2003 Mar;129(2):270-304. [doi: 10.1037/0033-2909.129.2.270] [Medline: 12696841]
34. Chandra S, Scharf D, Shiffman S. Within-day temporal patterns of smoking, withdrawal symptoms, and craving. Drug Alcohol Depend 2011 Sep 01;117(2-3):118-125 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.12.027] [Medline: 21324611]
35. Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, Kassel JD, et al. Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. J Abnorm Psychol 2002 Nov;111(4):531-545. [doi: 10.1037//0021-843x.111.4.531] [Medline: 12428767]
36. Thrul J, Ferguson SG, Bühler A. How do light and intermittent smokers differ from heavy smokers in young adulthood: the role of smoking restraint strategies. J Psychoactive Drugs 2016 Apr 27;48(3):153-158 [FREE Full text] [doi: 10.1080/02791072.2016.1172744] [Medline: 27120135]
37. Shiffman S, Dunbar MS, Li X, Scholl SM, Tindle HA, Anderson SJ, et al. Smoking patterns and stimulus control in intermittent and daily smokers. PLoS One 2014;9(3):e89911 [FREE Full text] [doi: 10.1371/journal.pone.0089911] [Medline: 24599056]
38. Magid V, Colder CR, Stroud LR, Nichter M, Nichter M, TERN Members. Negative affect, stress, and smoking in college students: unique associations independent of alcohol and marijuana use. Addict Behav 2009 Nov;34(11):973-975 [FREE Full text] [doi: 10.1016/j.addbeh.2009.05.007] [Medline: 19523773]
39. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
40. Rowan PJ, Cofta-Woerpel L, Mazas CA, Vidrine JI, Reitzel LR, Cinciripini PM, et al. Evaluating reactivity to ecological momentary assessment during smoking cessation. Exp Clin Psychopharmacol 2007 Aug;15(4):382-389. [doi: 10.1037/1064-1297.15.4.382] [Medline: 17696685]


## Abbreviations

**CPD:** cigarette per day
**EDiT:** Enabling Digital Transformation
**EMA:** ecological momentary assessment
**FTND:** Fagerström Test of Nicotine Dependence
**NA:** negative affect

*Edited by L Buis; submitted 23.02.21; peer-reviewed by K O'Connell, J Brigham; comments to author 27.05.21; revised version received 11.08.21; accepted 15.12.21; published 18.02.22.*

*Please cite as:*
*Zhai D, van Stiphout R, Schiavone G, De Raedt W, Van Hoof C*
*Characterizing and Modeling Smoking Behavior Using Automatic Smoking Event Detection and Mobile Surveys in Naturalistic Environments: Observational Study*
*JMIR Mhealth Uhealth 2022;10(2):e28159*
*URL: https://mhealth.jmir.org/2022/2/e28159*
*doi:10.2196/28159*
*PMID:35179512*

Original Paper

# The Content, Quality, and Behavior Change Techniques in Nutrition-Themed Mobile Apps for Children in Canada: App Review and Evaluation Study

Jacqueline Marie Brown[1*], MHSc; Beatriz Franco-Arellano[1*], PhD; Hannah Froome[1], MHSc; Amina Siddiqi[1], MSc; Amina Mahmood[1], BHSc; JoAnne Arcand[1], RD, PhD

Faculty of Health Sciences, Ontario Tech University, Oshawa, ON, Canada

[*]these authors contributed equally

**Corresponding Author:**
JoAnne Arcand, RD, PhD
Faculty of Health Sciences
Ontario Tech University
2000 Simcoe Street North
Oshawa, ON, L1H 7K4
Canada
Phone: 1 905 721 8668
Email: joanne.arcand@ontariotechu.ca

## Abstract

**Background:** Children increasingly use mobile apps. Strategies to increase child engagement with apps include the use of gamification and images that incite fun and interaction, such as food. However, the foods and beverages that children are exposed to while using apps are unknown and may vary by app type.

**Objective:** The aim of this study is to identify the app content (ie, types of foods and beverages) included in nutrition-themed apps intended for children, to assess the use of game-like features, and to examine app characteristics such as overall quality and behavior change techniques (BCTs).

**Methods:** This analysis used a cross-sectional database of nutrition-themed apps intended for children (≤12 years), collected between May 2018 and June 2019 from the Apple App Store and Google Play Store (n=259). Apps were classified into four types: *food games* or nongames that included *didactic nutrition guides*, *habit trackers*, and *other*. Food and beverages were identified in apps and classified into 16 food categories, as recommended (8/16, 50%) and as not recommended (8/16, 50%) by dietary guidelines, and quantified by app type. Binomial logistic regression assessed whether game apps were associated with foods and beverages not recommended by guidelines. App quality, overall and by subscales, was determined using the Mobile App Rating Scale. The BCT Taxonomy was used to classify the different behavioral techniques that were identified in a subsample of apps (124/259, 47.9%).

**Results:** A total of 259 apps displayed a median of 6 (IQR 3) foods and beverages. Moreover, 62.5% (162/259) of apps were classified as food games, 27.4% (71/259) as didactic nutrition guides, 6.6% (17/259) as habit trackers, and 3.5% (9/259) as other. Most apps (198/259, 76.4%) displayed at least one food or beverage that was not recommended by the dietary guidelines. Food game apps were almost 3 times more likely to display food and beverages not recommended by the guidelines compared with nongame apps (β=2.8; *P*<.001). The overall app quality was moderate, with a median Mobile App Rating Scale score of 3.6 (IQR 0.7). Functionality was the subscale with the highest score (median 4, IQR 0.3). Nutrition guides were more likely to be educational and contain informative content on healthy eating (score 3.7), compared with the other app types, although they also scored significantly lower in engagement (score 2.3). Most apps (105/124, 84.7%) displayed at least one BCT, with the most common BCT being *information about health consequences*.

**Conclusions:** Findings suggest nutrition-themed apps intended for children displayed food and beverage content not recommended by dietary guidelines, with gaming apps more likely to display not recommended foods than their nongame counterparts. Many apps have a moderate app quality, and the use of consequences (instead of rewards) was the most common BCT.

***(JMIR Mhealth Uhealth 2022;10(2):e31537)*** doi:10.2196/31537

**KEYWORDS**

mHealth; children; app quality; behavior change techniques; child nutrition; mobile apps; Canada; mobile phone

## Introduction

### Background

Establishing healthy eating patterns in early childhood promotes growth and development and reduces the risk of obesity and noncommunicable diseases [1-3]. This is also a time when habitual dietary patterns are established [4]. However, many children worldwide have poor quality diets [5-7]. In Canada, the average child has insufficient intakes of vegetables, fruit, and whole grain foods and consumes excess fat, sodium, and sugar [8,9], which is a dietary intake pattern strongly associated with childhood overweight, obesity, and chronic disease risk [10]. There are a multitude of factors that influence children's dietary attitudes, behaviors, and food choices, including intrinsic (eg, predisposed biological tendencies and gender) and extrinsic factors to children, such as the family (eg, mealtime and parenting style) and the community (eg, schools and media) [11]. Several strategies have been developed to improve child healthy eating habits including family and school-based interventions [12], nutrition policies [13,14], and the promotion of nutrition education, food skills, and food literacy [15]. These interventions can have profound health system and economic advantages [1,16].

Technology has evolved into a central part of everyday life [17,18]. In Canada and before the COVID-19 pandemic, over 25% of children spent more than 2.5 hours each day in front of a screen [19], with 99% of them having internet access outside of school. Almost a quarter of the children in grades 4 and 5 and half of the children in grade 7 owned a smartphone [20]. Studies have also demonstrated that children as young as 3 and 4 years old use their parents' smartphones between 25 and 50 minutes a day to watch television and videos, listen to music, and play games [20-23]. Children respond positively to fun and engaging challenges [24] that are captivating and motivating [25]. It is therefore understandable why games, advergames, and other digital activities are highly popular among children [26] and are substantively added to the digital marketplace [27].

Mobile health (mHealth), defined by the World Health Organization "as medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [28], is a convenient approach to support health promotion [29] and nutrition education among adults and children [17]. In particular, mHealth interventions have the potential to better engage youth about health-related topics, compared with traditional health interventions [17]. Among youth, mHealth apps have been found to increase fruit and vegetable intake [30-33], improve nutrition knowledge [32-34], promote food choice awareness and healthy eating behaviors [35,36], reduce sugar intake [37], and improve physical activity [38]. Furthermore, mHealth apps, as part of multicomponent interventions, can be effective tools to improve and support health-related behaviors [38]. Previous research has demonstrated that app quality, such as the ability to customize an app and ease of use, influence the overall effectiveness of mHealth apps on health and behavioral outcomes [39-41]. Additional evidence indicates that the integration of appropriate behavior change techniques (BCTs) [42] into mHealth apps further enhances their effectiveness [39,42,43]. Although there is strong evidence to support the efficacy of mHealth apps designed and evaluated by health researchers, minimal research exists on the content, quality, and use of evidence-based BCTs of publicly available mobile apps with health-related content.

Given its relevance, there has been increasing interest in studying the use and content of mobile apps and mHealth interventions that are available to the public [22,39,44-46]. For example, one study showed that diet and nutrition apps have a higher proportion of advertisements in comparison with other general health and wellness apps [22]. Reviews of pediatric weight management, healthy eating, and physical activity mobile apps found that most lacked any integration of expert recommendations [44], and less than 1% underwent scientific evaluation [47]. Another study found that children and adolescents are frequently exposed to the advertisement of unhealthy foods when using social media apps [45]. However, the specific foods and beverages displayed in mobile apps intended for children that are not from advertisements have not yet been examined [38]. Even less is known about foods and beverages displayed in nutrition-themed apps that contain the highly engaging game-like features that attract children and youth. This concept is highly relevant as not all nutrition-themed apps are considered mHealth apps. In addition, with new apps becoming available almost every day, it has become difficult for users, as well as for health professionals and researchers, to identify, evaluate, and use high-quality mobile apps to support healthy habits.

### Objectives

The aims of this study are 3-fold. First, we identified the app content (foods and beverages) included in different types of nutrition-themed apps intended for children and determined whether nutrition-themed apps with gaming features displayed more foods and beverages not recommended by dietary guidelines compared with nutrition-themed nongame apps. Second, we evaluated the overall quality of these apps using the Mobile App Rating Scale (MARS) [48], which is a "validated multidimensional measure of quality indicators" [48,49]. We also determined if app quality differed across the different types of nutrition-themed apps. Finally, we identified the different BCTs used in these apps, guided by a well-established taxonomy of such techniques [50].

## Methods

### Study Design

This research was a cross-sectional study that used a systematic search strategy and standardized evaluation process, modeled after comparable studies [51,52]. The Strengthening the

Reporting of Observational Studies in Epidemiology checklist is presented in Multimedia Appendix 1.

## App Selection

### Eligibility Criteria

Apps were eligible for inclusion in the analysis if they contained nutrition content relevant to children (ie, not targeted to parents), were rated by the app developer as being appropriate for an audience aged ≤12 years, were in the English language, were accessible to any user (ie, did not require an access code to use), were not affiliated with a brand or product, and were updated in the past 2 years. Excluded apps were simple food tracking apps (eg, calorie counting), chronic disease management apps (eg, diabetes), and apps that contained nutrition and food content irrelevant to dietary behaviors or education (eg, restaurant-themed time management games and word searches).

### App Search Strategy

Between May 2018 and June 2019, apps were identified from the Canadian Apple App Store and Google Play Store, which are app retailers containing the greatest number of publicly available apps in Canada [27]. The search methodology used in this study was adapted from comparable studies that used multiple keywords and terms to conduct their searches [51,52]. A search of app categories was conducted using 16 unique search terms as follows: nutrition game, eating game, diet game, food education, food game, nutrition education, child nutrition, kids nutrition, kids food, kids healthy eating, health food, child health, kids health, health game, health education, and child education. The data extracted from the identified apps were title, developer, number of downloads (when available), and cost. All app information was entered into a database that was used throughout the screening, selection, and evaluation processes.

### App Screening, Selection, and Classification

Identified apps first underwent screening for inclusion by 2 independent reviewers based on the title and developer. Duplicate apps, defined as apps appearing in both the Apple App Store and the Google Play Store, were identified and removed. Apps identified as relevant after the screening phase underwent a second independent review based on the detailed app description available in the respective app store. Those detailed descriptions determined if apps were eligible for inclusion in the analysis. Apps with a cost were purchased if they met the inclusion and exclusion criteria after reviewing the detailed descriptions. Apps were excluded from the study if they were removed from the marketplace before evaluation or if technical errors prevented a full evaluation (eg, app crashing). Disagreements at any stage of the app screening and evaluation process were resolved by a third independent reviewer.

The 4 app classifications (app types) were created using an inductive approach, which considered common app themes and characteristics observed during the review of apps, and 4 distinct classifications of apps were defined (Table 1). On the basis of their primary purpose, format, and core features, selected apps were classified into those four app types by two independent reviewers: food game, didactic nutrition guide, habit tracker, or other. Food games were defined as those that have implemented gamification techniques, such as rewards and competition, to engage the user in play involving food icons [53]. Nongame apps included didactic nutrition guides that provided information on food and nutrition to the user in written and picture format; habit trackers enabled users to log their food or drink intake [54] and apps classified as *other* did not contain any of the features nor had the primary purpose of the aforementioned app types.

**Table 1.** App types, definitions, and examples.

| App type | Definition | Examples |
|---|---|---|
| **Food game** | An app that implemented gamification techniques, such as rewards and competition, to engage the user in play involving food icons | Dr Panda Restaurant 2 by Dr Panda Ltd; Strawberry Shortcake Bake Shop by Budge Studios |
| **Nongame** | | |
| Didactic nutrition guide | An app that provided information on food and nutrition to the user in written and picture format | Nutrition Lookup by SparkPeople; SuperFoodsRx—Essential Guide by SuperFoods Partners, LLC |
| Habit tracker | An app that enabled users to log their food or drink intake | Fooducate—Nutrition Tracker by Fooducate, Ltd; Water Drink Reminder by Leap Fitness |
| Other | An app that did not contain the features or served the purpose of a food game, didactic nutrition guide, or habit tracker | Food & Cooking Genius by Brainscape; LaLa Lunchbox by LaLa Lunchbox |

### App Evaluation

Each app was downloaded and used for approximately 5 to 10 minutes for the reviewer to fully evaluate all aspects of the app content. In-app purchases were not evaluated in apps as these additional costs were determined to be largely inaccessible to the target audience of children. Apps were reviewed by 2 independent reviewers, and disagreements were resolved in consultation with a third independent reviewer.

## App Content Assessment

### Foods and Beverages Displayed

Foods and beverages displayed in apps were identified and classified into 16 food and beverage categories based on Canadian dietary guidelines, specifically the Canada's Food Guide (CFG) [55] and Canada's Dietary Guidelines [56]. Using the CFG, food and beverage categories were further classified as recommended (ie, foods that should be consumed more often) and not recommended (ie, foods that should be limited):

- Recommended food and beverages: fruit, vegetables, whole grain foods, unprocessed meat, fish, meat alternatives, milk products, and milk alternatives
- Not recommended food and beverages: refined grain foods, sugar drinks, desserts, chocolate and candies, salty snacks, pizza, fast foods, and processed meat

Apps were also assessed to determine if foods and beverages differed between *food game* apps and those without gaming features (ie, didactic nutrition guide, habit tracker, and others). In addition, textual healthy eating messages, such as "eat as many different colors as you can at each meal," "eat breakfast every day, breakfast gives you energy and helps you think and learn" and "make at least half of your grain products whole grain each day," were also identified in the apps.

### *Other Content Information*

The number of app downloads and app cost were also extracted from the Apple App Store or Google Play Store. The number of downloads was only extracted from apps available in the Google Play Store, as information on the number of downloads was not available from the Apple App Store.

### App Quality Assessment

App quality was determined for the apps using the MARS [48]. The MARS contains 23 items divided into 5 subscales that also contain specific domains. Subscales of engagement (entertainment, interest, customization, interactivity, and target group), functionality (performance, ease of use, navigation, and gestural design), aesthetics (layout, graphics, and visual appeal), and information (accuracy of app description, goals, quality of information, quantity of information, visual information, credibility, and evidence base) were used to assess the objective quality of included apps [48]. The fifth subgroup, subjective quality (Would you recommend this app? How many times do you think you would use this app? Would you pay for this app? What is your overall rating of the app?), was not included because the apps were evaluated by researchers, not by the target audience of children; therefore, the subjective scores would not reflect the views of the intended audience. Each domain was rated by researchers on a 5-point Likert scale: 1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent. If a domain was not present in the app, that domain was rated as *N/A* and was not included in the domain subscale score. The average of all scores from each evaluated domain was considered as the overall app quality (MARS).

### BCT Assessment

The use of different BCTs in apps was evaluated in a subsample of nongame nutrition apps, most likely to contain mHealth features, using the BCT Taxonomy (v1), developed by Michie et al [50]. This taxonomy identifies 93 hierarchically clustered techniques grouped within 16 behavioral clusters. Each app was evaluated for the presence or absence of each BCT listed in the taxonomy.

### Statistical Analyses

Data were tested for normality, and descriptive statistics were used to calculate the number and proportion of foods and beverages and other information displayed in the apps, both overall and by app type. To assess differences in foods and beverages between food game apps and nongame apps, data from didactic nutrition guides, habit trackers, and other apps were combined. A binomial logistic regression assessed whether food game apps displayed more foods not recommended by dietary guidelines compared with nongame apps. The proportion of foods and beverages in food game apps and nongame apps was calculated by food category and chi-square-tested for differences between both groups.

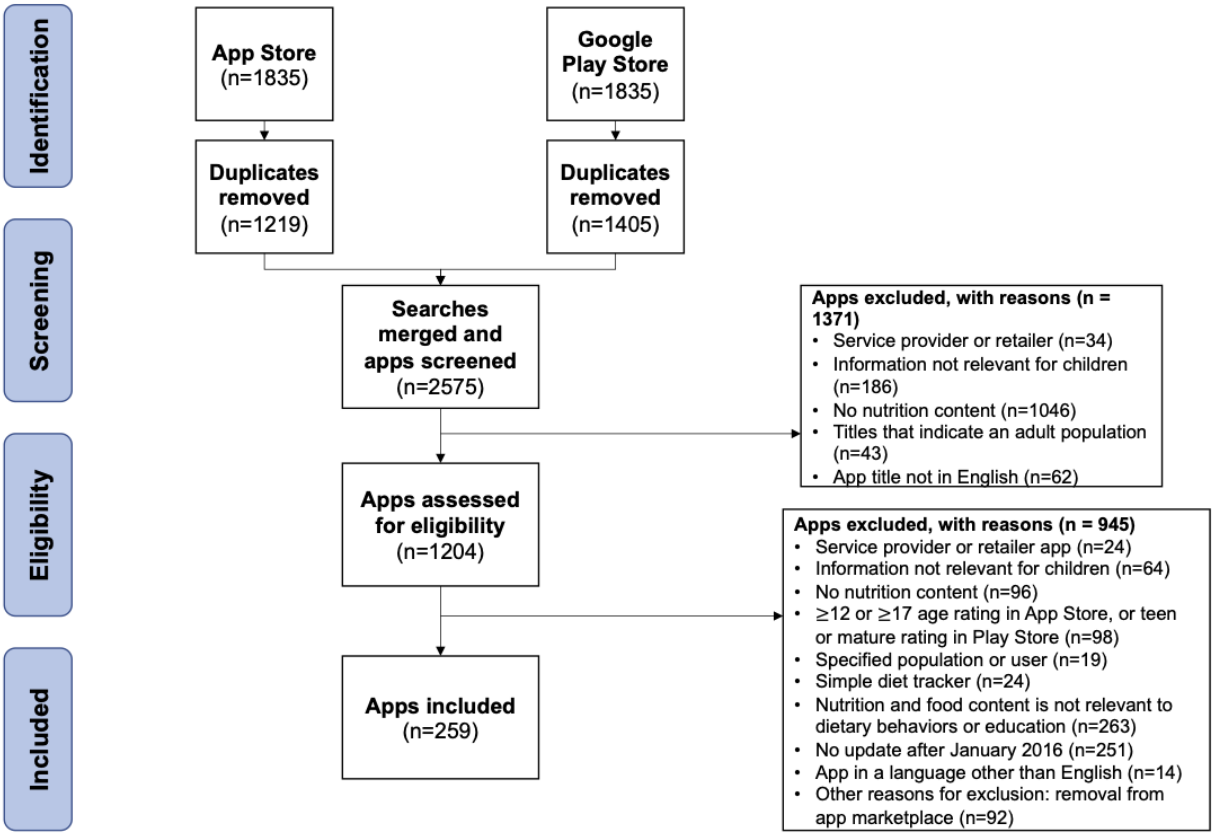
The median and IQR were calculated for the MARS score, subscales, and domains. The Kruskal-Wallis test was used to evaluate differences in the MARS scores and subscales between the 4 different app types. The frequency and proportion of use of the different BCTs were calculated by cluster label and by specific behavioral component for the subsample of apps using descriptive statistics. Statistical significance was set at $P<.05$, except for the between-group comparisons of the MARS score (and subscales) and the 4 different app types, where $\alpha$ was set at $P<.01$, to account for multiple comparisons. Statistical analyses were conducted using RStudio software (RStudio) [57].

## Results

### App Screening, Selection, and Classification

After removing duplicates, a total of 2575 unique apps were identified during the app search phase (Figure 1). From the 1204 apps that underwent title and developer screening review, 259 apps were eligible for inclusion in the analysis (interrater percent raw agreement=94.3% and Cohen $\kappa$=0.88). Multimedia Appendix 2 summarizes all included apps. Owing to the dynamic nature of the app marketplace, 60 apps were evaluated only by 1 reviewer, as these apps were removed from the marketplace during the app evaluation phase. For apps that were evaluated by 2 reviewers (199/259, 76.8%), app type classification yielded an interrater percent raw agreement of 91.7% and Cohen $\kappa$ of 0.87. From the 259 apps, 162 (62.5%) were classified as food games, 71 (27.4%) as didactic nutrition guides, 17 (6.6%) as habit trackers, and 9 (3.5%) as other.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for used to identify nutrition-themed apps intended for children.




## App Content

Among all 259 apps included in this analysis, apps included a median of 6 (IQR 3) food and beverage items. The most prevalent food and beverage items in the apps were fruit (200/259, 77.2%), milk products (185/259, 71.4%), vegetables (179/259, 69.1%), and unprocessed meats (175/259, 67.6%). The least prevalent food items were salty snacks (61/259, 23.6%), fast foods (57/259, 22%), pizza (51/259, 19.7%), and milk alternatives (38/259, 14.7%; Table 2). Despite the high prevalence of foods recommended by dietary guidelines, only 28.9% (75/259) of the apps included explicitly healthy eating messages.

**Table 2.** Types of food and beverages displayed in nutrition-themed apps for children.

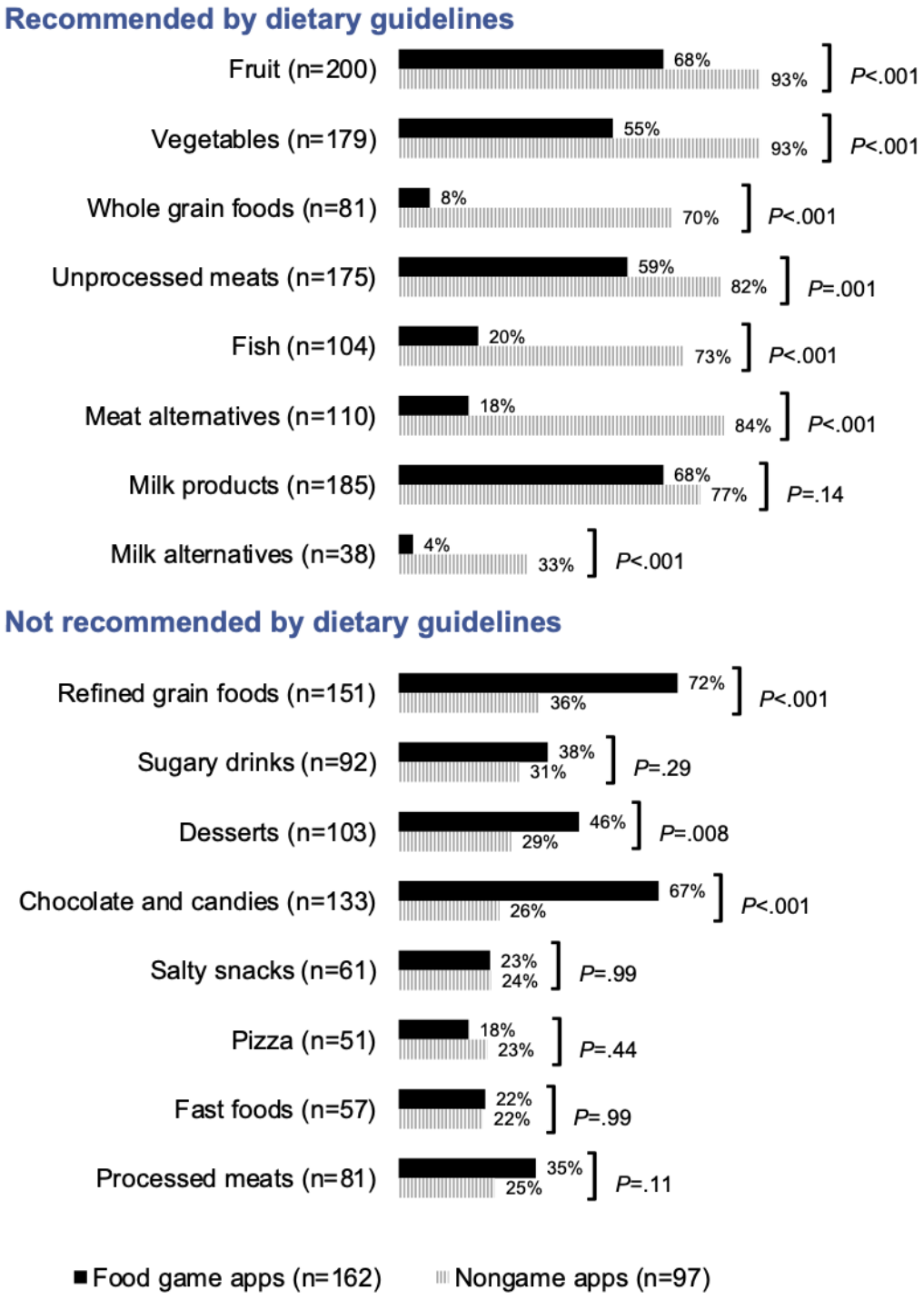
| Food and beverages | Value, n (%) | | | | |
|---|---|---|---|---|---|
| | All (n=259) | Food game (n=162) | Didactic nutrition guide (n=71) | Habit tracker (n=17) | Other (n=9) |
| **Presence of foods and beverages by category** | | | | | |
| **Recommended by dietary guidelines[a]** | | | | | |
| Fruit | 200 (77.2) | 110 (67.9) | 68 (95.8) | 14 (82.4) | 8 (88.9) |
| Vegetables | 179 (69.1) | 89 (54.9) | 69 (97.2) | 14 (82.4) | 7 (77.7) |
| Whole grain foods | 81 (31.2) | 13 (8) | 53 (74.6) | 14 (82.4) | 1 (11.1) |
| Unprocessed meats | 175 (67.6) | 95 (58.6) | 62 (87.3) | 13 (76.5) | 5 (55.6) |
| Fish | 104 (40.2) | 33 (20.4) | 56 (78.9) | 14 (82.4) | 1 (11.1) |
| Meat alternatives | 110 (42.5) | 29 (17.9) | 63 (88.7) | 14 (82.4) | 4 (44.4) |
| Milk products | 185 (71.4) | 110 (67.9) | 56 (78.9) | 14 (82.4) | 5 (55.6) |
| Milk alternatives | 38 (14.7) | 6 (3.7) | 18 (25.4) | 14 (82.4) | 0 (0) |
| **Not recommended by dietary guidelines[a]** | | | | | |
| Refined grain foods | 151 (58.3) | 116 (71.6) | 19 (26.78) | 12 (70.6) | 4 (44.4) |
| Sugary drinks | 92 (35.5) | 62 (38.3) | 15 (21.1) | 15 (88.2) | 1 (11.1) |
| Desserts | 103 (39.8) | 75 (46.3) | 10 (14.1) | 14 (82.4) | 4 (44.4) |
| Chocolate and candies | 133 (51.4) | 108 (66.7) | 9 (12.7) | 14 (82.4) | 2 (22.2) |
| Salty snacks | 61 (23.6) | 38 (23.4) | 6 (8.4) | 14 (82.4) | 3 (33.3) |
| Pizza | 51 (19.7) | 29 (17.9) | 7 (9.8) | 14 (82.4) | 1 (11.1) |
| Fast foods | 57 (22) | 36 (22.2) | 6 (8.4) | 13 (76.5) | 2 (22.2) |
| Processed meats | 81 (31.3) | 57 (35.2) | 8 (11.3) | 13 (76.5) | 3 (33.3) |
| **Displayed at least one food or beverage not recommended by dietary guidelines** | | | | | |
| 0 food and beverage | 61 (23.6) | 10 (3.9) | 45 (17.4) | 2 (0.8) | 4 (1.5) |
| ≥1 food and beverage | 198 (76.4) | 152 (58.7) | 26 (10) | 15 (5.8) | 5 (1.9) |
| Healthy messages | 75 (28.9) | 13 (8) | 51 (71.8) | 8 (47.1) | 3 (33.3) |
| App with cost | 35 (13.5) | 12 (7.4) | 16 (22.5) | 3 (17.6) | 4 (44.4) |
| Number of downloads | 121 (46.7) | 73 (45.1) | 34 (47.9) | 11 (64.7) | 3 (33.3) |

[a]Determined using categories and key messages provided by dietary guidelines (ie, Canada's Food Guide and Canada's Dietary Guidelines).

Overall, 46.7% (121/259) of apps had data available on the number of downloads. The median number of downloads was 500,000 (IQR 4,990,000), with a range of 50 to 89,000,000 downloads. Cost was evaluated for all apps, with 86.5% (224/259) of apps being free. The median cost for apps with a monetary charge was CAD $2.80 (US $2.24; 35/259, 13.5%), with a range between CAD $1 (US $0.80) and CAD $8.50 (US $6.80).

Food game apps, which comprised 62.5% (162/259) of apps overall, were almost 3 times as likely to display foods not recommended by dietary guidelines (β=2.8; *P*<.001), compared with nongame apps (ie, didactic nutrition guides, habit trackers, and other). In particular, high-sugar foods, such as chocolates and candies (*P*<.001) and desserts (*P*=.008) were significantly more likely to be displayed in food game apps, as shown in Figure 2 (detailed information in Multimedia Appendix 3). Importantly, food game apps also displayed significantly lower proportions of recommended foods and beverages in almost all recommended food categories, except for milk products, which was not significantly different between groups.

**Figure 2.** Proportion of foods and beverages displayed in food game apps and nongame apps by food category.




## App Quality

The overall app quality was moderate, with a median MARS score of 3.6 (IQR 0.7; Table 3). There were significant differences in overall quality between the 4 types of apps, with nutrition guides only having a median MARS rating of 3.2 ($P$<.001). Although nutrition guides were more likely to be educational and contain informative content on healthy eating (median score 3.7), compared with the other app types, they also scored significantly lower in overall engagement (score of 2.3) and in graphics and visual appeal (each with a score of 3). In addition, habit tracker apps were more likely to be engaging and aesthetic than were didactic nutrition guide apps. Overall, each subscale of the MARS was also moderate, with the highest ranked subscale being functionality (median 4, IQR 0.3), followed by aesthetics (median 3.7, IQR 0.7), information (median 3.6, IQR 1), and engagement (median 2.9, IQR 0.8). Although the information subscale received a moderate median

MARS score, this average is likely magnified as many apps did not contain goals (234/259, 90.3%) or visual information (238/259, 91.9%), nor were they evaluated in available peer-reviewed literature (258/259, 99.6%). Overall, larger differences between app types were seen in the engagement and aesthetics subscales.

**Table 3.** Quality of nutrition-themed apps for children, overall, and by app type[a].

| App type | Value, median (IQR) | | | | | *P* value[b] |
|---|---|---|---|---|---|---|
| | All (n=259) | Food game (n=162) | Didactic nutrition guide (n=71) | Habit tracker (n=17) | Other (n=9) | |
| Mobile App Rating Scale | 3.6 (0.7) | 3.4 (0.5) | 3.2 (0.4) | 3.5 (0.3) | 3.6 (0.7) | <.001 |
| **Engagement** | 2.9 (0.8) | 3.0 (0.8) | 2.3 (0.7) | 3.3 (0.5) | 3.1 (1.4) | <.001 |
| Entertainment | 3 (1.5) | 2.5 (2) | 3.0(0) | 3 (0.6) | 3.5(1) | .02 |
| Interest | 3 (1) | 3 (1.5) | 3 (0) | 3 (0.6) | 3 (1) | .24 |
| Customization | 3 (1) | 3 (0.5) | 1 (2) | 4 (0.5) | 3 (3) | <.001 |
| Interactivity | 3 (1.5) | 3 (0.5) | 1 (2) | 4 (0.5) | 3 (2) | <.001 |
| Target group | 3.5 (1) | 4 (0.5) | 3 (1) | 3 (0) | 4.5 (1) | <.001 |
| **Functionality** | 4.0 (0.3) | 4.0 (0.4) | 4.0 (0.3) | 4.0 (0.3) | 4.1 (0.8) | .01 |
| Performance | 4 (0.5) | 4 (0.5) | 4 (0) | 4 (0.6) | 4 (0.5) | .01 |
| Ease of use | 4 (0) | 4 (0.1) | 4 (0.5) | 4 (0) | 4.5 (1) | .002 |
| Navigation | 4 (0) | 4 (0) | 4 (0) | 4 (0) | 4 (1) | .03 |
| Gestural design | 4 (0.5) | 4 (0.5) | 4 (0) | 4 (0) | 4 (1) | .008 |
| **Aesthetics** | 3.7 (0.7) | 3.7 (0.7) | 3.3 (0.5) | 4 (0.6) | 4 (0.3) | <.001 |
| Layout | 4 (0.5) | 4 (0.5) | 4 (0.5) | 4 (0.5) | 4 (0) | .007 |
| Graphics | 3.5 (1) | 3.5 (1) | 3 (0.5) | 4 (0.6) | 4 (0) | <.001 |
| Visual appeal | 3.5 (1) | 3.5 (1) | 3 (0.8) | 4 (0.6) | 4 (1) | <.001 |
| **Information** | 3.6 (1) | 3.5 (1.3) | 3.7 (0.3) | 3.5 (0.4) | 3.3 (0.9) | .22 |
| Accuracy | 4 (0.5) | 4 (0.5) | 4 (0) | 4 (0) | 4 (0) | .19 |
| Goals | 3 (0.5) | 3 (1) | 3 (0) | 3 (0.5) | 3 (0.5) | .46 |
| Quality | 4 (0.8) | 3 (0.1) | 4 (0) | 4 (0) | 4. (0.5) | <.001 |
| Quantity | 4 (1) | 3 (1) | 4 (0.5) | 3.5 (0.5) | 3 (1) | <.001 |
| Visual | 4 (1) | 3.5 (1) | 3.5 (1) | 4 (0) | 4 (0) | .33 |
| Credibility | 1 (2) | 1 (0) | 3 (1) | 3 (2) | 1 (1.5) | <.001 |
| Evidence-based[c] | N/A[d] | N/A | N/A | N/A | N/A | N/A |

[a]The 23-item Mobile App Rating Scale was used to assess the quality of the included apps on 4 subscales of engagement (5 domains), functionality (4 domains), aesthetics (3 domains), and information (7 domains). Each domain was rated on a 5-point Likert scale: 1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent. If a domain was not present in the app, that domain was rated as *N/A*. The average of all scores from each evaluated domain was considered as the overall app quality.

[b]Significant difference was set at $P<.01$ to account for multiple comparisons and determined by the Kruskal-Wallis test.

[c]Not evaluated as the number of responses was <5.

[d]N/A: not applicable.

## BCT Assessment

BCTs were identified in 84.7% (105/124 subsample) of apps that were evaluated. Among those, 72.4% (76/105) had 1 or 2 BCTs, 22.9% (24/105) had between 3 and 10 BCTs, and we also found 4.8% (5/105) of apps with more than 10 BCTs. The most common BCT clusters among the 105 apps were *natural consequences* (100/105, 95.2%), *shaping knowledge* (49/105, 46.7%), and *goal setting and planning* (47/105, 44.8%). The most common individual BCTs identified in apps were *information about health consequences* (92/105, 87.6%), followed by *instructions on how to perform a behavior* (49/105, 46.7%; see Table 4 for detailed information).

**Table 4.** Frequency of behavior change techniques (BCTs) identified in a sample of apps intended for children (n=105)[a].

| BCT cluster label and component | Frequency, n (%) |
|---|---|
| **Goal setting and planning** | 47 (44.8) |
| Goal setting (behavior) | 16 (15.2) |
| Problem Solving | 1 (0.9) |
| Goal setting (outcome) | 12 (11.4) |
| Action planning | 7 (6.7) |
| Review behavior goals | 6 (5.7) |
| Review outcome goals | 5 (4.8) |
| **Feedback and monitoring** | 36 (34.3) |
| Feedback on behavior | 11 (10.5) |
| Self-monitoring of behavior | 12 (11.4) |
| Self-monitoring of outcome of behavior | 10 (9.5) |
| Feedback on outcome of behavior | 3 (2.9) |
| **Social support** | 3 (2.9) |
| Social support (unspecified) | 3 (2.9) |
| **Shaping knowledge** | 49 (46.7) |
| Instruction on how to perform a behavior | 49 (46.7) |
| **Natural consequences** | 100 (95.2) |
| Information about health consequences | 92 (87.6) |
| Information about social and environmental-consequences | 2 (1.9) |
| Information about emotional consequences | 6 (5.7) |
| **Comparison of behavior** | 11 (10.5) |
| Demonstration of the behavior | 11 (10.5) |
| **Associations** | 11 (10.5) |
| Prompts cues | 11 (10.5) |
| **Repetition and substitution** | 4 (3.8) |
| Behavior substitution | 2 (1.9) |
| Habit formation | 1 (0.9) |
| Graded tasks | 1 (0.9) |
| **Comparison of outcomes** | 4 (3.8) |
| Credible source | 4 (3.8) |
| **Regulation** | 2 (1.9) |
| Reduce negative emotions | 1 (0.9) |
| Conserving mental resources | 1 (0.9) |
| **Identity** | 2 (1.9) |
| Identification of self as role model | 2 (1.9) |

[a]Identified using the behavior change technique taxonomy developed by Michie et al [50].

# Discussion

## Principal Findings

This study is among the first to identify and empirically evaluate foods and beverages displayed in nutrition-themed apps intended for children in publicly available app stores. Most apps displayed foods and beverages not recommended by dietary guidelines (especially among those apps with game-like features), which had a moderate app quality. Importantly, this study also identified the use of consequences, rather than rewards, as the most common BCT in apps most likely to contain mHealth features.

Apps displayed a median of 6 food or beverage items to children, and the majority (198/259, 76.4%) displayed foods and beverages not recommended by dietary guidelines. This finding is concerning as it is likely a conservative estimate of exposure to food products that are not recommended by dietary guidelines in apps. This study excluded apps from branded products and did not evaluate in-app advertisements by food companies, which could have largely increased the number of displayed foods and beverages that are not recommended by dietary guidelines [45]. It was not surprising that the types of foods and beverages differed by app type. For example, nutrition guides, which are informative and educational in nature, displayed more fruit, vegetables, meat alternatives, and milk products, whereas habit tracker apps displayed large amounts of both foods recommended and not recommended by dietary guidelines. In addition, the nutrition guide apps were found to be of lower overall quality, which was largely the result of having fewer engaging features and lower quality graphics and visual appeal. Of notable concern was that food game apps were more likely to display *unhealthy* foods to children, such as chocolates, candies, and desserts. They also displayed significantly fewer recommended foods compared with nongame apps. This finding concurs with other studies that have found that the food content in apps may not align with dietary recommendations and lack evidence-based health information [40,44,58,59]. Childhood is the formative life stage when food and nutrition preferences, attitudes, and habits are learned [13], which are traits known to carry over into adulthood [60]. The latter highlights the importance of enhancing healthy eating and nutrition knowledge at an early life stage [17]. As children are more likely to be exposed and influenced by web-based media platforms [26,61,62], especially high-quality and engaging apps [24], it is critical that app developers limit the use of *unhealthy* foods in mobile game apps for children [26,63,64]. Although nongame apps displayed food and beverage items that are recommended by dietary guidelines, their limited ability to engage users indicates that they are unlikely to positively influence healthy eating behaviors.

Although nutrition guides may benefit children from an educational standpoint, this study found that they may be less attractive and engaging for children based on their design features. This finding is relevant because there is significant potential for mobile apps that can be used to deliver engaging and high-fidelity interventions to educate and motivate children about healthy eating practices [65]. Several studies have emphasized the potential positive role of mHealth apps as cost-effective [46], low burden interventions to promote healthy eating, self-monitoring, and behavior change [66,67]. For instance, serious games (ie, digital games designed for educational purposes) have been found to support children's increased vegetable and fruit intake [30-33], knowledge of macronutrients [32,33], food choice skills [35], and reduced sugar intake [37]. Moreover, healthy eating messaging can be incorporated into apps. Although our study found that only 8% (13/162) of food game apps contained messages aligned with dietary guidelines, 72% (51/71) of didactic nutrition guides and 42% (8/17) of habit trackers displayed healthy eating messages. These findings suggest that it is possible to use apps as a vehicle to support healthy eating messaging and nutrition education, as it has been used in other areas such as physical and mental health [68,69]. However, many apps on the marketplace would require modifications to align messaging with dietary guidelines and to include elements to increase their overall quality (visual appeal, graphics, engaging features, and BCTs). As nutrition guide and habit tracker apps may not be highly engaging for children, integrating gaming elements into these educational and informative apps may be more impactful in promoting the uptake of healthy eating knowledge and behaviors [41,70]. An example of a game-based nutrition education app is Foodbot Factory [71], which was designed to teach children about CFG and positively influenced children's knowledge of CFG guidelines [34].

Importantly, the efficacy of mHealth interventions can be significantly impacted by overall app quality [48] and the integration of suitable theory-based BCTs [50]. The MARS, for instance, has been used to identify high-quality medication reminder apps [72]. However, a lack of initial and sustained engagement has also been identified as a key constraint that limits digital nutrition promotion interventions [73]. In this study, we found that the overall quality of the apps, as evaluated by the MARS, was moderate, with apps receiving a median quality score of 3.6/5. These results underline the need to integrate engagement and motivation cues, factors known to strongly influence how long children will interact with an app [24]. Most apps also used *information about health consequences* as a BCT, which seems to be an unsupported tactic to induce behavior change (and less useful among children), compared with incentivization of positive health behaviors (ie, the use of points and rewards) [74]. Furthermore, limited evidence exists on user testing [73] and the assessment of app effectiveness in terms of user satisfaction [22]. Thus, the development of effective and engaging mHealth apps not only requires evidence-based content and appropriate BCTs but also necessitates the feedback of end users and evaluation of effectiveness through appropriately designed studies.

## Limitations

There are limitations to this study. It may be argued that a search of this kind cannot be truly systematic because of the dynamic nature of the app marketplace and limited search and data extraction abilities [75,76]. However, the use of multiple search terms to identify apps allowed this search to reach saturation and capture the most common nutrition apps targeted to children. Although paid apps were included in this study when free apps contained in-app purchases, only the free content was analyzed because the use of in-app purchases is likely to be less accessible to our target audience of children and adolescents; this decision may have resulted in an underestimation of the foods and beverages included. In addition, for this analysis, information on data privacy was not assessed. Assessments of data privacy policies and procedures are critically important, especially when apps are targeted to children; however, this information was not publicly available at the time of data collection. Although the Apple App Store added information on data privacy in June 2020, the Google Play Store has not, as of January 2022. Finally, because of its cross-sectional design, this study does not allow us to determine the relationship between exposure to foods and beverages displayed in apps (particularly those not

recommended by dietary guidelines) and children's health, or their knowledge, attitudes, and behaviors related to such foods. However, describing the exposure is a first step to investigating this relevant topic, and this should be a focus of future research.

### Comparison With Prior Work

An earlier comparable study by Schumer et al [22] examined publicly available *diet* and *nutrition* apps available in the Google Play Store (n=86 apps that were relevant for any age group), identified the focus of apps (eg, education, tracking, and planning), diet types (eg, Paleo diet), and app features (eg, goal setting and feedback). In capturing this information, the authors described initial factors that potential users evaluate when deciding to use an app [22]. An Australian study used a comparable approach to evaluate food and nutrition-related mobile apps to support healthy family food provision [40]. This study found that apps targeting parents with children had an app quality score of 3.5, similar to our results; however, many apps also had poor engagement. This study builds upon the previous work by conducting an in-depth evaluation of the different foods and beverages displayed in apps and whether or not these foods and beverages met dietary guidelines. This study's novelty further expands our understanding of apps intended solely for children and not parents or adults in general. In addition, our study goes beyond the findings of Schumer et al [22] in the assessment of apps intended for children by identifying those that were games, which is critical to examine as gaming apps are highly influential among youth [17,46]. A major strength of this study was the use of standardized evaluation and classification of apps, which was conducted independently by 2 reviewers, with a third independent reviewer to resolve disagreements. This process ensured a rigorous evaluation of the app content and classification. Another major strength was the use of the 2 major app stores, in contrast to the study by Schumer et al [22], which only involved 1 app store.

### Conclusions

This research demonstrated that nutrition-themed apps intended for children displayed many foods and beverages not recommended by dietary guidelines, and food game apps were more likely to display *unhealthy* foods and beverages compared with nongame apps. We also found that many of these apps in the subsample have a moderate app quality, and most of them use *information about health consequences* as a BCT. Nevertheless, given their popularity, nutrition-themed game apps have the potential to be used by health professionals, researchers, educators, and app developers to create evidence-based apps for children that align with dietary guidelines, which can be used to encourage healthy eating habits. Future research is required to broaden our understanding of how youth use and interact with apps containing nutrition content, their influence, and possible use for promoting nutrition education and healthy eating.

### Acknowledgments

This study was funded by an Ontario Research Fund—Research Excellence Grant (RE08-092). The authors thank Mohammed Tayyib Khan for his contributions to this research.

### Authors' Contributions

JA and JMB contributed to the conceptualization and design of the study. JMB, HF, AS, and AM collected and extracted data. BF-A analyzed the data and drafted the final manuscript. JA supervised and oversaw all aspects of the study. All authors provided critical revisions for intellectual content and read and approved the final version.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
Strengthening the Reporting of Observational Studies in Epidemiology Statement—checklist of items that should be included in reports of observational studies.
[DOCX File , 25 KB - mhealth_v10i2e31537_app1.docx ]

Multimedia Appendix 2
Summary of nutrition apps included in the analysis (n=259).
[DOCX File , 48 KB - mhealth_v10i2e31537_app2.docx ]

Multimedia Appendix 3
Proportion of foods and beverages displayed in food game apps and nongame apps by food category (n=259).
[DOCX File , 16 KB - mhealth_v10i2e31537_app3.docx ]

### References

1. Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et al. A future for the world's children? A WHO–UNICEF–Lancet Commission. Lancet 2020 Feb;395(10224):605-658. [doi: 10.1016/s0140-6736(19)32540-1]

2. Hurley K, Yousafzai A, Lopez-Boo F. Early Child Development and Nutrition: A Review of the Benefits and Challenges of Implementing Integrated Interventions. Adv Nutr 2016 Mar;7(2):357-363 [FREE Full text] [doi: 10.3945/an.115.010363] [Medline: 26980819]
3. Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ. Developmental origins of non-communicable disease: implications for research and public health. Environ Health 2012 Jun 27;11:42 [FREE Full text] [doi: 10.1186/1476-069X-11-42] [Medline: 22715989]
4. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: a systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
5. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
6. Global strategy on diet, physical activity and health. World Health Organization. 2017. URL: http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf [accessed 2017-08-21]
7. Ending childhoold obesity. World Health Organization. 2016. URL: http://apps.who.int/iris/bitstream/10665/204176/1/9789241510066_eng.pdf?ua=1 [accessed 2022-01-20]
8. Tugault-Lafleur C, Black J. Differences in the quantity and types of foods and beverages consumed by Canadians between 2004 and 2015. Nutrients 2019 Feb 28;11(3):526 [FREE Full text] [doi: 10.3390/nu11030526] [Medline: 30823448]
9. Kirkpatrick SI, Raffoul A, Lee KM, Jones AC. Top dietary sources of energy, sodium, sugars, and saturated fats among Canadians: insights from the 2015 Canadian Community Health Survey. Appl Physiol Nutr Metab 2019 Jun;44(6):650-658 [FREE Full text] [doi: 10.1139/apnm-2018-0532] [Medline: 30951373]
10. Di Cesare M, Sorić M, Bovet P, Miranda JJ, Bhutta Z, Stevens GA, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019 Nov 25;17(1):212 [FREE Full text] [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
11. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. Nutrients 2018 May 31;10(6):706 [FREE Full text] [doi: 10.3390/nu10060706] [Medline: 29857549]
12. Haines J, Haycraft E, Lytle L, Nicklaus S, Kok FJ, Merdji M, et al. Nurturing children's healthy eating: position statement. Appetite 2019 Jun 01;137:124-133 [FREE Full text] [doi: 10.1016/j.appet.2019.02.007] [Medline: 30797837]
13. Kent MP, Dubois L, Wanless A. Self regulation by industry in food marketing is having little impact during children's preferred television viewing. Can J Diabetes 2011 Jan;35(2):151. [doi: 10.1016/s1499-2671(11)52052-0]
14. Velazquez C, Black J, Kent MP. Food and beverage marketing in schools: a review of the evidence. Int J Environ Res Public Health 2017 Sep 12;14(9):1054 [FREE Full text] [doi: 10.3390/ijerph14091054] [Medline: 28895921]
15. Ronto R, Ball L, Pendergast D, Harris ND. Food literacy at secondary schools in Australia. J Sch Health 2016 Nov 06;86(11):823-831. [doi: 10.1111/josh.12440] [Medline: 27714873]
16. The heavy burden of obesity. Organisation for Economic Co-operation and Development. 2019. URL: https://www.oecd.org/health/the-heavy-burden-of-obesity-67450d67-en.htm [accessed 2022-01-20]
17. The Lancet Digital Health. Child and adolescent health in the digital age. Lancet Digit Health 2020 Mar;2(3):e102 [FREE Full text] [doi: 10.1016/S2589-7500(20)30029-7] [Medline: 33328079]
18. Prensky M. The games generations: how learners have changed. In: Digital Game-Based Learning. New York: McGraw-Hill Education; 2001.
19. Physical activity and screen time among Canadian children and youth, 2016 and 2017. Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/n1/pub/82-625-x/2019001/article/00003-eng.htm [accessed 2022-01-20]
20. Steeves V. Young Canadian in a wired world: Phase III: trends and recommendations. MediaSmarts. 2015. URL: https://mediasmarts.ca/sites/mediasmarts/files/publication-report/full/ycwwiii_trends_recommendations_fullreport.pdf [accessed 2022-01-20]
21. Guedes SD, Morais RL, Santos LR, Leite HR, Nobre JN, Santos JN. Children's use of interactive media in early childhood - an epidemiological study. Rev Paul Pediatr 2020;38:e2018165 [FREE Full text] [doi: 10.1590/1984-0462/2020/38/2018165] [Medline: 31778410]
22. 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]
23. Kabali HK, Irigoyen MM, Nunez-Davis R, Budacki JG, Mohanty SH, Leister KP, et al. Exposure and use of mobile media devices by young children. Pediatrics 2015 Dec 02;136(6):1044-1050. [doi: 10.1542/peds.2015-2151] [Medline: 26527548]
24. Kim SC, Yoon D, Han EK. Antecedents of mobile app usage among smartphone users. J Mark Commun 2014 Aug 22;22(6):653-670. [doi: 10.1080/13527266.2014.951065]
25. Ravyse WS, Blignaut AS, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virt Real 2016 Sep 20;21(1):31-58. [doi: 10.1007/s10055-016-0298-4]
26. Mulligan C, Kent MP, Christoforou AK, L'Abbé MR. Inventory of marketing techniques used in child-appealing food and beverage research: a rapid review. Int J Public Health 2020 Sep 25;65(7):1045-1055. [doi: 10.1007/s00038-020-01444-w] [Medline: 32840633]

27. Number of apps available in leading app stores as of 1st quarter 2020. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2020-09-18]
28. WHO Global Observatory for eHealth. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
29. Baranowski T, Frankel L. Let's get technical! Gaming and technology for weight control and health promotion in children. Child Obes 2012 Feb;8(1):34-37. [doi: 10.1089/chi.2011.0103] [Medline: 22799477]
30. Thompson D, Bhatt R, Vazquez I, Cullen KW, Baranowski J, Baranowski T, et al. Creating action plans in a serious video game increases and maintains child fruit-vegetable intake: a randomized controlled trial. Int J Behav Nutr Phys Act 2015 Mar;12:39 [FREE Full text] [doi: 10.1186/s12966-015-0199-z] [Medline: 25890060]
31. Cullen KW, Liu Y, Thompson DI. Meal-specific dietary changes from squires quest! II: a serious video game intervention. J Nutr Educ Behav 2016 May;48(5):326-330 [FREE Full text] [doi: 10.1016/j.jneb.2016.02.004] [Medline: 27169641]
32. Holzmann SL, Schäfer H, Groh G, Plecher DA, Klinker G, Schauberger G, et al. Short-term effects of the serious game "Fit, Food, Fun" on nutritional knowledge: a pilot study among children and adolescents. Nutrients 2019 Aug 30;11(9):2031 [FREE Full text] [doi: 10.3390/nu11092031] [Medline: 31480257]
33. Hermans RC, van den Broek N, Nederkoorn C, Otten R, Ruiter EL, Johnson-Glenberg MC. Feed the Alien! The effects of a nutrition instruction game on children's nutritional knowledge and food intake. Games Health J 2018 Jun;7(3):164-174. [doi: 10.1089/g4h.2017.0055] [Medline: 29634366]
34. Froome HM, Townson C, Rhodes S, Franco-Arellano B, LeSage A, Savaglio R, et al. The effectiveness of the foodbot factory mobile serious game on increasing nutrition knowledge in children. Nutrients 2020 Nov 06;12(11):3413 [FREE Full text] [doi: 10.3390/nu12113413] [Medline: 33172094]
35. DeSmet A, Liu Y, De Bourdeaudhuij I, Baranowski T, Thompson D. The effectiveness of asking behaviors among 9-11 year-old children in increasing home availability and children's intake of fruit and vegetables: results from the Squire's Quest II self-regulation game intervention. Int J Behav Nutr Phys Act 2017 Apr 21;14(1):51 [FREE Full text] [doi: 10.1186/s12966-017-0506-y] [Medline: 28431547]
36. Kato-Lin Y, Kumar UB, Prakash BS, Prakash B, Varadan V, Agnihotri S, et al. Impact of pediatric mobile game play on healthy eating behavior: randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 18;8(11):e15717 [FREE Full text] [doi: 10.2196/15717] [Medline: 33206054]
37. Majumdar D, Koch PA, Lee H, Contento IR, Islas-Ramos AD, Fu D. "Creature-101": a serious game to promote energy balance-related behaviors among middle school adolescents. Games Health J 2013 Oct;2(5):280-290 [FREE Full text] [doi: 10.1089/g4h.2013.0045] [Medline: 24761326]
38. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
39. Flaherty S, McCarthy M, Collins A, McAuliffe F. Can existing mobile apps support healthier food purchasing behaviour? Content analysis of nutrition content, behaviour change theory and user quality integration. Public Health Nutr 2017 Oct 30;21(2):288-298. [doi: 10.1017/s1368980017002889]
40. Mauch CE, Wycherley TP, Laws RA, Johnson BJ, Bell LK, Golley RK. Mobile apps to support healthy family food provision: systematic assessment of popular, commercially available apps. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11867 [FREE Full text] [doi: 10.2196/11867] [Medline: 30578213]
41. McLean G. Examining the determinants and outcomes of mobile app engagement - A longitudinal perspective. Comput Hum Behav 2018 Jul;84:392-403 [FREE Full text] [doi: 10.1016/j.chb.2018.03.015]
42. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
43. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
44. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]
45. Kent MP, Pauzé E, Roy E, de Billy N, Czoli C. Children and adolescents' exposure to food and beverage marketing in social media apps. Pediatr Obes 2019 Jun;14(6):e12508 [FREE Full text] [doi: 10.1111/ijpo.12508] [Medline: 30690924]
46. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
47. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: a scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]

48. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
49. 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 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
50. 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]
51. Nouri R, Kalhori SR, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
52. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
53. Schmidt-Kraepelin M, Thiebes S, Tran M, Sunyaev A. What's in the game? Developing a taxonomy of gamification concepts for health apps. In: Proceedings of the Hawaii International Conference on System Sciences. 2018 Presented at: Hawaii International Conference on System Sciences; 2018; Waikoloa Village, Hawaii. [doi: 10.24251/hicss.2018.150]
54. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada p. 3063-3072. [doi: 10.1145/1978942.1979396]
55. Canada's food guide. Government of Canada - Health Canada. 2019. URL: https://food-guide.canada.ca/en/ [accessed 2022-01-20]
56. Canada's dietary guidelines for health professionals and policy makers. Government of Canada - Health Canada. 2019. URL: https://food-guide.canada.ca/guidelines/ [accessed 2022-01-20]
57. RStudio. RStudio: Integrated Development for R. 2020. URL: http://www,rstudio.com/ [accessed 2022-01-20]
58. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
59. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
60. Smolin L, Grosvenor M, Gurfinkel D. Nutrition: Science and Applications, 3rd Canadian Edition. Hoboken, New Jersey, United States: John Wiley & Sons; 2020.
61. Sadeghirad B, Duhaney T, Motaghipisheh S, Campbell NR, Johnston BC. Influence of unhealthy food and beverage marketing on children's dietary intake and preference: a systematic review and meta-analysis of randomized trials. Obes Rev 2016 Oct 18;17(10):945-959. [doi: 10.1111/obr.12445] [Medline: 27427474]
62. Dorey E, McCool J. The role of the media in influencing children's nutritional perceptions. Qual Health Res 2009 May 03;19(5):645-654. [doi: 10.1177/1049732309334104] [Medline: 19380500]
63. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. Annu Rev Public Health 2008 Apr;29(1):253-272. [doi: 10.1146/annurev.publhealth.29.020907.090926] [Medline: 18031223]
64. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. Appetite 2013 Mar;62:209-215 [FREE Full text] [doi: 10.1016/j.appet.2012.04.017] [Medline: 22561190]
65. Hingle M, Patrick H. There are thousands of apps for that: navigating mobile technology for nutrition education and behavior. J Nutr Educ Behav 2016 Mar;48(3):213-218. [doi: 10.1016/j.jneb.2015.12.009] [Medline: 26965099]
66. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014 Sep;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
67. Brick TR, Mundie J, Weaver J, Fraleigh R, Oravecz Z. Low-burden mobile monitoring, intervention, and real-time analysis using the Wear-IT framework: example and usability study. JMIR Form Res 2020 Jun 17;4(6):e16072 [FREE Full text] [doi: 10.2196/16072] [Medline: 32554373]
68. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
69. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Dec 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]

70. Dovaliene A, Masiulyte A, Piligrimiene Z. The relations between customer engagement, perceived value and satisfaction: the case of mobile applications. Procedia Soc Behav Sci 2015 Dec;213:659-664 [FREE Full text] [doi: 10.1016/j.sbspro.2015.11.469]
71. Brown JM, Savaglio R, Watson G, Kaplansky A, LeSage A, Hughes J, et al. Optimizing child nutrition education with the foodbot factory mobile health app: formative evaluation and analysis. JMIR Form Res 2020 Apr 17;4(4):e15534 [FREE Full text] [doi: 10.2196/15534] [Medline: 32301743]
72. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
73. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
74. Melbye EL, Øverby NC, Øgaard T. Child consumption of fruit and vegetables: the roles of child cognitions and parental feeding practices. Public Health Nutr 2011 Oct 17;15(6):1047-1055. [doi: 10.1017/s1368980011002679]
75. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. A systematic review of pain-related apps. PLoS One 2014 Jul;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
76. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]


## Abbreviations

**BCT:** behavior change technique
**CFG:** Canada's Food Guide
**MARS:** Mobile App Rating Scale
**mHealth:** mobile health


*Edited by L Buis; submitted 24.06.21; peer-reviewed by I Gabashvili, U Bork; comments to author 28.10.21; revised version received 04.11.21; accepted 21.12.21; published 16.02.22.*

*Please cite as:*
*Brown JM, Franco-Arellano B, Froome H, Siddiqi A, Mahmood A, Arcand J*
*The Content, Quality, and Behavior Change Techniques in Nutrition-Themed Mobile Apps for Children in Canada: App Review and Evaluation Study*
*JMIR Mhealth Uhealth 2022;10(2):e31537*
*URL: https://mhealth.jmir.org/2022/2/e31537*
*doi:10.2196/31537*
*PMID:35171100*

Original Paper

# Mobile Apps for Hematological Conditions: Review and Content Analysis Using the Mobile App Rating Scale

Álvaro Narrillos-Moraza[1*], PharmD; Patricia Gómez-Martínez-Sagrera[1*], PharmD; Miguel Ángel Amor-García[1], PharmD; Vicente Escudero-Vilaplana[1], PharmD, PhD; Roberto Collado-Borrell[1], PharmD, PhD; Cristina Villanueva-Bueno[1], PharmD; Ignacio Gómez-Centurión[2], MD, PhD; Ana Herranz-Alonso[1], PharmD, PhD; María Sanjurjo-Sáez[1], PharmD

[1]Servicio de Farmacia, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[2]Servicio de Hematología, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[*]these authors contributed equally

**Corresponding Author:**
Vicente Escudero-Vilaplana, PharmD, PhD
Servicio de Farmacia, Hospital General Universitario Gregorio Marañón
Calle Doctor Esquerdo 43
Madrid, 28007
Spain
Phone: 34 915867714
Email: vicente.escudero@salud.madrid.org

## Abstract

**Background:** Hematological conditions are prevalent disorders that are associated with significant comorbidities and have a major impact on patient care. Concerning new tools for the care of these patients, the number of health apps aimed at hematological patients is growing. Currently, there are no quality analyses or classifications of apps for patients diagnosed with hematological conditions.

**Objective:** The aim of this study is to analyze the characteristics and quality of apps designed for patients diagnosed with hematological conditions by using the Mobile App Rating Scale (MARS).

**Methods:** We performed an observational, cross-sectional descriptive study of all smartphone apps for patients diagnosed with hematological conditions. A search was conducted in March 2021 using the following terms: *anemia*, *blood cancer*, *blood disorder*, *hematological cancer*, *hematological malignancy*, *hematological tumor*, *hematology*, *hemophilia*, *hemorrhage*, *lymphoma*, *leukemia*, *multiple myeloma*, *thalassemia*, *thrombocytopenia,* and *thrombosis*. The apps identified were downloaded and evaluated by 2 independent researchers. General characteristics were registered, and quality was analyzed using MARS scores. Interrater reliability was measured by using the Cohen κ coefficient.

**Results:** We identified 2100 apps in the initial search, and 4.19% (88/2100) of apps met the inclusion criteria and were analyzed. Of the 88 apps, 61% (54/88) were available on Android, 30% (26/88) were available on iOS, and 9% (8/88) were available on both platforms. Moreover, 7% (6/88) required payment, and 49% (43/88) were updated in the last year. Only 26% (23/88) of the apps were developed with the participation of health professionals. Most apps were informative (60/88, 68%), followed by preventive (23/88, 26%) and diagnostic (5/88, 6%). Most of the apps were intended for patients with anemia (23/88, 26%). The mean MARS score for the overall quality of the 88 apps was 3.03 (SD 1.14), ranging from 1.19 (lowest-rated app) to 4.86 (highest-rated app). Only 47% (41/88) of the apps obtained a MARS score of over 3 points (acceptable quality). Functionality was the best-rated section, followed by aesthetics, engagement, information, and app subjective quality. The five apps with the highest MARS score were the following: *Multiple Myeloma Manager*, *Hodgkin Lymphoma Manager*, *Focus On Lymphoma*, *ALL Manager*, and *CLL Manager*. The analysis by operating system, developer, and cost revealed statistically significant differences in MARS scores ($P$<.001, $P$<.001, and $P$=.049, respectively). The interrater agreement between the 2 reviewers was substantial (k=0.78).

**Conclusions:** There is great heterogeneity in the quality of apps for patients with hematological conditions. More than half of the apps do not meet acceptable criteria for quality and content. Most of them only provide information about the pathology, lacking interactivity and personalization options. The participation of health professionals in the development of these apps is low, although it is narrowly related to better quality.

*(JMIR Mhealth Uhealth 2022;10(2):e32826)* doi:10.2196/32826

**KEYWORDS**

blood; hematology; mHealth; mobile apps; quality; rating tool; mobile phone

# Introduction

## Background

The use of mobile technologies for health is increasing at an unstoppable rate. App capabilities for sharing health care information or real-time patient monitoring make them an important health tool because of their ease of use, broad reach, and wide acceptance [1]. At the beginning of 2021, more than 53,000 medical apps were available in the Android Play Store (one of the main download platforms) [2]. Medical apps have targeted a diverse number of conditions, such as diabetes [3,4], pain [5], rheumatic [6] and psychiatric disorders [7], COVID-19 [8-10], or cancer [11-13]. Apps for patients diagnosed with hematological conditions are also found on the main download platforms, although there is little information about them.

Hematological conditions comprise a wide range of disorders that can be classified as nonmalignant (anemia, hemorrhagic, or thrombotic disorders and conditions affecting blood-forming organs) and malignant (hematological cancers, such as Hodgkin and non-Hodgkin lymphoma, leukemia, or multiple myeloma, among others) [14]. These diseases meet all criteria for qualifying as a very important public health problem, with serious morbidities affecting patients worldwide [14-16]. Many of these conditions, such as hemophilia or anemia, are highly prevalent and become chronic. These patients could benefit from tools that improve treatment adherence or self-management guidelines, making medical apps an increasingly attractive option for this purpose [17,18].

Considering the large number of health apps available for patients with hematological conditions and the increasing interest in tools that encourage patient self-care, a proper review is needed. However, no clear consensus exists as to the appropriate method to assess the quality of health apps [19]. The Mobile App Rating Scale (MARS) is the most widely used scale for evaluating the quality and content of health apps. This allows the evaluation and comparison of apps by relating to their user engagement, functionality, aesthetics, and information quality [20,21]. In addition, it provides a quantitative and validated system that allows both users and health care professionals to avoid unreliable information.

## Objective

The aim of our study is to analyze the characteristics and quality of mobile apps for patients diagnosed with hematological conditions using the MARS.

# Methods

## Study Design

We performed an observational, descriptive, cross-sectional study of all smartphone apps for patients diagnosed with hematological conditions, including hematological malignancies, various types of anemia, and hemorrhagic and thrombotic diathesis, available on the Android and iOS platforms. The study followed the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) 2015 guidelines for systematic reviews [22].

## Eligibility Criteria

A search on the Apple App Store and Android Play Store was performed in February 2021 by 2 independent health professionals with experience in app analysis, design, and development (PGMS and ANM). The following search terms were used: *anemia*, *blood cancer*, *blood disorder*, *hematological cancer*, *hematological malignancy*, *hematological tumor*, *hematology*, *hemophilia*, *hemorrhage*, *lymphoma*, *leukemia*, *multiple myeloma*, *thalassemia*, *thrombocytopenia*, and *thrombosis*. The reviewers screened the title and download page of the apps. Only apps intended for patients or their caregivers and in English or Spanish were selected. Those apps potentially eligible were downloaded and installed on the appropriate, corresponding mobile device, regardless of the cost. iOS apps were installed on an iPhone 7 (version 14.4.2; Apple Inc) and Android apps on a Nexus 5X (Android version 8.1.0; Google LLC). Apps with nonscientific content; intended for health care professionals; duplicated; not specific for hematological conditions; specific to congresses, meetings, and charitable purposes; and those with restricted access were excluded from the review.

## Data Extraction and Quality Assessment

Apps were individually evaluated in isolation by the same 2 independent reviewers. Variables analyzed were app name, search term (for what the app was found), platform (Android or iOS), developer, hematological disorder, cost, app category (books and reference works, education, entertainment, health and fitness, health and wellness, lifestyle, medicine, simulation, and social media), date of the last update, language, and purpose. Concerning the developer, if hospitals, health authorities, universities, scientific societies, or patients' associations were involved in the design of an app, we classified them as *developed by a health organization*. The purpose was further classified into the following categories: diagnostic, informative, and preventive depending on whether the priority of the app was to run self-diagnosis, to provide generic data about one or several conditions, or to track treatment and symptoms, respectively. Grading was assessed by the same 2 independent reviewers according to the validated MARS. Data extraction, analysis, and grading were completed within 60 days.

The MARS is a multidimensional instrument that assesses the quality of mobile health apps. The quality assessment consists of a total of 23 items covering 5 dimensions. The dimensions are (1) engagement (5 items: entertainment, interest, customization, interactivity, and target adequacy), (2) functionality (4 items: performance, ease of use, navigation, and gestural design), (3) aesthetics (3 items: layout, graphics, and visual appeal), (4) information quality (7 items: accuracy

of app description, goals, quality of information, quantity of information, quality of visual information, evidence base, and credibility), and (5) subjective quality (4 items: recommendation, payment willingness, frequency of use, and overall rating). All items were rated on a 5-point scale (1=inadequate; 2=poor; 3=acceptable; 4=good; 5=excellent). Then, the overall quality of the app was obtained from the mean score of the domains [20,21].

### Data Analysis

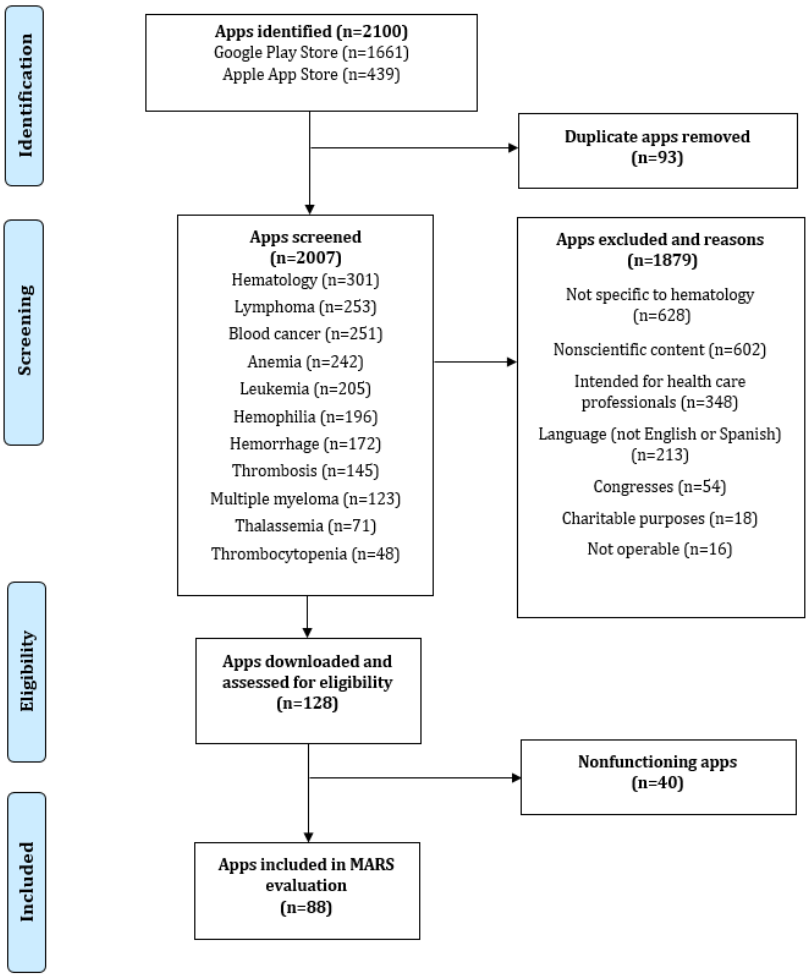
Quantitative variables were expressed as means and SDs and categorical variables as frequencies and percentages. Continuous variables were compared using the 2-tailed *t* test when the distribution was normal or the Mann-Whitney test when it was not. κ coefficient was used to measure the interrater reliability of the data analyzed by the 2 independent researchers [23]. Data were analyzed using Stata (version IC-16; StataCorp). A *P* value <.05 was considered statistically significant.

## Results

### Overview

A total of 2100 apps were retrieved from the Apple App Store and Android Play Store (1661 Android apps and 439 iOS apps). After screening the description and the screenshots available at the app platforms and deleting apps duplicated, 128 apps were selected as potentially eligible. After downloading and checking the fulfillment of the inclusion criteria, 88 apps were finally included in the descriptive analysis. A flow diagram illustrating the selection and exclusion of apps at various stages of the study is shown in Figure 1.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram and app selection. MARS: Mobile App Rating Scale.




### Characteristics and Purposes of Included Apps

In total, of the 88 apps, 8 (9%) were found on both digital distribution platforms, whereas 54 (61%) were obtained only from the Android Play Store, and 26 (30%) were only available at the Apple App Store. In addition, of the 88 apps, only 6 (7%) required payment (mean cost: mean €3.16 [US $3.60], SD €1.57 [US $1.79]). Table 1 shows the general characteristics of apps.

**Table 1.** General characteristics of the apps.

| Characteristics | Apps, n (%) |
|---|---|
| **Platform** | |
| Android | 54 (61) |
| iOS | 26 (30) |
| Android and iOS | 8 (9) |
| **Cost** | |
| No | 82 (93) |
| Yes | 6 (7) |
| **Category** | |
| Medicine | 35 (40) |
| Health and wellness | 33 (38) |
| Health and fitness | 8 (9) |
| Education | 7 (8) |
| Books and reference works | 1 (1) |
| Entertainment | 1 (1) |
| Lifestyle | 1 (1) |
| Simulation | 1 (1) |
| Social media | 1 (1) |
| **Date of the last update** | |
| 2012 | 1 (1) |
| 2016 | 2 (2) |
| 2017 | 7 (8) |
| 2018 | 12 (14) |
| 2019 | 19 (22) |
| 2020 | 34 (39) |
| 2021 | 9 (10) |
| Not updated | 4 (4) |
| **Language** | |
| English | 80 (92) |
| Spanish | 4 (4) |
| English and Spanish | 4 (4) |

Regarding purpose, most of the apps were informative (60/88, 68%), followed by preventive (23/88, 26%) and diagnostic (5/88, 6%). Of the 88 apps, a total of 43 apps (49%) were updated in the last year, and 23 apps (26%) were designed and developed with the participation of some kind of health care organization. The distribution of apps regarding hematological conditions was anemia (23/88, 26%), leukemia (12/88, 14%), hemophilia (11/88, 13%), thrombosis (8/88, 9%), thalassemia (7/88, 8%), hematological cancers (leukemia, lymphoma, or myeloma; 5/88, 6%), hemorrhage (5/88, 6%), lymphoma (4/88, 5%), leukemia or lymphoma (3/88, 3%), thrombocytopenia (3/88, 3%), multiple myeloma (2/88, 2%), hematological conditions (2/88, 2%), anemia or hemophilia (1/88, 1%), anemia or thalassemia (1/88, 1%), and hemochromatosis (1/88, 1%). The information on hematological conditions, purpose, app platform, free of cost, updates, developer, and language is shown in Tables 2 and 3.

**Table 2.** Characteristics of the apps analyzed. Apps are presented in alphabetical order, from those that start with "A" to those that start with "I."

| Name of the app (developer) | Hematological disease | Purpose | | | Platform | | Free | Updated in the last 12 months | Developed by a health organization | Language | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | I[a] | P[b] | D[c] | iOS | Android | | | | E[d] | S[e] |
| Alimentos para la anemia (Jotathat) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | | ✓ |
| All Blood Disease and Treatment A-Z (Patrikat Softech) | Blood disorders | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| ALL Manager (Point of Care) | Leukemia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| ALL Xplained (MedicineX) | Leukemia | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | |
| Anemia (Rouseapps) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Anemia (El Makaoui) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Anemia Care Diet & Nutrition (RecoveryBull) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Anemia Home Remedies (StatesApps) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Anemia Home Remedies (Salim Garba Usman) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Bleeder (Hannes Jung) | Hemophilia | | ✓ | | ✓ | | | ✓ | | ✓ | |
| Bleeding After Birth (Jaco Apps) | Hemorrhage | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Bleeding Disorder (Koodalappz) | Hemorrhage | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Blood Cancer (Digital Planete Space) | Hematological cancers | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Blood Cancer Tips (Free Apps For Everyone) | Hematological cancers | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Blood Clot Home & Natural Remedies (Salim Garba Usman) | Thrombosis | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Blood Count Reader free (Yurii Shevchenko) | Anemia | | | ✓ | | ✓ | ✓ | | | ✓ | |
| Blood Diseases (Medico_Guide) | Blood disorders | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Blood Group Genes (Gaurav Mathur) | Anemia or hemophilia | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | |
| Caprini DVT Risk (NorthShore University HealthSystem) | Thrombosis | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | |
| Childhood Leukemia: A Preventable Disease (FreeCreativity2019) | Leukemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| CIB—Coagulation Intervention Brigade (LFB Biomedicaments) | Hemorrhage | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | ✓ |
| CLL Manager (Point of Care) | Leukemia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| CLL Watch and Wait Tracker (Lymphoma Canada) | Leukemia or lymphoma | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| CML Life (Incyte Corporation) | Leukemia | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | |
| CML Today (Leukemia Patient Advocates Foundation) | Leukemia | ✓ | | | ✓ | | ✓ | | ✓ | | ✓ |

| Name of the app (developer) | Hematological disease | Purpose | | | Platform | | Free | Updated in the last 12 months | Developed by a health organization | Language | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | I[a] | P[b] | D[c] | iOS | Android | | | | E[d] | S[e] |
| Diario de INR (Web Factor BV) | Thrombosis | | ✓ | | ✓ | | | ✓ | | ✓ | ✓ |
| Don't Walk Alone (Lymphoma Canada) | Leukemia | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | |
| Easy Diagnosis—Thalassemia (Sarah Tinmaswala) | Thalassemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| EasyCoagLite (Loic Letertre) | Thrombosis | ✓ | | | ✓ | ✓ | ✓ | | | ✓ | |
| Focus On Lymphoma (Lymphoma Research Foundation) | Lymphoma | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| Folate & B12 Counter and Tracker (First Line Medical Communications) | Anemia | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | |
| Food For Anemia (MixLabApps) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| HaemActive—Fitness for People with haemophilia (NovoNordisk A/S) | Hemophilia | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Haemophilia Pal (Haemophilia Pal) | Hemophilia | | ✓ | | | ✓ | ✓ | | | ✓ | |
| Hemo Control (The Simulation Crew) | Hemophilia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| Hemophilia Disease (Bedieman) | Hemophilia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Hemophilia Support (MyHealthTeams) | Hemophilia | ✓ | | | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Hodgkin Lymphoma Manager (Point of Care) | Lymphoma | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| Home Remedies for Anemia (Anil Krishna) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| How To Cure Leukemia (Apps How To Apps) | Leukemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| iClot (Cranworth Medical Ltd) | Thrombosis | | ✓ | | | ✓ | ✓ | | | ✓ | |
| Increase A Low Platelet Count Naturally (FingertipApps) | Thrombocytopenia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| INR Care (Nikhil Patel) | Thrombosis | ✓ | | | ✓ | | | ✓ | | ✓ | |
| Iron Counter and Tracker (First Line Medical Communications) | Anemia | | ✓ | | ✓ | | | ✓ | ✓ | ✓ | |
| Iron Deficiency Anemia (Bedieman) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Iron Tracker—Hemochromatosis (IronTracker) | Hemochromatosis | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | |

[a]I: informative.

[b]P: preventive.

[c]D: diagnostic.

[d]E: English.

[e]S: Spanish.

**Table 3.** Characteristics of the apps analyzed. Apps are presented in alphabetical order, from those that start with "J" to those that start with "Z."

| Name of the app (developer) | Hematological disease | Purpose | | | Platform | | Free | Updated in the last 12 months | Developed by a health organization | Language | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | I[a] | P[b] | D[c] | iOS | Android | | | | E[d] | S[e] |
| Juntos contra la anemia (Andres Moran Tello) | Anemia | ✓ | | | | ✓ | ✓ | | ✓ | | ✓ |
| Leucemia—Sintomas Y Tratamiento—FAQ (Things To Do) | Leukemia | ✓ | | | | ✓ | ✓ | | | ✓ | ✓ |
| Leukemia: Symptoms And Treatment (The Reyv) | Leukemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Leukemia Disease (Bediemman) | Leukemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Leukemia Disease Treatment (Woochi Developer) | Leukemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| LLS CAR T (The Leukemia and Lymphoma Society) | Blood cancers | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| LLS Health Manager (The Leukemia and Lymphoma Society) | Blood cancers | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| LRF Understanding Lymphoma (Lymphoma Research Foundation) | Lymphoma | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| LRFFactSheets (Lymphoma Research Foundation) | Lymphoma | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| Microhealth Hemofilia (MicroHealth LLC) | Hemophilia | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mi Hemofilia (Rogelio Robles Tarano) | Hemophilia | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | |
| Multiple Myeloma Manager (Point of Care) | Multiple myeloma | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| My Blood Count (Sean Bottomley) | Anemia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| My HHT Tracker (Cure HHT) | Hemochromatosis | | ✓ | | ✓ | | ✓ | | | ✓ | |
| My INR (iMonitorMy) | Thrombosis | | ✓ | | | ✓ | ✓ | | | ✓ | |
| My Iron Manager (Good Dog Design Pty Ltd) | Anemia | | ✓ | | | ✓ | ✓ | ✓ | | ✓ | |
| Myeloma Cancer Guide (Everyone Learning Apps) | Multiple myeloma | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| myPROBE (Design2Code Inc) | Hemophilia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| myWAPPS (Design2Code Inc) | Hemophilia | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | |
| NCCN Patient Guides for Cancer (National Comprehensive Cancer Network) | Blood cancers | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| PA Pernicious Anaemia (B12 Global Limited) | Anemia | | ✓ | | ✓ | | | | | ✓ | |
| Pernicious-Anemia Advice (MoreFlow) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Pregnancy & Anaemia (Fumo) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |

| Name of the app (developer) | Hematological disease | Purpose | | | Platform | | Free | Updated in the last 12 months | Developed by a health organization | Language | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | I[a] | P[b] | D[c] | iOS | Android | | | | E[d] | S[e] |
| Recetas y consejos para combatir la anemia (App Free Enjoy) | Anemia | | ✓ | | | ✓ | ✓ | | | | ✓ |
| Recognize Anemia Disease (Media Clinic) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Recognize Hemophilia Disease (Media Clinic) | Hemophilia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Recognize Thalassemia Disease (Media Clinic) | Thalassemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Recognize Thrombocytopenia (Media Clinic) | Thrombocytopenia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Sickle Cell Anemia (Fumo) | Anemia | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | |
| Sickle Cell Anemia Home remedy (JGWS) | Anemia | | ✓ | | | ✓ | ✓ | | | ✓ | |
| Sickle Cell Disease (Kabirapp) | Anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| STB—Stop The Bleed (Uniformed Services University) | Hemorrhage | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | |
| SUSOKA (Subrata Saha) | Thalassemia | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | |
| Thalassemia Early Detection (ILIANA) | Thalassemia | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | |
| Thalassemia Disease (Bedieman) | Thalassemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| ThaliMe (Curatio Networks Inc) | Thalassemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| thalTracker (University Health Network) | Thalassemia | ✓ | | | ✓ | | ✓ | | | ✓ | |
| The Cancer App (Interactive Pharma solutions limited) | Blood cancers | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| The Seven Types of Anemia (Mrbeli) | Thalassemia or anemia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| Transplant Guidelines (National Marrow Donor Program/Be The Match) | Blood cancers | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | |
| Trombocytopenia Disease (Bedieman) | Thrombocytopenia | ✓ | | | | ✓ | ✓ | | | ✓ | |
| VTE Calc (Lindum Medical Ltd) | Thrombosis | ✓ | | | ✓ | | ✓ | | | ✓ | |

[a]I: informative.

[b]P: preventive.

[c]D: diagnostic.

[d]E: English.

[e]S: Spanish.

## Rating of Apps on the MARS

The specific MARS ratings for each app are shown in Tables 4 and 5. The mean score for the overall quality was 3.03 (SD 1.14), ranging from 1.19 (lowest rated app) to 4.86 (highest rated app). On average, the best-rated section was functionality (mean 3.44, SD 1.07), followed by aesthetics (mean 3.10, SD 1.23), engagement (mean 3.06, SD 1.32), information (mean 2.95, SD 1.09), and app subjective quality (mean 2.61, SD 1.28).

**Table 4.** Mobile App Rating Scale scores of the evaluated apps (rating out of 5). The first half (41/88, 47%) of the apps are presented here.

| Name of the app (developer) | Engagement, score | Functionality, score | Aesthetics, score | Information, score | Subjective quality, score | Overall |
|---|---|---|---|---|---|---|
| Multiple Myeloma Manager (Point of Care) | 5.00 | 5.00 | 5.00 | 4.57 | 4.75 | 4.86 |
| Hodgkin Lymphoma Manager (Point of Care) | 5.00 | 5.00 | 5.00 | 4.43 | 4.75 | 4.84 |
| Focus On Lymphoma (Lymphoma Research Foundation) | 4.90 | 4.88 | 4.83 | 4.43 | 5.00 | 4.81 |
| ALL Manager (Point of Care) | 5.00 | 4.88 | 4.67 | 4.43 | 4.88 | 4.77 |
| CLL Manager (Point of Care) | 5.00 | 4.75 | 4.67 | 4.43 | 4.75 | 4.72 |
| Transplant Guidelines (National Marrow Donor Program/Be The Match) | 4.80 | 5.00 | 4.50 | 4.43 | 4.50 | 4.65 |
| HaemActive—Fitness for people with haemophilia (NovoNordisk A/S) | 4.90 | 4.38 | 5.00 | 4.57 | 4.38 | 4.64 |
| Mi Hemofilia (Rogelio Robles Tarano) | 4.30 | 4.63 | 4.83 | 4.79 | 4.63 | 4.63 |
| My INR (iMonitorMy) | 4.60 | 4.88 | 4.17 | 4.64 | 4.38 | 4.53 |
| My Iron Manager (Good Dog Design Pty Ltd) | 4.60 | 4.50 | 4.50 | 4.71 | 4.13 | 4.49 |
| myWAPPS (Design2Code Inc) | 4.90 | 4.88 | 4.67 | 4.07 | 3.88 | 4.48 |
| CLL Watch and Wait Tracker (Lymphoma Canada) | 4.60 | 4.00 | 4.67 | 4.36 | 4.75 | 4.47 |
| Bleeder (Hannes Jung) | 4.70 | 4.50 | 4.50 | 4.14 | 4.50 | 4.47 |
| Microhealth Hemofilia (MicroHealth LLC) | 5.00 | 4.50 | 4.17 | 4.21 | 4.25 | 4.43 |
| Iron Tracker—Hemochromatosis (IronTracker) | 4.40 | 4.63 | 4.67 | 4.43 | 4.00 | 4.42 |
| STB—Stop The Bleed (Uniformed Services University) | 4.00 | 4.63 | 4.33 | 4.64 | 4.50 | 4.42 |
| PA Pernicious Anaemia (B12 Global Limited) | 4.70 | 4.50 | 4.50 | 4.14 | 4.13 | 4.39 |
| Hemophilia Pal (Haemophilia Pal) | 4.50 | 4.38 | 4.17 | 4.43 | 4.38 | 4.37 |
| ThaliMe (Curatio Networks Inc) | 4.60 | 4.50 | 4.67 | 3.79 | 3.88 | 4.29 |
| My HHT Tracker (Cure HHT) | 4.70 | 4.63 | 4.50 | 3.43 | 4.13 | 4.28 |
| CML Life (Incyte Corporation) | 4.40 | 4.75 | 4.67 | 3.71 | 3.63 | 4.23 |
| INR Care (Nikhil Patel) | 4.50 | 4.38 | 4.67 | 3.29 | 4.13 | 4.19 |
| Diario de INR (Web Factor BV) | 4.60 | 4.13 | 4.67 | 3.93 | 3.63 | 4.19 |
| NCCN Patient Guides for Cancer (National Comprehensive Cancer Network) | 3.10 | 4.75 | 4.67 | 4.36 | 4.00 | 4.17 |
| My Blood Count (Sean Bottomley) | 4.80 | 4.38 | 4.50 | 3.64 | 3.50 | 4.16 |
| LLS Health Manager (The Leukemia and Lymphoma Society) | 4.70 | 4.00 | 3.83 | 4.21 | 4.00 | 4.15 |
| thalTracker (University Health Network) | 4.60 | 4.50 | 4.50 | 3.86 | 3.25 | 4.14 |
| Don’t Walk Alone (Lymphoma Canada) | 4.90 | 3.38 | 4.17 | 4.21 | 3.38 | 4.01 |
| The Cancer App (Interactive Pharma Solutions Limited) | 4.30 | 4.25 | 4.50 | 3.71 | 3.13 | 3.98 |
| Hemophilia Support (MyHealthTeams) | 4.60 | 3.75 | 3.67 | 3.14 | 4.50 | 3.93 |
| CML Today (Leukemia Patient Advocates Foundation) | 3.70 | 4.38 | 3.50 | 4.29 | 3.63 | 3.90 |
| Pernicious Anemia Advice (MoreFlow) | 3.10 | 4.50 | 4.33 | 4.00 | 3.50 | 3.89 |
| ALL Xplained (MedicineX) | 3.10 | 4.38 | 4.00 | 3.71 | 2.75 | 3.59 |
| VTE Calc (Lindum Medical Ltd) | 3.80 | 4.38 | 3.17 | 3.86 | 2.63 | 3.56 |
| myPROBE (Design2Code Inc) | 3.80 | 4.25 | 4.00 | 2.71 | 2.13 | 3.38 |

| Name of the app (developer) | Engagement, score | Functionality, score | Aesthetics, score | Information, score | Subjective quality, score | Overall |
|---|---|---|---|---|---|---|
| Alimentos para la anemia (Jotathat) | 2.70 | 3.13 | 3.50 | 3.43 | 3.13 | 3.18 |
| Folate & B12 Counter and Tracker (First Line Medical Communications) | 3.10 | 4.13 | 2.50 | 3.36 | 2.63 | 3.14 |
| Blood Group Genes (Gaurav Mathur) | 4.20 | 3.25 | 4.00 | 2.21 | 2.00 | 3.13 |
| Iron Counter and Tracker (First Line Medical Communications) | 3.10 | 4.13 | 2.50 | 3.36 | 2.50 | 3.12 |
| All Blood Disease and Treatment A-Z (Patrikat Softech) | 2.60 | 3.88 | 3.00 | 3.21 | 2.50 | 3.04 |
| LRF Understanding Lymphoma (Lymphoma Research Foundation) | 3.10 | 3.38 | 3.33 | 3.36 | 2.00 | 3.03 |

**Table 5.** Mobile App Rating Scale scores of the evaluated apps (rating out of 5). The second half (47/88, 53%) of the apps are presented here.

| Name of the app (developer) | Engagement, score | Functionality, score | Aesthetics, score | Information, score | Subjective quality, score | Overall |
|---|---|---|---|---|---|---|
| LRFFactSheets (Lymphoma Research Foundation) | 3.40 | 3.38 | 3.17 | 3.14 | 1.88 | 2.99 |
| Juntos contra la anemia (Andres Moran Tello) | 3.00 | 3.13 | 3.17 | 3.07 | 2.38 | 2.95 |
| EasyCoagLite (Loic Letertre) | 3.30 | 2.88 | 3.00 | 2.79 | 2.50 | 2.89 |
| Hemo Control (The Simulation Crew) | 3.60 | 3.00 | 3.50 | 2.07 | 1.88 | 2.81 |
| Caprini DVT Risk (NorthShore University HealthSystem) | 3.70 | 3.38 | 2.67 | 2.57 | 1.38 | 2.74 |
| Recognize Thrombocytopenia (Media Clinic) | 2.30 | 3.88 | 2.00 | 2.50 | 1.75 | 2.49 |
| Thalasemia Early Detection (Iliana) | 3.00 | 3.25 | 2.67 | 2.21 | 1.25 | 2.48 |
| Recognize Hemophilia Disease (Media Clinic) | 2.30 | 3.88 | 1.67 | 2.50 | 1.88 | 2.44 |
| Sickle Cell Anemia Home remedy (JGWS) | 1.80 | 2.63 | 3.00 | 2.50 | 2.25 | 2.44 |
| Recognize Thalassemia Disease (Media Clinic) | 2.30 | 3.88 | 1.67 | 2.50 | 1.75 | 2.42 |
| Increase A Low Platelet Count Naturally (FingertipApps) | 1.80 | 3.38 | 2.33 | 2.43 | 2.13 | 2.41 |
| Recognize Anemia Disease (Media Clinic) | 2.30 | 3.25 | 2.17 | 2.50 | 1.75 | 2.39 |
| Anemia (RouseApps) | 2.00 | 3.38 | 2.00 | 2.57 | 1.88 | 2.36 |
| LLS CAR T (The Leukemia and Lymphoma Society) | 2.40 | 3.25 | 2.83 | 2.07 | 1.25 | 2.36 |
| Blood Clot Home & Natural Remedies (Salim Garba Usman) | 2.00 | 3.25 | 2.67 | 2.21 | 1.63 | 2.35 |
| Anemia Home Remedies (StatesApps) | 1.80 | 3.63 | 2.17 | 2.29 | 1.75 | 2.33 |
| Bleeding After Birth (JacoApps) | 1.90 | 3.13 | 2.83 | 2.14 | 1.63 | 2.33 |
| The Seven Types of Anemia (MrBeli) | 2.20 | 3.25 | 2.00 | 2.36 | 1.75 | 2.31 |
| Sickle Cell Disease (Kabirapp) | 1.90 | 2.38 | 2.83 | 2.00 | 2.25 | 2.27 |
| Bleeding Disorder (Koodalappz) | 2.10 | 2.50 | 2.83 | 2.29 | 1.63 | 2.27 |
| Blood Diseases (Medico_Guide) | 1.80 | 2.63 | 2.27 | 2.29 | 2.00 | 2.18 |
| Childhood Leukemia: A Preventable Disease (FreeCreativity2019) | 1.80 | 2.63 | 1.83 | 2.21 | 2.38 | 2.17 |
| Trombocytopenia Disease (Bedieman) | 1.80 | 2.75 | 2.00 | 2.50 | 1.75 | 2.16 |
| Food For Anemia (MixLabApps) | 1.80 | 3.13 | 2.33 | 2.14 | 1.38 | 2.16 |
| Iron Deficiency Anemia (Bedieman) | 1.80 | 3.00 | 2.00 | 2.21 | 1.75 | 2.15 |
| Hemophilia Disease (Bedieman) | 1.80 | 2.75 | 2.00 | 2.50 | 1.63 | 2.14 |
| Thalassemia Disease (Bedieman) | 1.70 | 2.75 | 2.00 | 2.50 | 1.63 | 2.12 |
| Easy Diagnosis—Thalassemia (Sarah Tinmaswala) | 1.90 | 3.25 | 1.67 | 2.00 | 1.38 | 2.04 |
| Home Remedies for Anemia (Anil Krishna) | 1.80 | 2.63 | 2.17 | 1.93 | 1.63 | 2.03 |
| SUSOKA (Subrata Saha) | 2.40 | 2.38 | 2.17 | 1.57 | 1.38 | 1.98 |
| Leucemia—Sintomas Y Tratamiento—FAQ (Things To Do) | 1.60 | 2.13 | 2.00 | 2.43 | 1.50 | 1.93 |
| Anemia Care Diet & Nutrition (RecoveryBull) | 2.00 | 2.13 | 2.00 | 2.00 | 1.50 | 1.93 |
| Sickle Cell Anemia (Fumo) | 1.80 | 2.38 | 1.83 | 1.79 | 1.63 | 1.88 |
| iClot (Cranworth Medical Ltd) | 1.90 | 2.00 | 2.17 | 1.93 | 1.38 | 1.87 |
| Pregnancy & Anaemia (Fumo) | 1.80 | 2.38 | 2.00 | 1.79 | 1.38 | 1.87 |
| Anemia (El Makaoui) | 1.50 | 2.75 | 2.00 | 1.64 | 1.25 | 1.83 |
| Anemia Home Remedies (Salim Garba Usman) | 1.80 | 2.00 | 1.83 | 1.86 | 1.63 | 1.82 |

| Name of the app (developer) | Engagement, score | Functionality, score | Aesthetics, score | Information, score | Subjective quality, score | Overall |
|---|---|---|---|---|---|---|
| CIB—Coagulation Intervention Brigade (LFB Biomedicaments) | 1.40 | 1.25 | 2.50 | 1.21 | 1.25 | 1.52 |
| Leukemia Disease (Bedieman) | 1.50 | 1.50 | 1.33 | 1.71 | 1.13 | 1.43 |
| Recetas y consejos para combatir la anemia (App Free Enjoy) | 1.60 | 1.50 | 1.50 | 1.43 | 1.00 | 1.41 |
| Leukemia: Symptoms And Treatment (The Reyv) | 1.50 | 1.75 | 1.17 | 1.43 | 1.13 | 1.39 |
| Myeloma Cancer Guide (Everyone Learning Apps) | 1.20 | 1.75 | 1.50 | 1.50 | 1.00 | 1.39 |
| Leukemia Disease Treatment (Woochi Developer) | 1.40 | 1.50 | 1.17 | 1.43 | 1.25 | 1.35 |
| Blood Cancer (Digital Planete Space) | 1.40 | 1.71 | 1.17 | 1.29 | 1.00 | 1.31 |
| How To Cure Leukemia (Apps How To Apps) | 1.10 | 1.75 | 1.17 | 1.14 | 1.00 | 1.23 |
| Blood Cancer tips (Free Apps For Everyone) | 1.00 | 1.63 | 1.17 | 1.14 | 1.00 | 1.19 |
| Blood-Count Reader free (Yurii Shevchenko) | 1.80 | 1.13 | 1.00 | 1.00 | 1.00 | 1.19 |

Comparison by app distribution platform (Apple App Store and Android Play Store) revealed a mean MARS score of 3.85 (SD 0.35) for apps developed for iOS (n=34) and 2.67 (SD 0.30) for apps developed for Android (n=62), resulting in a statistically significant difference ($P$<.001). Apps whose development had been supported by a health organization obtained better scores (mean 3.75, SD 0.29; n=23) than those that had not (mean 2.78, SD 0.31; n=65; $P$<.001). Finally, another statistically significant difference ($P$=.049) was found when the overall MARS scores were analyzed considering whether the apps were free (mean 2.97, SD 0.30; n=82) or required payment (mean 3.92, SD 0.29; n=6; $P$=.049). The comparison by different characteristics is shown in Table 6.

The mean κ coefficient score for the five MARS domains was 0.78. κ values between 0.61 and 0.81 indicate that interrater agreement between the 2 reviewers was substantial. The only item with a score less than 0.61 was ease of use (Table 7).

**Table 6.** Results of the Mobile App Rating Scale evaluation: comparison by different characteristics.

| Category | Operating system | | | Developer | | | Cost | | |
|---|---|---|---|---|---|---|---|---|---|
| | Android (n=62), score | iOS (n=34), score | *P* value | No health organization (n=65), score | Health organization (n=23), score | *P* value | Free (n=82), score | Payment (n=6), score | *P* value |
| Engagement | 2.59 | 4.16 | <.001 | 2.76 | 3.88 | <.001 | 2.98 | 4.12 | .04 |
| Functionality | 3.09 | 4.01 | <.001 | 3.23 | 4.02 | .002 | 3.38 | 4.29 | .04 |
| Aesthetics | 2.66 | 3.91 | <.001 | 2.82 | 3.90 | .002 | 3.04 | 3.89 | .10 |
| Information | 2.64 | 3.52 | <.001 | 2.70 | 3.67 | <.001 | 2.90 | 3.70 | .08 |
| Subjective quality | 2.26 | 3.34 | <.001 | 2.37 | 3.29 | .002 | 2.54 | 3.58 | .052 |
| Overall | 2.67 | 3.85 | <.001 | 2.78 | 3.75 | <.001 | 2.97 | 3.92 | .049 |

**Table 7.** Kappa score and interrater reliability for the Mobile App Rating Scale domains.

| Domain | Weighted Cohen κ | Agreement, % |
|---|---|---|
| **Engagement** | 0.82 | 93.1 |
| Entertainment | 0.63 | 86.7 |
| Interest | 0.72 | 90.4 |
| Customization | 0.90 | 95.2 |
| Interactivity | 0.84 | 92.6 |
| Target group | 0.73 | 90.4 |
| **Functionality** | 0.69 | 90.6 |
| Performance | 0.67 | 88.5 |
| Ease of use | 0.54 | 87.1 |
| Navigation | 0.64 | 87.9 |
| Gestural design | 0.71 | 90.3 |
| **Aesthetics** | 0.80 | 93.1 |
| Layout | 0.76 | 91.7 |
| Graphics | 0.76 | 90.7 |
| Visual appeal | 0.78 | 92.4 |
| **Information** | 0.80 | 93.5 |
| Accuracy of the app in the description (Apple App Store and Android Play Store) | 0.77 | 93.1 |
| Goals | 0.78 | 92.8 |
| Quality of information | 0.73 | 91.0 |
| Quantity of information | 0.67 | 88.6 |
| Visual information | 0.63 | 86.4 |
| Evidence base | 0.91 | 96.3 |
| Credibility | 0.84 | 94.3 |
| **Subjective quality** | 0.80 | 92.8 |
| Would you recommend this app to people who might benefit from it? | 0.78 | 91.5 |
| Would you pay for this app? | 0.86 | 94.2 |
| How many times do you think you would use this app in the next 12 months if it was relevant to you? | 0.77 | 90.3 |
| What is your overall star rating of the app? | 0.79 | 92.2 |

# Discussion

## Principal Findings

This is the first study to provide a systematic search and ranking of apps for patients diagnosed with hematological conditions available in the Apple App Store and Android Play Store, using the MARS as a standardized methodology for the classification, assessment, and validation of these apps.

We found that there were more apps available in the Android Play Store than in the Apple App Store, as mentioned in other studies [8,11,24], which can imply that uploading an app into the Android Play Store is an easier process. We observed that almost half of the apps (43/88, 49%) had been updated in the last year, as previously reported [25]. Considering hematology as a medical field that is constantly growing in complexity and extending its therapeutic arsenal, this low rate of app content actualization is insufficient [26].

Of 88 apps, only 23 (26%) were designed with the participation of some kind of health organization. The absence of health care professionals in the development of health apps continues to be raised time and time again. Amor-García et al [11] observed that only 15.2% of apps for patients with genitourinary cancer involved health professionals in their design process. When reviewing apps for medication management, Tabi et al [27] observed a similar result (14.6%). It would seem crucial that health care professionals be involved in the creation of medical apps; however, this scarcely happens. Moreover, the fact that most health-related apps are free favors accessibility [27].

Our results expose the high prevalence of informative apps (60/88, 68%), as reported by other authors [6,11]. The majority of these apps provide generic data about one or several pathologies, including symptoms, diagnostics, and treatment, focusing solely on education. One-third of the total of informative apps is intended for patients with anemia, which

highlights the interest in anemia self-management, as it is the most common blood disorder globally [18]. Preventive apps are less numerous (23/88, 26%), although their quality and performance are significantly higher. These apps focus on handling the pathology after diagnosis, allowing for treatment and laboratory values tracking and recording of symptoms and adverse events. We found these types of apps the most appropriate and useful for patients with hematological conditions because many blood conditions require chronic and complex pharmacologic treatment [28,29]. Only 5 diagnostic apps were evaluated. It is worth mentioning *STB—Stop The Bleed*, an app designed to help anyone learn how to safely and effectively deal with life-threatening bleeding, which has demonstrated the potential of mobile apps in emergency scenarios [30]. The other 4 diagnostic apps are screening tools based on hematological parameters, questionnaires, and gene traits. Its objectives are to predict blood groups or certain hereditary pathologies, such as hemophilia or thalassemia. The main limitation is again the lack of evidence-based content, which in this case could mislead patients into not seeking professional advice. The potential of apps to be implemented as remote diagnostic tools for hematological conditions is very high. This is the case of AnemoCheck Mobile, an app that estimates hemoglobin levels by analyzing the color of fingernail beds and detects anemia, serving as a completely noninvasive anemia screening tool [31].

The MARS has demonstrated its potential as a simple, reliable, and flexible health care app-quality rating scale [21]. It analyzes the quality of an app by evaluating 23 items, grouped into 5 domains, and rating on a 5-point scale. Our study showed a mean score of 3.03, considering a score of 3 as *acceptable*. This result is similar to the scores showed by other authors using the MARS to evaluate health apps for other conditions. The mean score found by Salazar et al [5] for apps designed for chronic pain management was 3.17, and Kwan et al [6] showed a mean score of 3.48 for apps targeted at patients with spondyloarthritis, out of 18 and 5 apps evaluated, respectively. Knitza et al [24] reviewed 28 rheumatology apps and obtained an overall MARS score of 3.85. The median overall MARS score of the analysis of 34 apps targeted toward supporting heart failure symptom monitoring was 3.4 [32]. In a larger sample study, Amor-García et al [11] evaluated 46 apps for patients with genitourinary cancers and found a mean score of 2.98. It is worth noting that our study encompasses a higher number of apps evaluated than any of the studies cited. Thus, the overall quality of health apps in digital platforms is moderate, and there remains considerable scope for improvement. Of the 88 apps, 41 (47%) hematological apps obtained a score of at least 3 points, meaning that more than half of the apps for hematological conditions do not meet acceptable criteria for quality and content. Moreover, of the 88 apps, only 28 (32%) exceeded 4 points in the overall score.

MARS ratings ranged from 1.19 (*Blood Count Reader*) to 4.86 (*Multiple Myeloma Manager*), indicating the highly inconsistent quality of apps. The apps with the highest scored were *Multiple Myeloma Manager*, *Hodgkin Lymphoma Manager*, *Focus On Lymphoma*, *ALL Manager*, and *CLL Manager*. All of them were exclusive to the Apple App Store, except *Focus On Lymphoma*, which was available in both platforms. These apps showed high scores in the engagement and functionality domains. The main characteristic that defines these top-rated apps was the active patient participation, offering wide treatment and symptom monitoring options, reminders, and schedules edition. The five apps with the highest score had a plain preventive purpose, whereas informative apps scored lower on the MARS despite being more frequent.

The comparison by operating system showed a statistically significant difference favoring iOS apps over Android apps in all 5 MARS domains, a tendency that has been observed in a similar evaluation about genitourinary apps [11]. The reason could be that the Apple App Store has stricter standards to include apps.

Although we observed that only 26% (23/88) of the apps involved the participation of health professionals in their design, their quality was significantly higher. The lack of health professional involvement is a constant that has already been highlighted by several authors, expressing their concern about app content and compromising patient safety [33-35]. However, 4 of the best apps (*Multiple Myeloma Manager*, *Hodgkin Lymphoma Manager*, *ALL Manager*, and *CLL Manager*) were developed by @Point of Care, a platform consisting of nonmedical stakeholders and dedicated to creating medical apps for patients and clinicians. @Point of Care has designed apps focused on diverse pathologies, some of them obtaining considerably high MARS scores in other studies similar to ours [11]. The analysis by cost revealed another statistically significant difference, positioning payment apps ahead of free apps in terms of quality, although the fact that only 6 hematological apps were not free and all of them were developed for iOS can destabilize the comparison.

Functionality was the domain that scored the highest on the MARS test, as described by other authors [11,36]. This implies that the apps are easy to navigate and efficient. Leaving subjective quality aside, engagement and information were the domains with the lowest MARS scores. Engagement reflects the capacity of the app to be personalized by the user. Patients usually search for a health app that allows for medication management, clinical and analytical parameter register, and symptom tracking [28,37]. Patients with hematological conditions would benefit significantly from this type of assistance, as several blood conditions demand constant patient monitoring and high adherence to treatment for a better health outcome [18,30]. *My INR*, *INR Care*, and *Diario de INR* are apps that allow anticoagulated patients to record and track their international normalized ratio readings and antivitamin K dosages. They could help improve adherence and avoid potential complications, such as the risk of bleeding or clots. *HaemActive* is a fitness app especially tailored to patients with hemophilia, who require special exercises that imply a minimal risk of bleeding. The app includes weekly training planning, explainer videos, and easy customization. In addition, patients expressed their interest in using health apps to communicate with their physicians [28,38]. Concerning the information domain, there is 1 specific item assessing the *evidence base*, which explores the extent to which the app has been scientifically tested. However, this item was excluded from all calculations, as no clinical studies to support the effectiveness and safety of any of the apps could be found. Thus, empirical studies should be

conducted for apps to determine their clinical impact on outcomes for patients diagnosed with hematological conditions [13].

### Recommendations for Health App Development

The number of health apps available and studies reviewing their quality is steadily growing, which will help health professionals to recommend apps to patients. This activity acquires even further relevance, considering the still little control from regulatory authorities over health app development. We have observed that the main issues that need to be addressed when designing health apps are as follows: no participation of health organizations in app development, questionable sources of information, and deficient interactivity and personalization options [35]. Production of medical apps from nonmedical stakeholders has benefits in terms of creativity in the design of apps. However, it must be combined with clinician assistance to boost the credibility of medical information with such apps. Concerning patients with hematological conditions, registering analytical information, treatment prescribed, and symptoms is highly recommendable for apps to help them in their care.

### Limitations

First, only apps available in the Android Play Store and Apple App Store, with contents in English or Spanish and accessed from a Spanish IP address were included, assuming the possibility of having missed some other apps dedicated to hematological conditions. Another limitation could be that app quality was assessed using the MARS, which is limited by the subjectivity of the evaluators. Nevertheless, this issue is partially addressed by the high interrater reliability of the data analyzed by the 2 independent researchers. We believe that this evaluation should allow health care professionals and patients to identify which apps meet minimum standards of quality and safety in their content.

### Conclusions

We provide the first systematic review of apps related to hematological conditions, identifying 88 apps and rating them using the MARS. The study shows great heterogeneity among their quality. Many of these apps emerge as tools for consulting information, being the most frequent functionality, although not the highest rated. A very small number of them offer a comprehensive self-management approach incorporating evidence-based strategies. Only 26% (23/88) of the apps were developed with the assistance of health care professionals. The top 5 rated apps—*Multiple Myeloma Manager*, *Hodgkin Lymphoma Manager*, *Focus On Lymphoma*, *ALL Manager*, and *CLL Manager*—allowed for active patient participation and app personalization. Higher scores in quality were observed in iOS apps, apps developed by health organizations, and payment apps.

### Conflicts of Interest

None declared.

### References


1. World Health Organization. mHealth Economics 2017: current status and future trends in mobile health. Research2Guidance. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2021-05-05]
2. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 1st quarter 2021. Statista. 2021. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2021-07-07]
3. Kebede MM, Pischke CR. 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;10:135 [FREE Full text] [doi: 10.3389/fendo.2019.00135] [Medline: 30881349]
4. Kalhori SR, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of English mobile applications for gestational diabetes: app review using mobile application rating scale (MARS). Curr Diabetes Rev 2021;7(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
5. 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]
6. Kwan YH, Ong WJ, Xiong M, Leung YY, Phang JK, Wang CT, et al. Evaluation of mobile apps targeted at patients with spondyloarthritis for disease monitoring: systematic app search. JMIR Mhealth Uhealth 2019 Oct 28;7(10):e14753 [FREE Full text] [doi: 10.2196/14753] [Medline: 31661080]
7. Terhorst Y, Rathner E, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
8. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
9. Singh HJ, 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]

10. Salehinejad S, Kalhori SR, Gohari SH, 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]
11. Amor-García MA, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arija A, et al. Assessing apps for patients with genitourinary tumors using the mobile application rating scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
12. Ali EE, Chan SS, Leow JL, Chew L, Yap KY. User acceptance of an app-based adherence intervention: perspectives from patients taking oral anticancer medications. J Oncol Pharm Pract 2019 Mar;25(2):390-397 [FREE Full text] [doi: 10.1177/1078155218778106] [Medline: 29792123]
13. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
14. Burns R, Leal J, Sullivan R, Luengo-Fernandez R. Economic burden of malignant blood disorders across Europe: a population-based cost analysis. Lancet Haematol 2016 Aug;3(8):362-370. [doi: 10.1016/S2352-3026(16)30062-X] [Medline: 27476787]
15. Boström EA, Lira-Junior R. Non-malignant blood disorders and their impact on oral health: an overview. Curr Oral Health Rep 2019 Apr 24;6(2):161-168. [doi: 10.1007/s40496-019-0211-9]
16. Atrash HK, Parker CS. The public health response to blood disorders. Am J Prev Med 2010 Apr;38(4 Suppl):451-455. [doi: 10.1016/j.amepre.2010.01.006] [Medline: 20331942]
17. Boccalandro EA, Dallari G, Mannucci PM. Telemedicine and telerehabilitation: current and forthcoming applications in haemophilia. Blood Transfus 2019 Sep;17(5):385-390 [FREE Full text] [doi: 10.2450/2019.0218-18] [Medline: 30747703]
18. Mann D, Riddell L, Lim K, Byrne LK, Nowson C, Rigo M, et al. Mobile phone app aimed at improving iron intake and bioavailability in premenopausal women: a qualitative evaluation. JMIR Mhealth Uhealth 2015 Sep 28;3(3):e92 [FREE Full text] [doi: 10.2196/mhealth.4300] [Medline: 26416479]
19. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. NPJ Digit Med 2021 Mar 10;4(1):46 [FREE Full text] [doi: 10.1038/s41746-021-00410-x] [Medline: 33692488]
20. 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]
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. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
24. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
25. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
26. Finnes HD. 2019-2020 Drug updates in hematologic malignancies. J Adv Pract Oncol 2021 Apr;12(3):279-283 [FREE Full text] [doi: 10.6004/jadpro.2021.12.3.12] [Medline: 34084575]
27. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
28. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
29. Shah N, Jonassaint J, De CL. Patients welcome the sickle cell disease mobile application to record symptoms via technology (SMART). Hemoglobin 2014;38(2):99-103. [doi: 10.3109/03630269.2014.880716] [Medline: 24512633]
30. Roncero AP, Marques G, Sainz-De-Abajo B, Martín-Rodríguez F, Vegas CD, Garcia-Zapirain B, et al. Mobile health apps for medical emergencies: systematic review. JMIR Mhealth Uhealth 2020 Dec 11;8(12):e18513 [FREE Full text] [doi: 10.2196/18513] [Medline: 33306037]
31. Mannino RG, Myers DR, Tyburski EA, Caruso C, Boudreaux J, Leong T, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. Nat Commun 2018 Dec 04;9(1):4924 [FREE Full text] [doi: 10.1038/s41467-018-07262-2] [Medline: 30514831]

32. Masterson CR, 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]
33. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
34. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
35. Morillo-Verdugo R, Margusino-Framiñán L, Monte-Boquet E, Morell-Baladrón A, Barreda-Hernández D, Rey-Piñeiro XM, et al. Spanish society of hospital pharmacy position statement on telepharmacy: recommendations for its implementation and development. Farm Hosp 2020 Jul 01;44(4):174-181 [FREE Full text] [doi: 10.7399/fh.11515] [Medline: 32646349]
36. Jupp JC, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer 2018 Nov;65(11):e27278. [doi: 10.1002/pbc.27278] [Medline: 29943893]
37. Martínez-Bautista MJ, Mangues-Bafalluy I, Cajaraville-Ordoñana G, Carreras-Soler MJ, Clopés-Estela A, Moreno-Martínez E. Survey of oncohematological pharmaceutical care situation in Spain. Farm Hosp 2019 Nov 01;43(6):194-201 [FREE Full text] [doi: 10.7399/fh.11239] [Medline: 31705643]
38. Collado-Borrell R, Escudero-Vilaplana V, Ribed A, Gonzalez-Anleo C, Martin-Conde M, Romero-Jimenez R, et al. Effect of a mobile app for the pharmacotherapeutic follow-up of patients with cancer on their health outcomes: quasi-experimental study. JMIR Mhealth Uhealth 2020 Oct 16;8(10):e20480 [FREE Full text] [doi: 10.2196/20480] [Medline: 33064100]


## Abbreviations

**MARS:** Mobile App Rating Scale
**PRISMA-P:** Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols


*Edited by L Buis; submitted 03.09.21; peer-reviewed by R Romero-Jimenez, B Chaudhry; comments to author 18.10.21; revised version received 28.10.21; accepted 10.12.21; published 16.02.22.*

*Please cite as:*
*Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García MÁ, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, Gómez-Centurión I, Herranz-Alonso A, Sanjurjo-Sáez M*
*Mobile Apps for Hematological Conditions: Review and Content Analysis Using the Mobile App Rating Scale*
*JMIR Mhealth Uhealth 2022;10(2):e32826*
*URL: https://mhealth.jmir.org/2022/2/e32826*
*doi:10.2196/32826*
*PMID:35171109*

Original Paper

# Classification of Smoking Cessation Apps: Quality Review and Content Analysis

Suin Seo[1], MPH; Sung-Il Cho[1], MD; Wonjeong Yoon[1], MPH; Cheol Min Lee[2], MD

[1]Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

[2]Department of Family Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea

**Corresponding Author:**
Sung-Il Cho, MD
Department of Epidemiology, Graduate School of Public Health
Seoul National University
1, Gwanak-ro, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 2 880 2717
Fax: 82 2 743 8240
Email: persontime@hotmail.com

## Abstract

**Background:** Many people use apps for smoking cessation, and the effectiveness of these apps has been proven in several studies. However, no study has classified these apps and only few studies have analyzed the characteristics of these apps that influence their quality.

**Objective:** The purpose of this study was to analyze the content and the quality of smoking cessation apps by type and identify the characteristics that affect their overall quality.

**Methods:** Two app marketplaces (App Store and Google Play) were searched in January 2018, and the search was completed by May 2020. The search terms used were “stop smoking,” “quit smoking,” and “smoking cessation.” The apps were categorized into 3 types (combined, multifunctional, and informational). The tailored guideline of Clinical Practice Guideline for Treating Tobacco Use and Dependence was utilized for evaluating app content (or functions), and the Mobile App Rating Scale (MARS) was used to evaluate the quality. Chi-square test was performed for the general characteristics, and one-way analysis of variance was performed for MARS analysis. To identify the general features of the apps that could be associated with the MARS and content scores, multiple regression analysis was done. All analyses were performed using SAS software (ver. 9.3).

**Results:** Among 1543 apps, 104 apps met the selection criteria of this study. These 104 apps were categorized as combined type (n=44), functional type (n=31), or informational type (n=29). A large amount of content specified in the guideline was included in the apps, most notably in the combined type, followed by the multifunctional and informational type; the MARS scores followed the same order (3.64, 3.26, and 3.0, respectively). Regression analysis showed that the sector in which the developer was situated and the feedback channel with the developer had a significant impact on both the content and MARS scores. In addition, problematic apps such as those made by unknown developers or copied and single-function apps were shown to have a large market share.

**Conclusions:** This study is the first to evaluate the content and quality of smoking cessation apps by classification. The combined type had higher-quality content and functionality than other app types. The app developer type and feedback channel with the app developer had a significant impact on the overall quality of the apps. In addition, problematic apps and single-function apps were shown to have a large market share. Our results will contribute to the use and development of better smoking cessation apps after considering the problems identified in this study.

***(JMIR Mhealth Uhealth 2022;10(2):e17268)*** doi:10.2196/17268

**KEYWORDS**

smoking cessation; app; type; content and functions; MARS; quality; score; mobile phone

## Introduction

Smoking is a repetitive addictive behavior [1,2], and environmental conditions promoting its cessation are important. A smoke-free environment achieved through antismoking campaigns and use of assistive devices such as computers and smartphones throughout the day could be helpful [3-5]. In particular, smartphones allow continuous monitoring of smoking and facilitate cessation because users have access to them at all times. Approximately 8.2 billion mobile cellular telephones, including smartphone subscriptions, were reported worldwide in 2020, which exceeds the world's population [6]. Apps are the most common features of smartphones. In the first quarter of 2021, Google Play had 3.48 million apps available for download and the iPhone App Store had 2.22 million apps [7]. In the first quarter of 2021, the total downloads from App Store and Google Play app amounted to an estimated 36.6 billion [8]. This illustrates the degree to which smartphone apps are used. Thus, a firmly established smoking cessation app could have a major impact.

Smartphone apps have been increasingly used for promoting behavioral changes and have proved to be effective with regard to physical activity and noncommunicable diseases. They greatly influence people's daily lives, particularly with respect to goal-based behavioral modifications [9-11]. As the number of people using a smartphone app to quit smoking has increased, there has been a concomitant increase in the number of smoking cessation apps available [12]. In a comprehensive review of studies evaluating the effects of smoking cessation apps, it was found that an evidence to determine the effect of the apps was not enough [13]. However, a few smoking cessation apps have been proven to be effective. Some studies have found the effects of individual apps with distinct characteristics. Smokers who use decision-aid apps are more likely to be continuously abstinent compared to those using information-only apps at 1 month (relative risk [RR] 1.68, 95% CI 1.25-2.28), 3 months (RR 2.08, 95% CI 1.38-3.18), and 6 months (RR 2.02, 95% CI 1.08-3.81) [14]. An evidence-based app with customized functions and information is more effective than a web-based self-help booklet for smoking cessation [15]. Thus, apps providing information and various functions are more effective than apps that provide information only.

Although many studies have evaluated smoking cessation apps and compared the effects between individual apps with distinct characteristics, no study has classified all the apps by their characteristics. The objective of this study was to evaluate the content and quality of smoking cessation apps by type and determine which type of apps have better content and quality. Furthermore, we aimed to identify the app characteristics associated with the content (or functions) and quality thereof. To this end, first, we identify the types of smoking cessation apps. Second, we determine which type of smoking cessation apps have more accurate content and high quality for continuous use.

## Methods

### App Search Strategy

Two app markets (App Store and Google Play), which account for over 80% of all the available mobile apps, were searched for this study [16]. English-language apps available to Koreans were searched by using the terms "stop smoking," "quit smoking," and "smoking cessation," as done in previous studies [12,17-19]. The app search was completed by May 2020.

### App Selection

We reviewed all the smoking cessation apps that we obtained with our search terms. We excluded apps that were not concerned with smoking cessation, not in English, not designed for smartphones (eg, Tablet and iPad apps), or not showing proper functionality, as well as those targeting specific groups, designed for commercial purposes, related to the overall health behavior and not just smoking cessation, or including only photos, videos, or games. As the purpose of this study was to compare apps by function, single-function apps that have only 1 function (eg, counter or hypnosis) were excluded. Apps with only 2 assistive functions or 1 function plus information were also excluded from the analysis because the functions of these apps were very limited similar to single-function apps. The main function of these apps is "counter (tracker)" and the other function is very minor (eg, free notes, little information, unidentified chatting). Since these criteria are not applicable in many items of evaluation tools, evaluation has no meaning. In the analysis of content and function, questions for an "advise" category, an "assess" category, and an "assist category-support provided" do not apply to single-function apps (eg, personalized advice, user could indicate lack of readiness to quit, users could interact with other users for mutual support [app community]). In the analysis of quality (using Mobile App Rating Scale [MARS]), some questions of the information quality category do not apply to single-function apps (eg, Credibility: Does the app come from a legitimate source? specified in App Store description or within the app itself). Apps that met our criteria were included in the final analysis (Multimedia Appendix 1).

### App Categorization

The apps were categorized into 3 types at the eligibility assessment stage by reference to prior studies [14,15]. Those 2 studies compared information-only apps (self-help booklets) with apps, including both information and specific functions to motivate the user to stop smoking. The effectiveness of the information-only apps for quitting smoking was lower than that of the apps consisting of both information and specific functions. It is necessary to classify the overall smoking cessation apps as a feature confirmed in the individual app effect evaluation. Therefore, in this study, we initially classified apps into 2 types: information-only apps and apps with functions. The availability of information has a significant impact on health behavior changes [11]. The apps with functions were further subdivided into apps that provided information and apps that did not, thereby resulting in 3 categories of apps in our study. The 3 categories were formally designated as informational, multifunctional, and combined type. Informational apps provide information only similar to an electronic book and have no

function (eg, counters, alarms, games, community features). Multifunctional apps have 3 or more assistive functions but provide little information. Combined apps provide both information and at least 2 assistive functions.

### Critical Appraisal of the Apps (Quality Assessment)

Two evaluation tools were used to assess the apps. A tailored guideline [20] for use of the 5As (Ask, Advise, Assess, Assist, and Arrange follow-ups), as recommended in the Clinical Practice Guideline for Treating Tobacco Use and Dependence [21], was modified to evaluate the content and functions of the apps [22]. The MARS was used to evaluate the functional quality of the apps [23]. The apps were rated by 2 independent researchers using a standardized rating form. After using each app for a minimum of 15 minutes, raters evaluated them twice, that is, once each using a content and function analysis form and a MARS form. Apps on both markets were analyzed using both iPhones and Android phones for including all functions. Each assessment took approximately 40-50 minutes. The consistency of the assessments was measured according to the interrater reliability. Discrepancies between content analyses were resolved by consensus between the 2 raters. This consensus process was not necessary for the MARS analysis because the MARS score was the average of the 2 assessments.

### Analysis of Content and Function

The revised guideline for evaluating apps [20] from the Clinical Practice Guide for Treating Tobacco Use and Dependence [21] was used for content and function analysis. This guideline breaks down the interventions into 5 major parts: ask, advise, assess, assist, and arrange follow-ups (the 5As). In this study, in the "ask" category, the smoking status of the users was confirmed. In the "advise" category, the raters identified whether or not the app included advice on quitting smoking. In the "assess" category, the raters identified whether the app possessed a function to evaluate the user's readiness for quitting smoking. In the "assist" category, the various functions of the apps designed to help users quit smoking were evaluated. Finally, in the "arrange follow-ups category," raters identified whether the app could track the user's smoking cessation status. By referring to the World Health Organization's "A guide for tobacco users to quit" [22], we added new content to the "assist" category, namely, social benefits, health risk of smoking, and confidence (motivation). The modified guideline for content and function analysis consisted of 39 questions. Each question used a yes/no scale; therefore, the total possible score was 39 points. The interrater reliability between 2 raters was assessed and the kappa value was 0.75.

### Analysis of Quality (Using MARS)

The MARS is a mobile health (mHealth) app quality assessment tool that provides a multidimensional measure of app quality according to the following indicators: engagement, functionality, aesthetics, and information quality. It is a 23-item, expert-based rating scale that can be used to systematically evaluate the quality of mHealth apps on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) [23]. The engagement category is concerned with factors such as fun, interest, customizability, interactivity (eg, sending alerts, messages, reminders and feedback, sharing), and suitability of material. The functionality category is concerned with technical functions such as functioning, ease of learning, navigation, flow, logic, and gestural design. The aesthetics category subsumes graphic design, visual appeal, color scheme, and stylistic consistency. Information quality is concerned with whether the information (eg, text, feedback, measurements, references) is from a credible source. The interrater reliability between the 2 raters was assessed by the intraclass correlation and had a value of 0.6 (95% CI 0.4-0.74).

### Statistical Analyses

General characteristics were analyzed by the chi-square test. Each dimension on MARS was analyzed by one-way analysis of variance The analyses of content score (ie, the score of content and function analysis) and MARS scores (ie, the score of MARS analysis) were conducted by app type. To identify characteristics of apps associated with the MARS and content scores, multiple regression analysis was performed. The sector in which the developer was situated, whether they had an affiliation with health care professionals, the app platform, payment type, and feedback type represented the characteristics of interest. All analyses were performed using SAS software (ver 9.3; SAS Institute).

## Results

### General Characteristics of the Apps by Type

A total of 1543 apps (App Store, n=701; Google Play, n=842) were identified via the search terms, of which 940 duplicated apps were excluded. Thus, 603 apps (App Store, n=305; Google Play, n=260; both markets, n=38) were preliminarily screened, and 212 irrelevant apps were excluded. The remaining 391 relevant apps (App Store, n=174; Google Play, n=181; both markets, n=36) were screened according to selection criteria. Finally, 104 apps (App Store, n=30; Google Play, n=39; both markets, n=35) were included in the analysis (Figure 1). Table 1 shows the general characteristics of the apps by type. Of the 104 apps assessed, there were 44 combined apps, 31 multifunctional apps, and 29 informational apps.

**Figure 1.** Study flow diagram.

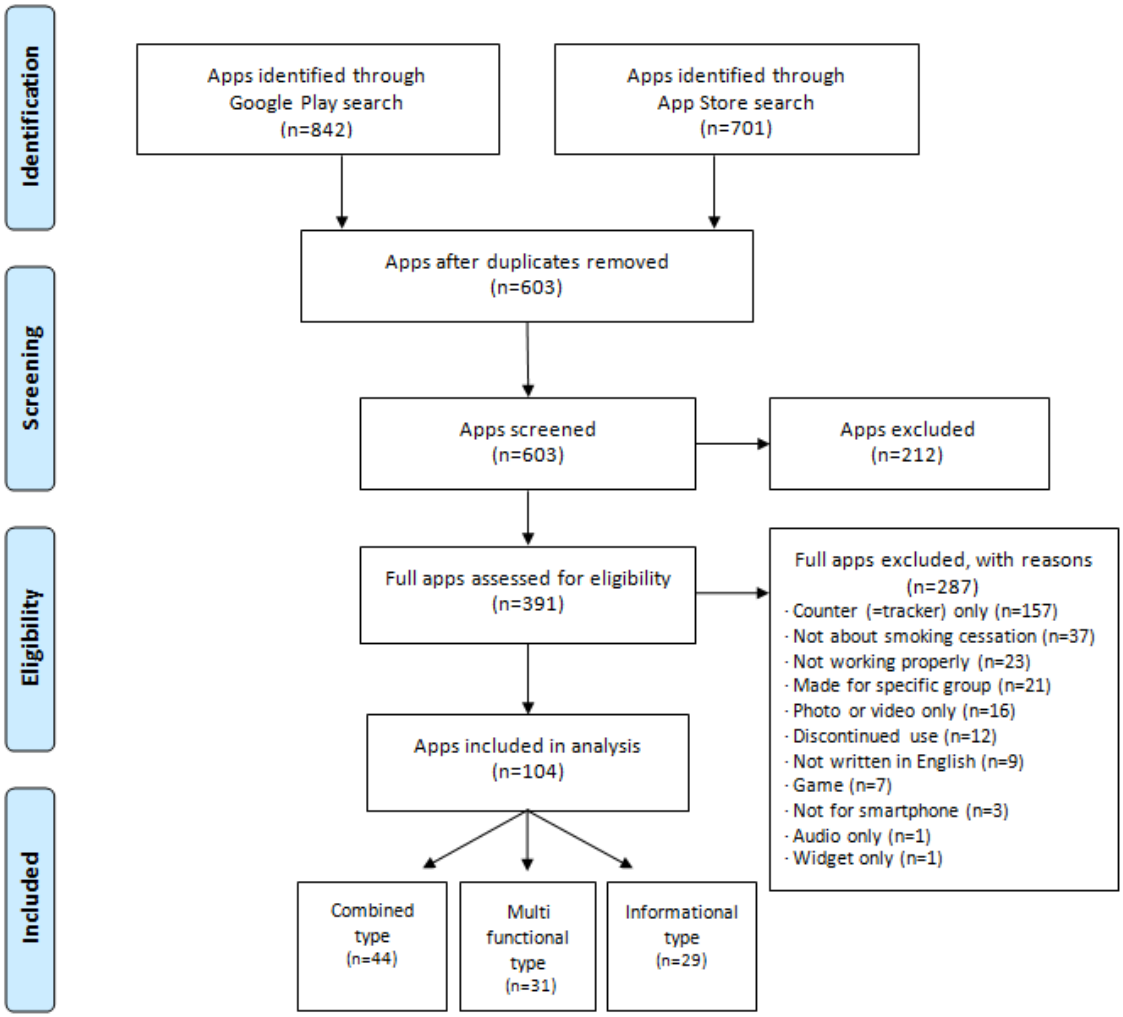


**Table 1.** General characteristics of the apps by type (N=104).

| Category, items | Combined type (n=44) | Multifunctional type (n=31) | Informational type (n=29) | $\chi^2$ *(df)* | *P* value |
|---|---|---|---|---|---|
| **Platform, n (%)** | | | | 18.1 (4) | .001 |
| Android only | 10 (23) | 14 (45) | 15 (52) | | |
| iPhone only | 12 (27) | 6 (19) | 12 (41) | | |
| Both | 22 (50) | 11 (36) | 2 (7) | | |
| **Last update[a], n (%)** | | | | 14.3 (8) | .07 |
| <1 month | 7 (16) | 5 (16) | 0 (0) | | |
| 1-6 months | 8 (18) | 7 (23) | 4 (14) | | |
| 6-12 months | 13 (30) | 9 (29) | 7 (24) | | |
| >12 months | 14 (32) | 8 (26) | 10 (34) | | |
| Unknown | 2 (5) | 2 (7) | 8 (28) | | |
| Advertisements in the app, n (%) | 9 (20) | 14 (45) | 17 (59) | 14.3 (4) | .007 |
| **Payment type, n (%)** | | | | 25.6 (4) | <.001 |
| Free | 29 (66) | 16 (52) | 19 (66) | | |
| In-app purchase | 15 (34) | 10 (32) | 0 (0) | | |
| Prepaid | 0 (0) | 5 (16) | 10 (34) | | |
| Price (prepaid app)[b], mean (SD) | N/A[c] | 2330 (2730) | 3910 (8360) | N/A | N/A |
| **Developer sector, n (%)** | | | | 57.6 (10) | <.001 |
| Government or university | 13 (30) | 0 (0) | 1 (3) | | |
| Government with commercial | 4 (9) | 0 (0) | 0 (0) | | |
| Commercial | 21 (48) | 20 (65) | 4 (14) | | |
| Nongovernment organization | 0 (0) | 2 (6) | 0 (0) | | |
| Unknown | 6 (14) | 9 (29) | 24 (83) | | |
| Affiliation of developer with health care professionality, n (%) | 25 (57) | 2 (6) | 1 (3) | 34.7 (2) | <.001 |
| **Feedback channel with developer, n (%)** | | | | 20.6 (4) | <.001 |
| Within the app | 19 (43) | 13 (42) | 2 (7) | | |
| Contact information of developer provided | 11 (25) | 3 (10) | 3 (10) | | |
| Market level only | 14 (32) | 15 (48) | 24 (83) | | |

[a]Updated on May 2020.

[b]US $1=1100 won; the cost of in-app purchases was calculated as the average cost if prices differed among items.

[c]N/A: not applicable.

The platform, presence of advertisements in the app, payment type, developer sector and health professional affiliation status, and feedback channel with developer showed significant differences by app type. Regarding differences in platforms, of the 44 combined apps, 22 (50%) were available in both Google Play and the App Store. Of the remainder, more were available in the App Store (12/44, 27%) than in Google Play (10/44, 23%). Of the 31 multifunctional apps, 11 (36%) were available in both markets. The remainders were more frequently available in Google Play (14/31, 45%) than in App Store (6/31, 19%) in contrast to the combined type. Informational apps were least frequently available in both markets (2/29, 7%); 15 (52%) of the 29 informational apps were available in Google Play only, and 12 (41%) were available in the App Store only. Advertisements were found in 17 (59%) of the 29 informational type, 14 (45%) of the 31 multifunctional type, and 9 (20%) of the 44 combined type. With respect to payment for the apps, all of the 44 combined apps were initially free, of which 15 (34%) required payment within the app to utilize all functions (in-app purchase). Regarding the 31 multifunctional apps, 5 (16%) were purchased and 26 (84%) were free. Of the 31 free multifunctional apps, 10 (32%) had in-app purchase functionality. Regarding the 29 informational apps, 10 (34%) were prepaid, and in contrast to the other 2 categories of apps, none of the free informational apps (19/29, 66%) had in-app purchase functionality.

Regarding developer sector, for the combined type, the largest proportion of the apps was developed within the private sector (21/44, 48%) followed by government or university independently (13/44, 30%) and then by collaboration between the government and a professional app development company (4/44, 9%). For 6 (14%) of the 44 combined type apps, no developer was identified. Regarding the 31 multifunctional apps, the majority were commercially developed (20/31, 65%), and only 2 (6%) apps were developed by nongovernment organizations. For nearly a third (9/31, 29%) of the multifunctional apps, the sector in which they were developed was not known. For informational apps, only a small proportion of the developers identified were commercial developers (4/29, 14%) and only 1 app (1/29, 3%) was developed by government. For the majority of the informational apps (24/29, 83%), the affiliation of the developer was unknown. Apps developed by health professionals were most prevalent in the combined type (25/44, 57% vs 2/31, 6% and 1/29, 3% for the multifunctional and informational type, respectively).

When we assessed the apps in terms of the feedback channel with the developer, a significant difference between app types was seen again. Among the combined type, apps that had an option to communicate with a developer within the apps were the most common (19/44, 43%); 14 (32%) did not have the option and communication was only possible at the market level (ie, Google Play or App Store) and 11 (25%) included the developer's contact information within the apps or as a link. Among the multifunctional type, the largest proportion (15/31, 48%) of the apps enabled communication with a developer only at the market level, while 13 (42%) provided the option to communicate using the apps and 3 (10%) included the developer's contact details within the apps or as a link. Among the informational type, the vast majority of apps enabled communication with a developer only at the market level (24/29, 83%); only 3 (10%) included the developer's contact details within the apps or as a link, and 2 (7%) apps enabled communication with a developer directly through the apps.

## Content and Function of the Apps by Type

Table 2 shows how well the content and functions of the apps matched the guideline. For all 5As components, the content of combined apps was the most consistent with the guideline. Regarding the "ask," "assist," and "arrange follow-up" components, multifunctional apps were more consistent with the guideline than informational apps. Regarding the "advice" and "assess" components, informational apps were more consistent with the guideline than multifunctional apps. All apps addressed the "assist" component. "Arrange follow-ups" was offered in some respect by the 44 combined apps (44/44, 100%), 30 multifunctional apps (30/31, 97%), and 10 informational apps (10/29, 34%); a substantial proportion of the apps asked smoking status (ask; 44/44, 100%; 30/31, 97%; and 0/29, 0%; respectively) and offered advices to quit smoking (advice; 38/44, 86%; 11/31, 35%; and 29/29, 100%; respectively), but very few fulfilled the "assess" component (23/44, 52%; 1/31, 3%; and 18/29, 62%; respectively), that is, readiness to change and interest in quitting.

**Table 2.** Content and functions of the apps by type (N=104).

| Function | Combined type (n=44), n (%) | Multifunctional type (n=31), n (%) | Informational type (n=29), n (%) |
|---|---|---|---|
| **Ask (app assessed smoking status)** | | | |
| Overall | 44 (100) | 30 (97) | 0 (0) |
| **Current smokers** | | | |
| Number of cigarettes smoked per day | 41 (93) | 29 (94) | 0 (0) |
| Time until first cigarette of the day | 9 (20) | 3 (10) | 0 (0) |
| Smoke when sick | 6 (14) | 2 (6) | 0 (0) |
| Reasons to smoke/quit smoking | 22 (50) | 4 (13) | 0 (0) |
| **Smoking triggers** | | | |
| Time of day-related smoking triggers | 21 (48) | 4 (13) | 0 (0) |
| Other smoking triggers | 13 (30) | 4 (13) | 0 (0) |
| **Advise (app advised the user to quit smoking)** | | | |
| **Overall** | | | |
| General advice | 38 (86) | 11 (35) | 29 (100) |
| Personalized advice (using user-provided info) | 8 (18) | 2 (6) | 0 (0) |
| **Assess (app assessed the user's readiness to quit)** | | | |
| **Overall** | 23 (52) | 1 (3) | 18 (62) |
| User could indicate lack of readiness to quit | 14 (32) | 1 (3) | 0 (0) |
| Barriers to quitting were addressed[a] | 23 (52) | 1 (3) | 18 (62) |
| **Assist (app assisted the user with the quit attempt)** | | | |
| Overall | 44 (100) | 31 (100) | 29 (100) |
| **Setting a quit date** | | | |
| Users were asked to pick a quit date | 41 (93) | 24 (77) | 0 (0) |
| Users received support for their quit attempt | 44 (100) | 30 (97) | 7 (24) |
| Users received feedback on their quit attempt[b] | 4 (9) | 1 (3) | 0 (0) |
| **Reward** | | | |
| Reminders about money saved since quitting | 40 (91) | 23 (74) | 0 (0) |
| Reminders about health benefits accrued | 19 (43) | 22 (71) | 10 (34) |
| Information about social benefits[b] | 13 (30) | 4 (13) | 5 (17) |
| **Risk** | | | |
| Information about health risks of smoking[b] | 27 (61) | 7 (23) | 9 (31) |
| **Support provided** | | | |
| Distraction from urges, reminders about number of cigarettes not smoked since quitting | 43 (98) | 29 (94) | 10 (34) |
| Users could interact with other users for mutual support (app community)[a] | 12 (27) | 12 (39) | 0 (0) |
| Web community[b] | 2 (5) | 2 (6) | 0 (0) |
| Referral to Quitline or other support groups | 18 (41) | 3 (10) | 3 (10) |
| Recorded personalized message to be played back later | 24 (55) | 18 (58) | 0 (0) |
| Reminders of their motivations during difficult times | 19 (43) | 13 (42) | 0 (0) |
| Motivation alarm[b] | 17 (39) | 4 (13) | 0 (0) |

| Function | Combined type (n=44), n (%) | Multifunctional type (n=31), n (%) | Informational type (n=29), n (%) |
|---|---|---|---|
| Personalized motivation alarm[b] | 4 (9) | 3 (10) | 0 (0) |
| Encouragement (to improve self-confidence, helpful quotes)[b] | 20 (45) | 8 (26) | 10 (34) |
| **Information provided** | | | |
| Information on self-confidence[b] | 4 (9) | 0 (0) | 13 (45) |
| Information on counseling, treatment, meds, Quitline, etc | 42 (95) | 1 (3) | 29 (100) |
| Links to resources | 19 (43) | 1 (3) | 2 (7) |
| **Arrange follow-ups (app followed up with the user regarding the quit attempt)** | | | |
| **Overall** | 44 (100) | 30 (97) | 10 (34) |
| Checked-in prior to quit attempt | 30 (68) | 17 (55) | 0 (0) |
| Checked-in after quit attempt | 44 (100) | 20 (65) | 2 (7) |
| If relapsed, encouraged user to set a new quit date | 34 (77) | 27 (87) | 1 (3) |
| If relapsed, possible to add smoking number[a] | 19 (43) | 7 (23) | 0 (0) |
| If relapsed, offered encouragement that quitting takes practice | 7 (16) | 1 (3) | 0 (0) |
| Information provided about relapse[b] | 9 (20) | 1 (3) | 8 (28) |

[a]Modification of the original tool.

[b]New additions to the original tool.

The "ask" component could not be addressed by the informational type, which provides information only and has no specific function. In terms of the questions for the current smokers, most apps in the other two types asked about the number of cigarettes smoked per day (41/44, 93% of combined apps; 29/31, 94% of multifunctional apps), but very few apps in those types asked about smoking when sick (6/44, 14% and 2/31, 6%, respectively). Few multifunctional apps asked about the smoking triggers (4/31, 13%). The advice provided by the apps was largely generic and very rarely personalized (8/44, 18% of combined apps; 2/31, 6% of multifunctional apps; and none of the informational apps). Barriers to readiness to quit was assessed or explained in nearly a half of the combined apps (23/44, 52%), 18 (62%) of the 29 informational apps, and 1 (3%) of the 31 multifunctional apps. However, there were almost no apps in which users could indicate their barriers on their own, except for the combined apps (combined apps 14/44, 32%; multifunctional apps 1/31, 3%; and informational apps 0).

Among the combined apps, the "assist" content was typically in the form of basic support regarding quit attempts (44/44, 100%; at least one function facilitating smoking cessation), reminders about the number of cigarettes not smoked or days of not smoking (43/44, 98%), asking the user to pick a quit date (41/44, 93%), and basic information on smoking cessation (42/44, 95%; eg, electronic books providing facts about smoking). Assist-related functions that could exploit smartphone technology to provide personalized content such as a personalized motivation alarm (4/44, 9%), tailored feedback on quit attempts (4/44, 9%), or information promoting self-confidence (4/44, 9%) were rarely utilized. Among the multifunctional apps, the "assist" content commonly consisted of basic support regarding quit attempts (30/31, 97%) and reminders about the number of cigarettes not smoked or days of not smoking (29/31, 94%). Similar to the combined apps, personalized feedback on quit attempts (1/31, 3%) and a personalized motivation alarm (3/31, 10%) were rarely utilized. Additionally, few apps provided information on basic smoking cessation information (1/31, 3%) and links to resources (1/31, 3%).

Among informational apps, basic information was provided by 29 (100%) of the apps, and 10 (34%) included general information about self-confidence, the health benefits of cessation, and the number of cigarettes smoked or days of not smoking. Less than 5 (17%) of the 29 informational apps provided information about social benefits, support regarding the user's quit attempt or Quitline or other support groups, or links to resources.

The "arranging follow-up" component, which only apps can offer, typically consist of a "check-in" prior to a quit attempt (30/44, 68% of combined apps; 17/31, 55% of multifunctional apps), check-in after a quit attempt (44/44, 100% and 20/31, 65%, respectively), and if relapsed, encouragement to set a new quit date (34/44, 77% and 27/31, 87%, respectively). Few apps provide support regarding relapse in the form of a reminder that quitting takes practice (7/44, 16% and 1/31, 3%, respectively). Informational apps hardly arranged follow-ups.

## Quality (MARS Score) of the Apps by Type

Table 3 shows the mean MARS scores by app type. The mean MARS scores of the combined, multifunctional, and informational apps were comparable at 3.64, 3.26 and 3.0, respectively. The mean scores on the 4 MARS dimensions were calculated. The functionality dimension scores were the highest among the subscores (3.97, 3.83, and 3.86 for combined, multifunctional, and informational type, respectively) and scores

on the engagement dimension were the lowest (3.52, 3.1, and 2.23 for combined, multifunctional, and informational type, respectively). Information quality score was the lowest in the multifunctional type (3.04). For the combined and multifunctional type, the mean scores on all dimensions were 3 or higher, whereas they were less than 3 for the informational type, except in functionality. The difference in all the dimension scores by type was statistically significant except the functionality score.

**Table 3.** Quality of the apps by type (N=104).

| Mobile App Rating Scale component scores | Combined type (n=44), mean (SD) | Multifunctional type (n=31), mean (SD) | Informational type (n=29), mean (SD) | *F* test *(df)* | *P* value |
|---|---|---|---|---|---|
| **Total score** | 3.64 (0.38) | 3.26 (0.48) | 3.0 (0.32) | 29.21 (2) | <.001 |
| Engagement | 3.52 (0.57) | 3.1 (0.57) | 2.23 (0.39) | 68.65 (2) | <.001 |
| Functionality | 3.97 (0.53) | 3.83 (0.58) | 3.86 (0.42) | 1.04 (2) | .36 |
| Aesthetics | 3.60 (0.69) | 3.08 (0.78) | 2.97 (0.52) | 10.37 (2) | <.001 |
| Information quality | 3.53 (0.56) | 3.04 (0.47) | 2.93 (0.38) | 15.52 (2) | <.001 |

## App Features Affecting Quality (MARS Score) or Content and Function (Content Score)

Table 4 shows the results of a multivariate analysis of app features, MARS scores, and content scores. Feedback channel with developer and developer sector had a significant impact on both scores. Both MARS and content scores were higher when feedback with a developer was possible within the app compared to that when feedback was only available at the app market level. The MARS score was higher for apps developed as a collaboration between government and commercial institutions compared to that when the developer was unknown. The content score was higher for apps developed by government or a university or commercial institution compared to that when the developer was unknown. Platform type was found to have a significant impact on MARS score, with iPhone apps having higher MARS scores than Android apps.

**Table 4.** Multiple regression analysis of quality (Mobile App Rating Scale score) or content and function (content score) (N=104).

| Category, items | Mobile App Rating Scale score | | | Content score | | |
|---|---|---|---|---|---|---|
| | β | SE | *P* value | β | SE | *P* value |
| Intercept | 2.91 | 0.07 | <.001 | 9.22 | 0.95 | <.001 |
| **Developer sector** | | | | | | |
| Government | .17 | 0.21 | .42 | 6.98 | 2.67 | .01 |
| University | .28 | 0.20 | .17 | 7.14 | 2.60 | .01 |
| Government with commercial | .52 | 0.23 | .03 | 4.87 | 2.99 | .11 |
| Nongovernmental organization | .31 | 0.28 | .27 | 2.69 | 3.61 | .46 |
| Commercial | .12 | 0.09 | .18 | 3.50 | 1.17 | <.001 |
| Unknown | Ref[a] | Ref | Ref | Ref | Ref | Ref |
| Affiliation of developer with health care professionals | .11 | 0.13 | .38 | 2.82 | 1.66 | .09 |
| Affiliation of developer with non–health care professionals | Ref | Ref | Ref | Ref | Ref | Ref |
| **Platform** | | | | | | |
| Both | .17 | 0.11 | .13 | 1.40 | 1.41 | .33 |
| iPhone | .26 | 0.11 | .02 | –.20 | 1.40 | .88 |
| Android | Ref | Ref | Ref | Ref | Ref | Ref |
| **Payment type** | | | | | | |
| Prepaid | –.12 | 0.13 | .36 | –.26 | 1.62 | .87 |
| In-app purchase | .21 | 0.10 | .04 | 1.40 | 1.34 | .30 |
| Free | Ref | Ref | Ref | Ref | Ref | Ref |
| **Feedback channel with developer** | | | | | | |
| Within the app | .35 | 0.10 | <.001 | 3.96 | 1.24 | <.001 |
| Contact information of developer provided | .17 | 0.13 | .19 | 2.01 | 1.67 | .23 |
| Market level only | Ref | Ref | Ref | Ref | Ref | Ref |

[a]Ref: reference value.

# Discussion

## Main Results

This study classified existing smoking cessation apps into 3 types and then evaluated their content (or functions) and quality. The characteristics associated with content and MARS scores were also analyzed. Combined type apps had the highest content and MARS scores among the 3 app types. Multifunctional type apps had higher MARS scores than informational type apps. Content and function analysis showed that multifunctional apps better represented the function-related components of “ask, assist, and arrange follow-ups” than did informational apps. On the contrary, informational apps better represented the information-related components of “advise and assess” than did multifunctional apps. Some previous studies that analyzed the content of smoking cessation apps by using the 5As guideline reported that very few apps actually conformed to the guideline [12,17,18]. In one such study, the percentage of apps consistent with the (modified) 5As guideline did not exceed 50% except for the “assist” dimension [20]. In a study analyzing smoking cessation apps using MARS, a mean MARS score of 2.88 was reported [24]. The apps analyzed in our study better adhered to the 5As guideline and had higher MARS scores than those reported in previous studies [12,17,18,20,23], possibly because the functions of the apps may have been improved in the interim. Other studies used clinical guidelines to analyze app content [12,17,18]; therefore, guideline adherence may have been lower than that when using guidelines specifically designed for apps [20]. It has been shown that decision-aid apps with multiple functions such as motivational messages, a quit diary, and a quitting benefits tracker but including scant information on quitting strategies are more effective for smoking cessation than information-only apps [14]. Another study found that evidence-based apps with customized functions and information were more effective than a self-help booklet [15]. Our results support these studies where the combined apps, which are similar to decision-aid apps and evidence-based apps, had higher scores than informational apps.

Mobile-based interventions have advantages over standard interventions for smoking cessation [25]. One such advantage of smartphone apps is the potential to deliver a user-centered interactive intervention. Combined apps are considered able to better exploit this advantage, and the content scores and MARS scores of this type of apps support that view. Multifunctional apps are somewhat in line with this. In this study, the combined type had the best general characteristics followed by the

multifunctional and informational type. Combined apps were more likely to be developed by trusted institutions (government or universities) or health professionals and to not include advertisements and be available for free, followed by multifunctional and informational apps.

Regression analysis showed that the sector in which a developer is situated and the feedback channel with developers are important. Our findings suggest that governments or universities should ideally be the creators of smoking cessation apps. Most of the apps produced by governments and universities were of the combined type, which had high content and MARS scores. In a previous study [24], 3 of 6 apps that received high scores were developed by the government; the other 3 were produced by research institutions. In detail, the regression analysis showed that government- or university- or commercial institution–developed app content scores were high. The MARS scores were high for apps developed collaboratively by government and commercial institutions. Although not described in the results, 4 apps created via such collaborations had a mean MARS score of 3.96, which is above the average score of combined type apps (3.64). Apps developed collaboratively by government and commercial institutions benefit from the reliable information provided by the former and the technical expertise of the latter.

In this study, apps providing an option for feedback with a developer had higher content and MARS scores. Further, users themselves provide more feedback when they are able to do so directly within the app, where such feedback can help improve app content and MARS scores. Although the content and MARS scores do not necessarily indicate smoking cessation efficacy, the interaction between users and developers is beneficial when creating a smoking cessation app or other mHealth apps; therefore, a feedback function should be included.

### Secondary Features Offered to Users

Informational apps had generally low content scores because they provide only information and have no functions. The low MARS scores of informational apps may have been influenced by the high proportion of unknown developers for these apps (24/29, 83%); informational apps can be created easily by almost anyone. It is easy to obtain general information pertaining to smoking cessation from the internet, and nameless developers have no responsibility to maintain or improve their apps. Although there were more than 49 apps of informational type in the finally selected apps, more than 20 apps were excluded from the final analysis. App instability [24] can be an issue—also due to an absence of management by unknown developers. The biggest problem with respect to informational apps was that they were mostly copies of other apps. Ten informational type apps included in the final analysis were copied apps that had different developers.

The large market share of counter and hypnosis apps is also problematic. Counter apps estimate the money saved by smoking cessation or the number of cigarettes not smoked. Hypnosis apps promote smoking cessation through video- or audio-based hypnosis. Of the 394 apps downloaded after the screening phase, 157 were counter apps (40%). Since hypnosis apps were excluded at the screening phase, the exact number of such apps was not counted, but they also appeared to have a large market share. In fact, when searching for apps using the term “quit smoking,” 11 apps (6 counter and 5 hypnosis type) in Google Play and 6 apps (5 counter and 1 hypnosis type) in the App Store were noted among the top 20 apps in each market. Single-function apps might be useful for users whose digital literacy might be limited or who are prescribed the apps by a health care provider for a specific purpose. However, their function is quite insufficient for the general public. A 2011 study reported low content scores for single-function apps (counter and hypnosis apps) [17]. The result of a Cochrane review showed that the effectiveness of hypnosis for smoking cessation is not clear [26]. Further, it has been found that the transference relationship between the therapist and patient influences the smoking cessation efficacy of hypnosis, but it is difficult to build such a relationship through an app [26]. In a recent study, the effect of a meditation app for quitting smoking was not revealed [27]. If problematic apps (made by an unknown developer or copied) and single-function apps have a large market share, users may find it difficult to access better apps. In this study, we were able to identify the problematic apps as above through the process of the app exclusion.

### Limitations

First, owing to app volatility, many apps were suspended during the analysis. High-quality apps should be maintained through collaboration between government and commercial institutions. Second, as the purpose of this study was to compare apps by function, single-function apps, which have only 1 function (eg, counter or hypnosis), were excluded. Since those are not applicable in many items of evaluation tools, evaluation has no meaning. In future research, it will be possible to analyze and evaluate only these apps. Lastly, although we identified apps with high MARS scores and with a large amount of content as recommended by the Clinical Practice Guideline for Treating Tobacco Use and Dependence, we cannot definitively conclude that these apps would be effective with respect to smoking cessation. Further studies should be conducted to confirm the smoking cessation efficacy of different classes of apps.

## Conclusions

This study can be a guide for users and developers of smoking cessation apps. This is the first study to evaluate the content and quality of smoking cessation apps by classification. The combined type had higher-quality content and functionality than the multifunctional type and the informational type. However, the other two types also had their own advantages. The classification of the apps can help users choose the appropriate type of smoking cessation app for their needs. The identification of characteristics that affected the app scores in this study may help in the development of a smoking cessation app (development collaboration between health professionals, academic researchers and industry, inclusion of a communication option with a developer within the app, etc). Additionally, we identified that problematic apps such as those made by unknown developers or copied and single-function apps had a large market share. This could promote discussion on the possible regulation of problematic apps. Public health–related apps should be more rigorously examined before

being released to the market. In summary, our results will contribute to the use and development of better smoking cessation apps.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
App list.
[PDF File (Adobe PDF File), 379 KB - mhealth_v10i2e17268_app1.pdf ]

## References


1. Gender, women, and the tobacco epidemic. World Health Organization (WHO). 2010. URL: https://www.who.int/publications-detail-redirect/9789240004849 [accessed 2021-08-13]
2. DiFranza JR, Savageau JA, Fletcher K, O'Loughlin J, Pbert L, Ockene JK, et al. Symptoms of tobacco dependence after brief intermittent use: the Development and Assessment of Nicotine Dependence in Youth-2 study. Arch Pediatr Adolesc Med 2007 Jul;161(7):704-710. [doi: 10.1001/archpedi.161.7.704] [Medline: 17606835]
3. Bala MM, Strzeszynski L, Topor-Madry R. Mass media interventions for smoking cessation in adults. Cochrane Database Syst Rev 2017 Nov 21;11:CD004704 [FREE Full text] [doi: 10.1002/14651858.CD004704.pub4] [Medline: 29159862]
4. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
5. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
6. Global and regional ICT data: Mobile-cellular subscriptions. International Telecommunication Union (ITU). 2021. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2021-08-13]
7. Number of apps available in leading app stores as of 1st quarter 2021. Statistica. 2021. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2021-08-13]
8. Combined global Apple App Store and Google Play app downloads from 1st quarter 2015 to 4th quarter 2021 (in billions). Statista. 2021. URL: https://www.statista.com/statistics/604343/number-of-apple-app-store-and-google-play-app-downloads-worldwide/ [accessed 2021-08-13]
9. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
10. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
12. Bennett ME, Toffey K, Dickerson F, Himelhoch S, Katsafanas E, Savage CL. A review of android apps for smoking cessation. J. Smok Cessat 2014 Feb 04;10(2):106-115. [doi: 10.1017/jsc.2014.1]
13. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
14. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
15. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]
16. App stores - Statistics & Facts. Statista. 2019. URL: https://www.statista.com/topics/1729/app-stores/ [accessed 2020-05-03]
17. 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]
18. Abroms LC, Lee Westmaas J, 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]

19. Cheng F, Xu J, Su C, Fu X, Bricker J. Content analysis of smartphone apps for smoking cessation in China: empirical study. JMIR Mhealth Uhealth 2017 Jul 11;5(7):e93 [FREE Full text] [doi: 10.2196/mhealth.7462] [Medline: 28698170]
20. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
21. Clinical Practice Guideline Treating Tobacco UseDependence 2008 Update Panel, Liaisons,Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
22. A Guide for Tobacco Users to Quit. Geneva: World Health Organization; 2014.
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 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
24. 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]
25. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
26. Barnes J, Dong CY, McRobbie H, Walker N, Mehta M, Stead LF. Hypnotherapy for smoking cessation. Cochrane Database Syst Rev 2010 Oct 06(10):CD001008. [doi: 10.1002/14651858.CD001008.pub2] [Medline: 20927723]
27. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to quit: A randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]


## Abbreviations

**5As:** ask, advise, assess, assist, and arrange follow-ups
**MARS:** Mobile App Rating Scale
**mHealth:** mobile health
**RR:** relative risk


*Edited by L Buis; submitted 01.12.19; peer-reviewed by C Lucchiari, L Martinengo, Y Liao; comments to author 10.03.20; revised version received 05.05.20; accepted 02.12.21; published 17.02.22.*

*Please cite as:*
*Seo S, Cho SI, Yoon W, Lee CM*
*Classification of Smoking Cessation Apps: Quality Review and Content Analysis*
*JMIR Mhealth Uhealth 2022;10(2):e17268*
*URL: https://mhealth.jmir.org/2022/2/e17268*
*doi:10.2196/17268*
*PMID:35175213*

Original Paper

# Exploring and Characterizing Patient Multibehavior Engagement Trails and Patient Behavior Preference Patterns in Pathway-Based mHealth Hypertension Self-Management: Analysis of Use Data

Dan Wu[1*], BSc; Xiaoyuan Huyan[2*], MBBS; Yutong She[1], BSc; Junbin Hu[3], MBBS; Huilong Duan[1], PhD; Ning Deng[1,4,5], PhD

[1]College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China

[2]The First Health Care Department, The Second Medical Center & National Clinical Research Center for Geriatric Diseases, Chinese People's Liberation Army General Hospital, Beijing, China

[3]Health Community Group of Yuhuan People's Hospital, Kanmen Branch, Taizhou, China

[4]Alibaba–Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou, China

[5]Binjiang Institute of Zhejiang University, Hangzhou, China

[*]these authors contributed equally

**Corresponding Author:**
Ning Deng, PhD
College of Biomedical Engineering and Instrument Science
Ministry of Education Key Laboratory of Biomedical Engineering
Zhejiang University
38 Zheda Rd, Zhouyiqing Bldg 512
Yuquan Campus
Hangzhou, 310000
China
Phone: 86 571 2295 2693
Email: zju.dengning@gmail.com

## Abstract

**Background:** Hypertension is a long-term medical condition. Mobile health (mHealth) services can help out-of-hospital patients to self-manage. However, not all management is effective, possibly because the behavior mechanism and behavior preferences of patients with various characteristics in hypertension management were unclear.

**Objective:** The purpose of this study was to (1) explore patient multibehavior engagement trails in the pathway-based hypertension self-management, (2) discover patient behavior preference patterns, and (3) identify the characteristics of patients with different behavior preferences.

**Methods:** This study included 863 hypertensive patients who generated 295,855 use records in the mHealth app from December 28, 2016, to July 2, 2020. Markov chain was used to infer the patient multibehavior engagement trails, which contained the type, quantity, time spent, sequence, and transition probability value (TP value) of patient behavior. K-means algorithm was used to group patients by the normalized behavior preference features: the number of behavioral states that a patient performed in each trail. The pages in the app represented the behavior states. Chi-square tests, Z-test, analyses of variance, and Bonferroni multiple comparisons were conducted to characterize the patient behavior preference patterns.

**Results:** Markov chain analysis revealed 3 types of behavior transition (1-way transition, cycle transition, and self-transition) and 4 trails of patient multibehavior engagement. In perform task trail (PT-T), patients preferred to start self-management from the states of task blood pressure (BP), task drug, and task weight (TP value 0.29, 0.18, and 0.20, respectively), and spent more time on the task food state (35.87 s). Some patients entered the states of task BP and task drug (TP value 0.20, 0.25) from the reminder item state. In the result-oriented trail (RO-T), patients spent more energy on the ranking state (19.66 s) compared to the health report state (13.25 s). In the knowledge learning trail (KL-T), there was a high probability of cycle transition (TP value 0.47, 0.31) between the states of knowledge list and knowledge content. In the support acquisition trail (SA-T), there was a high probability of self-transition in the questionnaire (TP value 0.29) state. Cluster analysis discovered 3 patient behavior preference

patterns: PT-T cluster, PT-T and KL-T cluster, and PT-T and SA-T cluster. There were statistically significant associations between the behavior preference pattern and gender, education level, and BP.

**Conclusions:** This study identified the dynamic, longitudinal, and multidimensional characteristics of patient behavior. Patients preferred to focus on BP, medications, and weight conditions and paid attention to BP and medications using reminders. The diet management and questionnaires were complicated and difficult to implement and record. Competitive methods such as ranking were more likely to attract patients to pay attention to their own self-management states. Female patients with lower education level and poorly controlled BP were more likely to be highly involved in hypertension health education.

***(JMIR Mhealth Uhealth 2022;10(2):e33189)*** doi:10.2196/33189

**KEYWORDS**

hypertension; mobile health; patient behavior; engagement; data analysis

# Introduction

## Background

Hypertension is a serious medical condition that affects health-related quality of life and increases the risks to the heart, brain, kidney, etc [1]. Controlling hypertension requires patients to achieve management goals by insistently adhering to long-term self-management plans, including regularly checking blood pressure (BP), taking medication, being physically active on a regular basis, eating more fruit and vegetables, and reducing alcohol consumption, etc. These plans should be established based on hypertension management guidelines and guidance from health care providers [2].

The development of mobile technology has promoted the implementation of out-of-hospital mobile health (mHealth) services [3-6]. Extensive evidence supports the effect of mHealth services in disease control [4,5,7-11], including promoting patient engagement in health care services and helping patients develop positive behavior in their daily self-management [12-16]. In comparison with transitional hypertension management methods, mHealth services can effectively improve patient engagement in hypertension self-management [17,18]. Cechetti et al [19] designed an mHealth app with a gamification method for hypertension management, which can effectively promote patient engagement in their self-management. However, while recent studies have demonstrated the effectiveness of some mHealth services, others have performed poorly [12-15]. There have been mixed results for using mHealth services to support patient self-management of hypertension in the community [20,21]. Thus, hypertension mHealth services pose new design challenges in management strategies, partly because the mechanism by which patients engage in their self-management is not clear.

Patient engagement is a broad concept that combines patient activation with interventions of health care services designed to increase activation and promote positive patient behavior [22]. Positive self-management behavior of patients engaging in mHealth services is essential for bringing an improvement in health outcomes [12,17,18,23-26]. Goyal et al [27] found a significant relationship between an increased number of daily blood glucose readings and improved glycated hemoglobin. Toto-Ramos et al [28] found that hypertensive patients with sustained engagement in mHealth services experienced significant reduction in BP. Thus, understanding self-management behavior of real-world patients in their daily lives can help to reveal patient behavior in natural settings. Compared to attracting patients from clinical trials who are more likely to overcome the burden associated with research work [29], this provides deeper insight into real patient self-management behaviors.

Hypertension management requires long-term efforts, and patient behavior in hypertension management is dynamic and continuous [30,31]. Understanding the longitudinal characteristics of patients engaged in self-management is important for long-term successful management. Moreover, there are many different dimensions of behavior with patients who engage in mHealth services [12,30,32,33], such as measuring, viewing, and recording, which are associated with individual inherent preferences and habits of patients [34-36]. Patient engagement behavior has been measured as amount, duration, breadth, and depth of using mHealth services [12]. Rahman et al [37] measured patient engagement by 3 key use features: duration and frequency of using the mHealth app plus the number of use records. Sanatkar et al [33] measured 5 use features of patient engagement: number of user log-ins, number of daily trackers used, numbers of learning activities started and completed, and number of reminders received. However, the multidimensional and dynamic behavioral processes that change over time cannot be captured simply by analyzing count data captured at 1 time point in the cross-sectional data analysis. Longitudinal change in patient multibehavior can be identified through analyzing the time series data. More comprehensive understanding of the multiple self-management behavior and individual behavior preferences of patients engaging in mHealth self-management over a long period requires further research. Data mining techniques have been successfully adopted in the analysis of longitudinal events, such as human behavior navigation [26,38-40], information search behavior [41], phase-type distribution in the health care industry [42], and lifetime health care costs [43]. These would allow us to infer the characteristics of patient behavior throughout the entire period of self-management.

## Objective

The aim of this study was to explore the trails of patient multibehavior engagement in the pathway-based mHealth hypertension self-management and identify the characteristics of patients with different behavior preferences. This included 3 objectives. The first objective was to discover the trails of patient multibehavior engagement. The second objective was

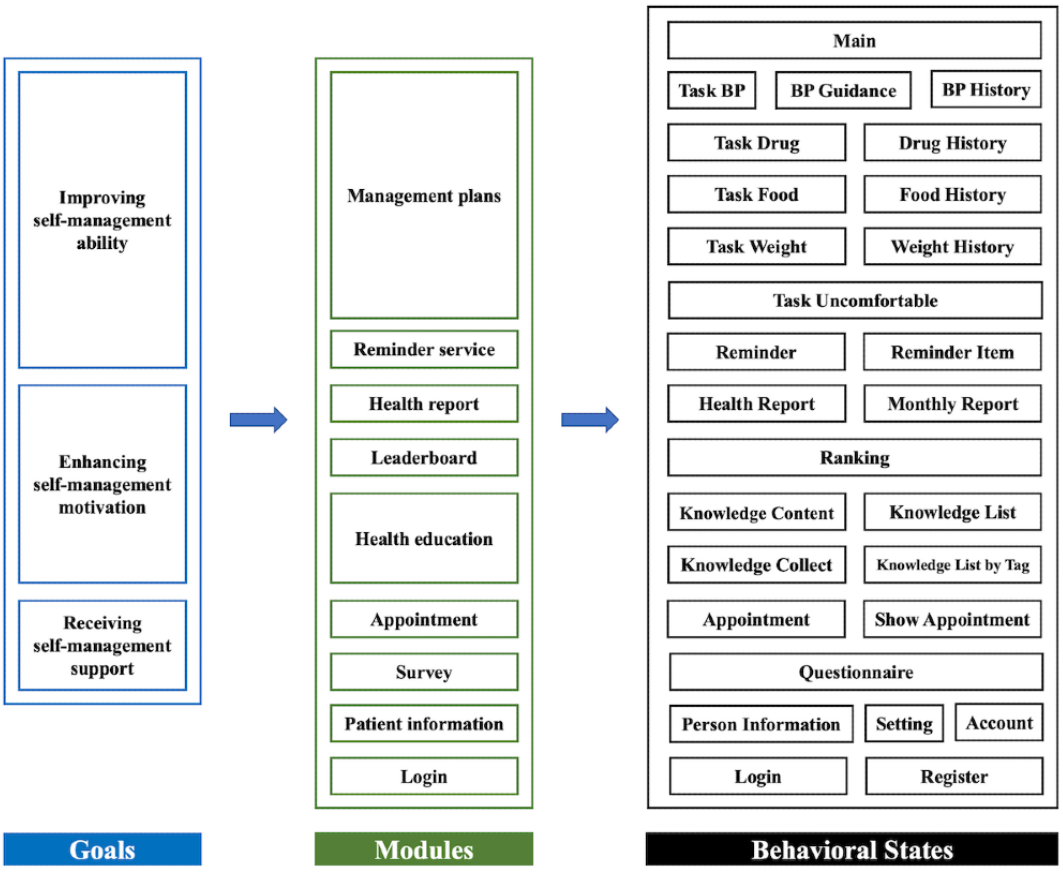
to explore patient behavior preference patterns. The third objective was to identify the association between behavior preference patterns and the demographic and physiological characteristics. Identifying the multidimension and longitudinal characteristics of patient behavior within the mHealth hypertension management offers new opportunities for personalizing management goals and plans to reduce nonadherence and enhance possible effectiveness.

## Methods

### Description of the mHealth Hypertension Management App

We used data from the Blood Pressure Assistant, a pathway-based hypertension self-management app in the Digital Care Study for Hypertension Management [44]. The app was designed in accordance with a customized care pathway in compliance with the Chinese guideline for hypertension management [45] and was available for patients in the General Hospital of Ningxia Medical University. The care pathway involves 2 roles in hypertension management: health care providers and patients. The care pathway defines 3 goals for patients—improve self-management ability, enhance self-management motivation, and receive self-management support (see Figure 1)—and comprises 9 modules and 28 behavioral states (see Multimedia Appendix 1 for the detailed behavioral states). The 9 modules generate intervention plans and patient self-management plans. The term of state comes from the Markov decision process, a mathematical model of sequential decision. The behavioral state is a description of the patient's behavior in a hypertension self-management environment or scene and is expressed by pages in the mHealth app. In this paper, we chose to use state to represent the patient behavior.

**Figure 1.** Components of pathway-based mHealth hypertension self-management. BP: blood pressure.




In this care pathway, patient actions on different pages of the app reflect specific patient behaviors. The behavioral states are represented by different pages in the app. Through the app, each patient registers and enters basic demographic information, after which the patient is assigned to a health care provider who enrolls patients they manage into the mHealth hypertension management program at an online community [45]. The health care provider is responsible for formulating a tailored management plan, reviewing patients' uploaded data, and conducting follow-up. Patients can use the app to seek help from health care providers, check their management plans, and record self-management data including BP monitoring, medication, physical activity, and diet. These patient behaviors are represented on the appointment, main, task BP, task drug, and other pages of this mHealth app. Health care providers can track current BP readings of patients through a web-based platform to adjust management plans and use mobile phones for patient follow-ups to assist in BP control.

### Data Collection

#### Ethics

Ethical approval was granted by the ethics committee for conducting human research at the General Hospital of Ningxia Medical University (NXMU-GH-2017-273). All patients in this

study signed the informed consent forms for their anonymized data to be used in routine evaluations to monitor and improve health care services.

### *Sample*

Since this app was launched in December 2015, 1159 patients have used it to self-manage their BP. We selected patients based on the following inclusion criteria: aged 18 to 80 years, diagnosed with hypertension, performed self-management between December 28, 2016, and July 2, 2020 (the main functions of the app were consistent during this period, ensuring patient behavior was not affected by the changes in app functions).

### *Data Extraction*

All data were stored and extracted from the server of the Blood Pressure Assistant, which contains the self-management plans, demographic information, uploaded self-management data from patients, and follow-up records of health care providers. We extracted 3 types of data from the database: demographics, physiological records, and patient use records. Demographics included patient identification, data of birth, gender (male or female), and education level (below high school, high school, university and above). Physiological records included patient identification, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), and uploaded date (year, month, day, minute, and second). Patient use records included patient identification, page name, time page is entered and exited, and stay time. These data help determine longitudinal and multidimensional behavioral characteristics of hypertensive patients and characteristics of patients with different behavior preferences.

## Data Analysis

### *Identifying Patient Multibehavior Engagement Trails*

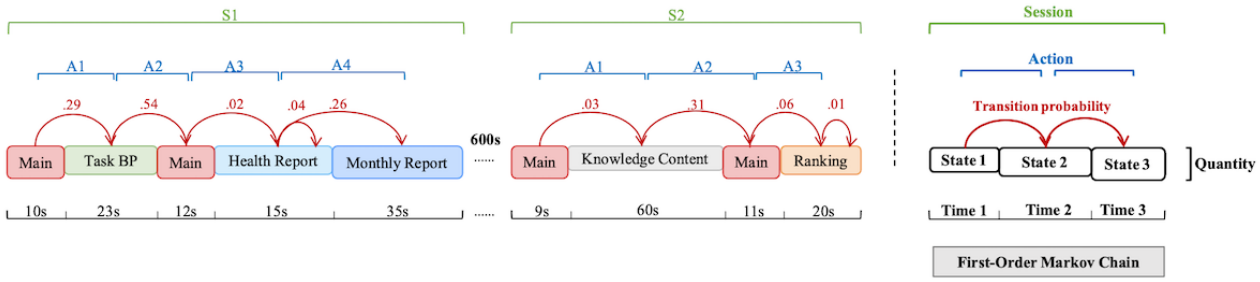
Patient multibehavior engagement trails were indicated by the type, quantity, time spent, sequence, and transition probability value (TP value) of patient behavioral states (see Figure 2). The time spent and sequence were calculated from the time pages were entered and exited.

The data of patient use behavior were analyzed through first-order Markov chain analysis, a method to model stochastic processes, which is suitable for analyzing user interaction with mHealth apps [46,47]. In the mHealth field, interaction between users and apps is determined through a process of connecting multiple continuous actions together. In the case of this study, we assumed that a next state depends only on the current state and not on the history of previous states. Thus, we used the first-order Markov chain called a memoryless model, containing the state space S , action, session, and transition matrix P. The discrete state space S was defined by the $n$ different behavioral states (represented by different pages in the app): $S=\{s_1,...,s_n\}$. A start state $s_0$ and an exit state $s_{n+1}$ were created when the patient started and left the app. The action was defined by the transition from one behavioral state to another ($s_i \rightarrow s_j$). The session was discriminated by the time interval between 2 consecutive states greater than 600 seconds ($s_i - s_j > 600$ $s$). The transition matrix P contained each element on row $i$ and column $j$ ($p_{i,j}$) and indicated the transition probability that a patient moves from state $s_i$ to state $s_j$. The transition probability $p_{i,j}$ was defined as

where $N_{i,j}$ was the number of transitions $s_i \rightarrow s_j$. The quantity and time spent of the behavioral state are represented by the width and length of the rectangle (see Figure 2).

The transition matrix P was displayed by a heat map, which allowed us to understand the characteristics of patient multibehavior transition. Analysis was carried out using Python (version 3.8, Python Software Foundation). The 1% relative quantity of states and the 0.15 transition probability were used as cutoff values to improve readability. We then combined quantity, time, transition probability, and the pathway to plot the main trails of patient multibehavior engagement.

**Figure 2.** Example of patient multibehavior engagement trail represented by first-order Markov chain. BP: blood pressure.




### *Cluster Analysis of Patients With Different Behavior Preferences*

Cluster analysis was conducted to group patients according to patient behavior preference features: the number of behavioral states that a patient performed in each trail. First, we normalized the behavior preference features of each patient. The sum of behavior preference features of each patient for different behavior trails was 1. A K-means algorithm was then used to cluster patients through these normalized patient behavior preference features. Euclidean distance was used to calculate the similarity of features between patients. Finally, we used a silhouette score to determine the optimal number of clusters. Higher silhouette scores indicate tighter clusters, where each cluster is completely separate from the others.

### *Characterizing the Clusters*

An optimal clustering result was reached based on the silhouette score of different numbers of clusters. For each cluster, we analyzed the demographic features (age, gender, and education level), and physiological features (mean BP [SBP and DBP] and mean HR). Statistical analysis was conducted in SPSS (version 24, IBM Corp). A chi-square test was performed to evaluate the statistical significance of associations between the clusters and discrete variables (gender and education level). A z-test was used to conduct pairwise comparisons of the differences among the proportion of discrete variables between the clusters. Analyses of variance (ANOVAs) were used to evaluate the statistical significance of associations between the clusters and continuous variable (age, BP, and HR). Bonferroni multiple comparisons were further conducted to examine the differences between the clusters. The cutting value ($P \leq .05$) was used to determine whether the difference was statistically significant.

## Results

### Patient Multibehavior Engagement Trails

### *Quantity and Time Characteristics of Each Behavioral State*

We restricted analysis to the remaining 863 patients, with 295,855 records from the app. Tables 1 and 2 depict the quantity and time spent in each behavioral state. It was clear that the main state (125,487 [42.42%]; 1,745,286 s [13.91 s]) was visited most in quantity and total time spent; it was the default starting page for a session. The relative number of 14 states exceeded 1%, and the mean time spent of 12 states exceeded 20 seconds. In 5 self-management tasks, the state of task BP (48,935 [16.54%]; 1,253,924 s [25.62 s]) was visited more, and the total time spent was also longer. The mean time spent on task food (149,164 s [35.87 s]) state was more than other self-management tasks. The states of ranking (7856 [2.66%]) and knowledge content (7330 [2.48%]) had relatively high quantities of visits. The states of knowledge content (615,430 s [83.96 s]) and show appointment (261,371 s [162.75 s]) had relatively long mean time spent. In 9 modules, the mean time spent in management plans (contained 5 self-management task parts: 230.21 s) was longer than health education (contained 4 knowledge parts: 124.06 s) and appointment (contained 2 appointment parts: 168.72 s).

**Table 1.** Quantity of each behavioral state.

| Behavioral state[a] | Quantity, n (%) |
|---|---|
| Main | 125,487 (42.42) |
| Task blood pressure | 48,935 (16.54) |
| Task weight | 33,560 (11.34) |
| Task drug | 26,509 (8.96) |
| Ranking | 7856 (2.66) |
| Knowledge content | 7330 (2.48) |
| Blood pressure history | 6070 (2.05) |
| Knowledge list | 4871 (1.65) |
| Task food | 4159 (1.41) |
| Questionnaire | 4128 (1.40) |
| Health report | 4055 (1.37) |
| Reminder item | 3287 (1.11) |
| Appointment | 3209 (1.08) |
| Weight history | 3150 (1.06) |
| Reminder | 2434 (0.82) |
| Drug history | 2286 (0.77) |
| Show appointment | 1606 (0.54) |
| Task uncomfortable | 1345 (0.45) |
| Person information | 1267 (0.43) |
| Monthly report | 1144 (0.39) |
| Knowledge collect | 1065 (0.36) |
| Knowledge list by tag | 742 (0.25) |
| Blood pressure guidance | 721 (0.24) |
| Setting | 207 (0.07) |
| Food history | 174 (0.06) |
| Account | 125 (0.04) |
| Log-in | 113 (0.04) |
| Register | 20 (0.01) |

[a]Descending by quantity (relative) of behavioral states.

**Table 2.** Time spent in each behavioral state.

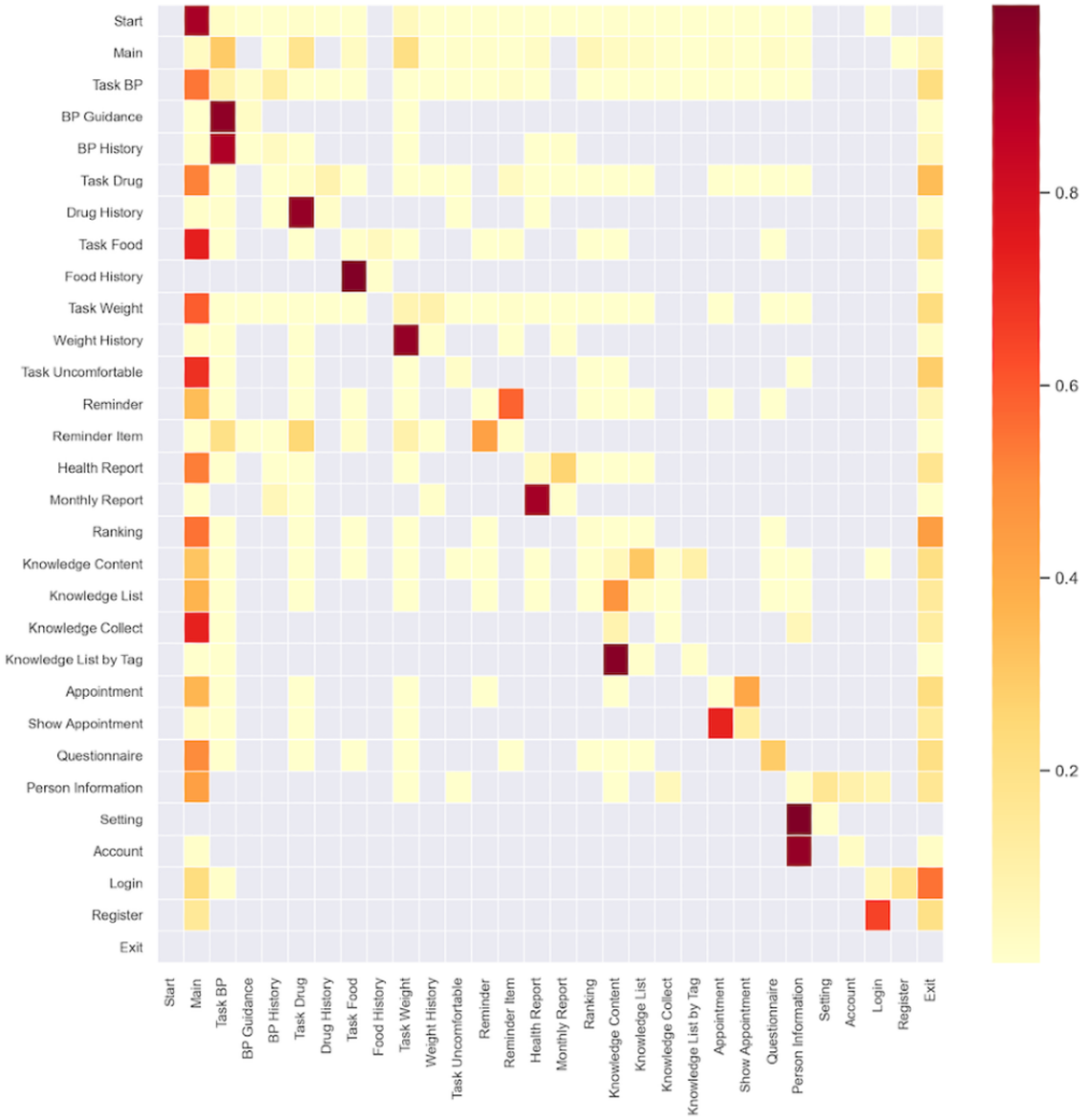
| Behavioral state[a] | Time spent (total), s | Time spent (mean), s |
|---|---|---|
| Show appointment | 261,371 | 162.75 |
| Knowledge content | 615,430 | 83.96 |
| Task food | 149,164 | 35.87 |
| Monthly report | 39,590 | 34.61 |
| Register | 639 | 31.95 |
| Blood pressure history | 185,240 | 30.52 |
| Blood pressure guidance | 19,772 | 27.42 |
| Task blood pressure | 1,253,924 | 25.62 |
| Log-in | 2575 | 22.78 |
| Task drug | 600,104 | 22.64 |
| Task weight | 746,268 | 22.24 |
| Knowledge list | 100,995 | 20.73 |
| Ranking | 154,468 | 19.66 |
| Food history | 3391 | 19.49 |
| Task uncomfortable | 23,474 | 17.45 |
| Account | 2110 | 16.88 |
| Weight history | 52,218 | 16.58 |
| Reminder item | 52,493 | 15.97 |
| Main | 1,745,286 | 13.91 |
| Knowledge list by tag | 10,129 | 13.65 |
| Health report | 53,730 | 13.25 |
| Drug history | 28,309 | 12.38 |
| Questionnaire | 31,579 | 7.65 |
| Setting | 1451 | 7.00 |
| Person information | 7779 | 6.14 |
| Appointment | 19,157 | 5.97 |
| Knowledge collect | 6094 | 5.72 |
| Reminder | 8426 | 3.46 |

[a]Descending by time spent (mean) of behavioral states.

### *Patient's Behavior Transition Matrix*

The heatmap was used to demonstrate the transition matrix of the Markov chain (see Figure 3). The various shades of color represent the probability of transition from one behavioral state to another, and the transparent color indicate that these 2 behavioral states cannot be transitioned. Horizontal and vertical coordinates indicate behavioral states. The rows summed up to 1. In most cases, the session started from the main state (0.91), and the transition probability from other behavioral states to the main state was also high, which was intended by design. When patients were in the main state, the states of task BP (0.29), task drug (0.18), and task weight (0.20) were the most visited among all behavioral states. Patients had a high probability if exiting the app from the following states: task drug (0.34), ranking (0.44), and log-in (0.55).

**Figure 3.** Transition matrix of patient multibehavior engagement. Probability of transitioning from behavioral state A[row] to B[column]. BP: blood pressure.

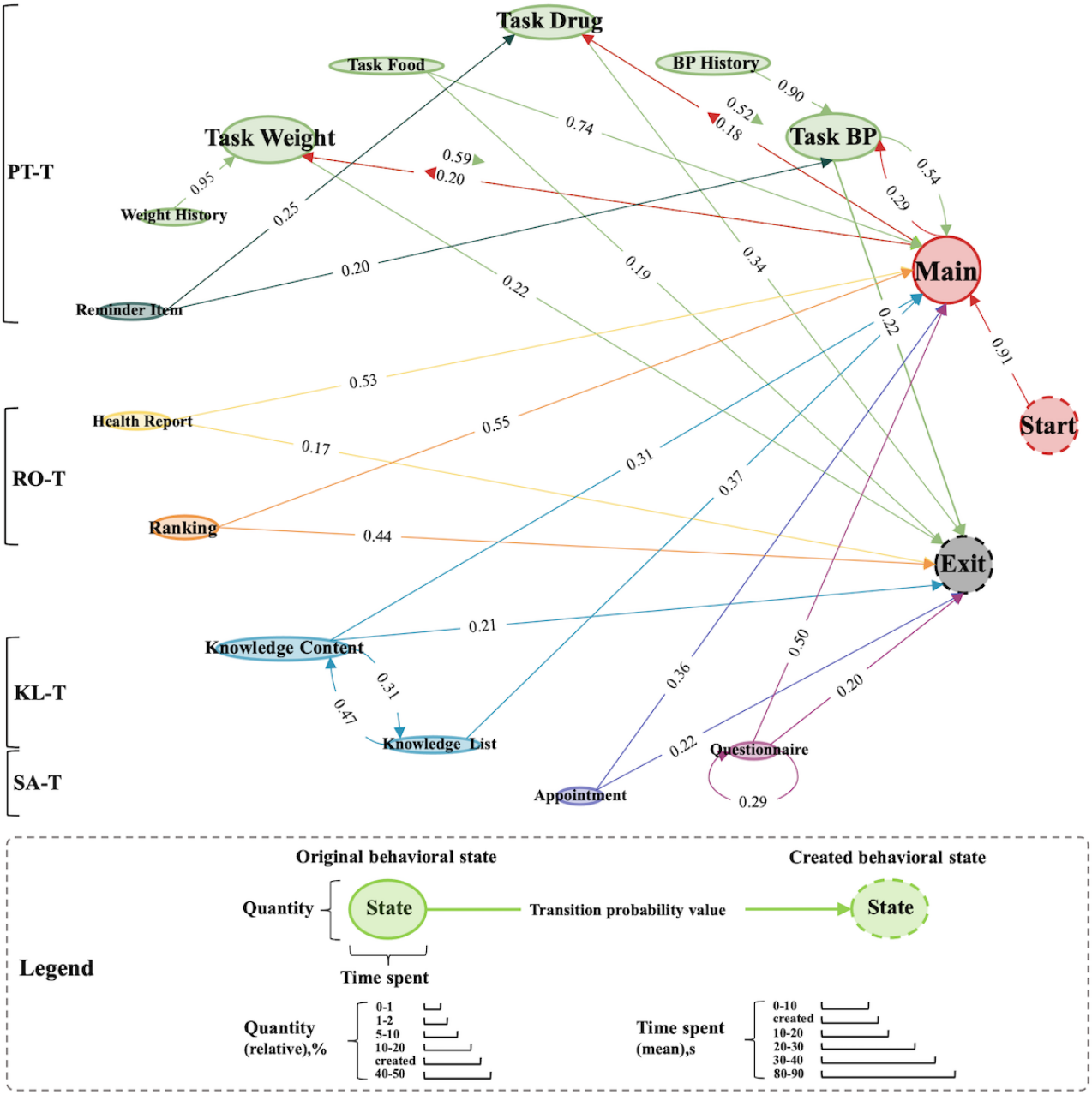



We found that the 1-way transition ($s_i \rightarrow s_j$) from one behavioral state to another had a high probability in the following states: from BP guidance (0.96) or BP history (0.90) to task BP, from drug history to task drug (0.95), from food history to task food (0.99), from weight history to task weight (0.95), from reminder item to task BP (0.20) or task drug (0.25), from knowledge list by tag to knowledge content (0.98), and from account to person information (0.95). The cycle-transitions ($s_i \Leftrightarrow s_j$) between the 2 behavioral states had a high probability in the following states: between reminder and reminder item (0.58, 0.43), between health report and monthly report (0.26, 0.92), between knowledge content and knowledge list (0.31, 0.47), between appointment and show appointment (0.41, 0.73), between person information and setting (0.16, 1.00), and between log-in and register (0.17, 0.65). We found that the self-transition ($s_i \not\rightarrow s_i$) in a behavioral state was very common, and a high self-transition can be seen in the questionnaire state (0.29).

### *Four Types of Patient Multibehavior Engagement Trails*

There were 28 different types of behavior states and 283 possible behavior transitions in the original design in mHealth hypertension management app. The trails of patient multibehavior engagement in the pathway-based hypertension self-management were visualized by the main transitions between the different behavioral states (see Figure 4). The main trails of patient behavior included the type, quantity, time spent, sequence, and TP value. The size of a node indicated the quantity and time spent of behavioral states; a node or line of the same or similar color indicated a trail. Visual inspection of the transition between different behavioral states in mHealth hypertension self-management revealed several types of trails with different behavioral characteristics.

**Figure 4.** Four types of patient multibehavior engagement trails. BP: blood pressure.




A first trail type can be labeled as a perform task trail (PT-T; green and dark green lines). After launching the app, patients go from the main state to the self-management task states to conduct management plans. Among the 5 self-management tasks, patients were more likely to transfer to the states of task BP, task drug, and task weight (TP value 0.29, 0.18, 0.20) than the states of task food, and task uncomfortable (TP value 0.03, 0.01). Moreover, some patients entered the states of task BP and task drug (TP value 0.20, 0.25) from the reminder item state, which happened when patients needed the app to remind themselves to measure BP and take medication on time. Patients also checked their BP history and weight history when they conduct the self-management tasks. After finishing the self-management tasks, there was a high probability that patients would return to the main state. The mean time spent was longer in the task food (35.87 s) state than that in the states of task BP, task drug, and task weight, and patients spent more time checking BP history (30.52 s) than weight history (16.58 s).

Second, a result-oriented trail (RO-T) can be distinguished (yellow and orange lines). Patients had a high probability in the states of health report and ranking in this trail, and the purpose was to check their self-management outcomes and their ranking among all hypertensive patients. The mean time spent in the ranking (19.66 s) state was longer than the health report (13.25 s) state, which showed that when competing with other patients, patients were more willing to spend time to understand the results of self-management. Eventually, this trail usually returned to the main state.

The third trail type was labeled a knowledge learning trail (KL-T; blue lines). In this trail, patients preferred to read the health education content. In addition, there was a high probability of cycle transition (TP value 0.47, 0.31) between

the knowledge list state and the knowledge content state. This situation occurred when patients switched to a new piece of knowledge content after reading a previous piece of knowledge content. Patients spent an average of 83.96 s to reading a piece of knowledge content, and they spent an average of 20.73 s in the knowledge list state to select a piece of knowledge content to read. The end of this trail was sometimes the main state.

The last trail type was labeled a support acquisition trail (SA-T; purple and rese red lines). The states of questionnaire and appointment indicated the behavior of patients providing information to health care providers and seeking support from health care providers. There was a high probability of self-transition in the questionnaire (TP value 0.29) state, indicating the patients entered the questionnaire state, then switched to another app but did not exit the app, and then reopened the app on the questionnaire page. The end of this trail was sometimes the main state.

## Patient Behavior Preference Patterns

A total of 863 patients were selected for cluster analysis. We found that the silhouette score was the highest with 3 clusters of patients (see Figure 5). Hence, we accepted the 3-cluster output of K-means for further analysis. The patient behavior preferences in 4 trails were significantly different (PT-T $P<.001$, RO-T $P=.06$, KL-T $P<.001$, SA-T $P<.001$).

There were 3 distinctive patterns of patient behavior preferences (see Figure 6). The sum of each patient's behavior preferences for the 4 behavior trails was 1. The first cluster (PT-T) comprised 694 patients. Their behavior preference was particularly focused on PT-T (0.81), which indicated that this patient group preferred to perform self-management tasks. The second cluster (PT-T and KL-T) comprised 96 patients, who were active in PT-T (0.37) and KL-T (0.53). They were more likely to read knowledge about hypertension than to conduct self-management tasks. The third cluster (PT-T and SA-T) comprised 73 patients whose behavior preferences were PT-T (0.30) and SA-T (0.37). These patients were more willing to seek help from health care providers than to perform self-management tasks.

**Figure 5.** Comparison of the silhouette score for different number of clusters (range 2-9).

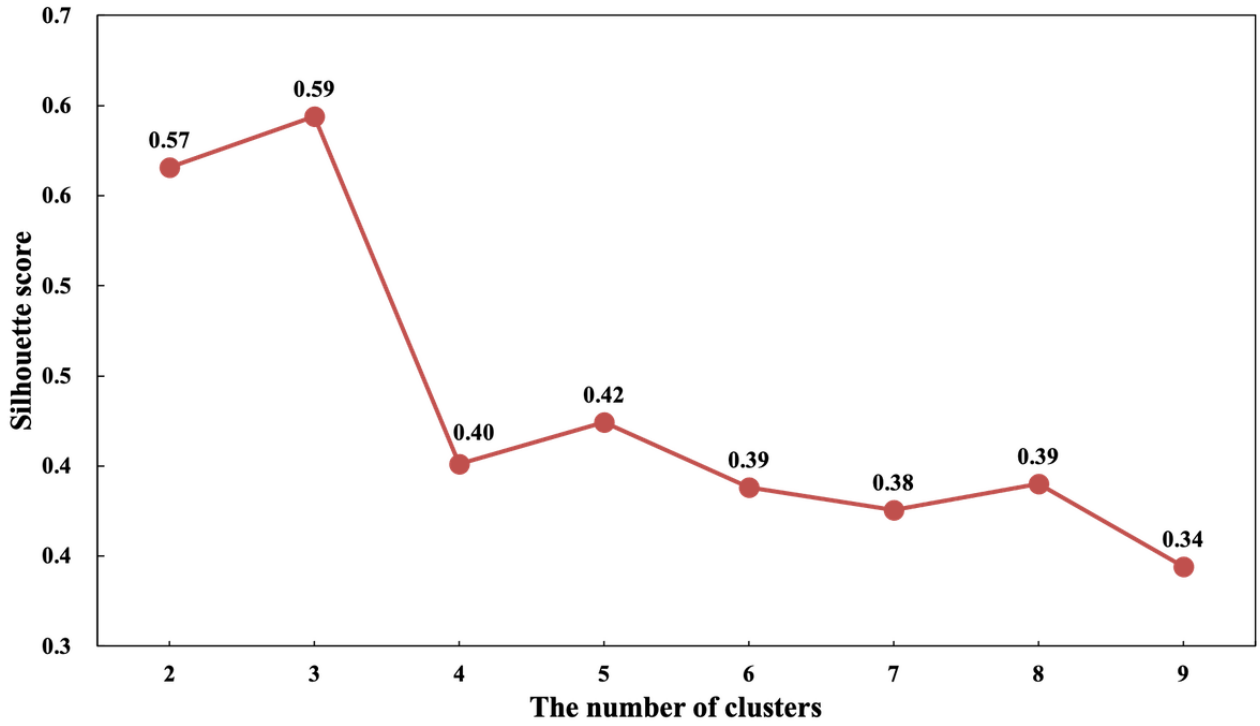
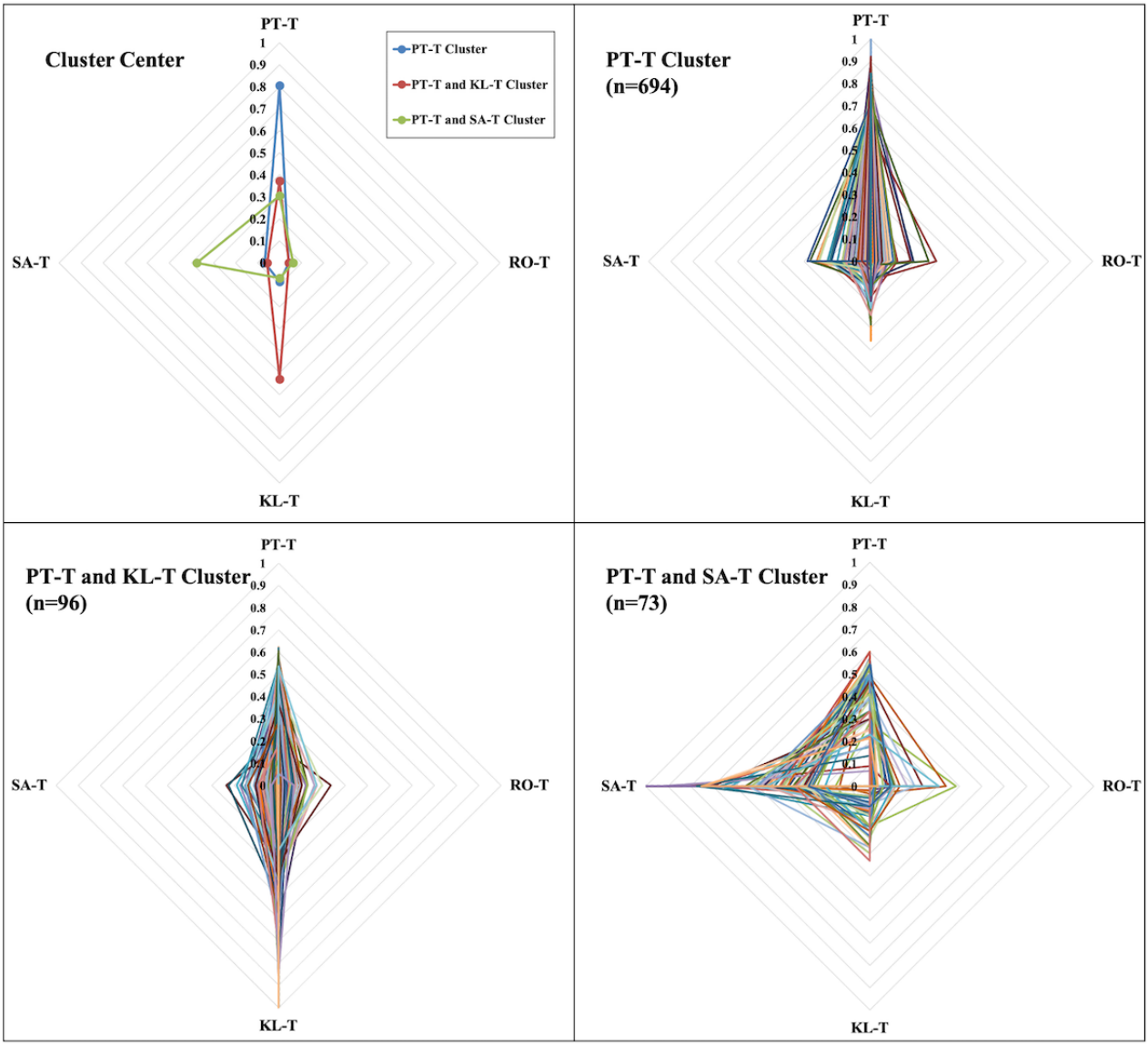


**Figure 6.** Three patterns of patient behavior preference.




## Demographics and Physiological Characteristics of the 3 Behavior Preference Patterns

There were statistically significant associations between the behavior preference pattern and gender, education level, and BP, but there were not associations between the behavior preference pattern and age and HR (see Table 3). In the PT-T cluster and the PT-T and SA-T cluster, there were much more male patients than female patients, but the proportion of male and female patients was equal in the PT-T and KL-T cluster. Compared with the PT-T cluster, the PT-T and KL-T cluster was characterized by significantly fewer male patients, lower education level, and higher BP. The PT-T and SA-T cluster had significantly lower BP than the PT-T and KL-T cluster.

**Table 3.** Descriptive characteristics of the behavior preference pattern. Multiple comparisons of the 4 clusters (at the .05 level).

| Characteristic | PT-T[a] cluster (n=694) | PT-T and KL-T[b] cluster (n=96) | PT-T and SA-T[c] cluster (n=73) | *P* value |
|---|---|---|---|---|
| Age, mean (SD) | 53.20 (9.93) | 52.23 (9.99) | 50.86 (10.34) | .13 |
| **Gender, n (%)** | —[d] | — | — | .003 |
| Male | 469 (67.58) | 48 (50.00)[e] | 49 (67.12) | — |
| Female | 225 (32.42) | 48 (50.00)[e] | 24 (32.88) | — |
| **Education, n (%)** | — | — | — | .05 |
| < High school | 249 (35.88) | 43 (44.79) | 34 (46.57) | — |
| High school | 105 (15.13) | 21 (21.88) | 11 (15.07) | — |
| ≥ University | 321 (46.25) | 30 (31.25)[e] | 25 (34.25) | — |
| Don’t know | 19 (2.74) | 2 (2.08) | 3 (4.11) | — |
| SBP[f], mean (SD) | 132.71 (15.43) | 137.44 (17.96)[e] | 130.77 (13.18)[g] | .02 |
| DBP[h], mean (SD) | 86.46 (11.37) | 90.60 (12.88)[e] | 84.87 (11.29)[g] | .004 |
| HR[i], mean (SD) | 72.41 (14.23) | 73.20 (17.13) | 73.23 (15.83) | .84 |

[a]PT-T: perform task trail.

[b]KL-T: knowledge learning trail.

[c]SA-T: support acquisition trail.

[d]Not applicable.

[e]Given cluster is significantly different from the PT-T cluster.

[f]SBP: systolic blood pressure.

[g]Given cluster is significantly different from the PT-T and KL-T cluster.

[h]DBP: diastolic blood pressure.

[i]HR: heart rate.

# Discussion

## Principal Findings

In this study, we used a stochastic model to describe longitudinal trails of patient multibehavior engagement with mHealth hypertension self-management and further analyzed the characteristics of patient groups with different behavior preferences. In a sample of 863 patients in mHealth hypertension self-management, we identified 3 types of behavior transition (1-way transition, cycle transition, and self-transition), 4 trails of patient multibehavior engagement (PT-T, RO-T, KL-T, and SA-T), and 3 behavior preference patterns (PT-T cluster, PT-T and KL-T cluster, and PT-T and SA-T cluster).

These insights revealed what actual patients do in daily hypertension self-management and how patients use the mHealth app to conduct self-management [32,43,47]. This may facilitate tailored and precise behavioral intervention strategies with specific content, methods, and time points for patient groups with specific behavior preferences. For example, for patient groups who measure and record BP but not take medication regularly, we will send a pop-up medication reminder in the app after completing the BP recording, rather than randomly and frequently reminding patients to measure BP and take medication. After they follow this reminder, we will provide rewards and push the knowledge about the advantages of regular medication to improve their adherence. This may promote more accurate chronic disease management strategies to achieve disease control.

## Patient Multibehavior Engagement Trails

The dynamic behavioral characteristics were identified by the transition probability between 2 continuous behaviors [39,47], which suggested the transitions in daily self-management behavior of hypertension patients. When patients started self-management, they were more likely to focus on BP, medications, and weight conditions. Patients prefer to pay attention to BP and medications from reminders, which can help them manage their conditions. Tao et al [48] found that the use of electronic reminders was associated with a significant improvement in patient adherence to medication. The cycle transition was often performed between 2 different behaviors of same module, such as reminder, health report, knowledge, and appointment, etc. This kind of behavior transition information gives us opportunities to provide specific behavioral interventions at an accurate time between 2 consecutive behaviors. In addition, self-transition was frequent when patients completed questionnaires. Patients were willing to engage in the questionnaire, but the time spent was very short. It may be because the questionnaire was too long and patients had no patience when completing it or the content of the questionnaire was difficult to understand. The clarity and conciseness of questionnaires should be improved so patients can better engage in this management part [49].

The trails of patient multibehavior engagement revealed the high probability sequences of patient behavior in mHealth hypertension self-management. This was a longitudinal and multidimensional process analysis rather than a cross-sectional analysis at a point in time. The 4 trails contained the main scenes of hypertension management in daily life [50] and described the behavioral conditions of patient self-management. In PT-T, for 5 types of self-management tasks of improving ability, patients preferred to start self-management by focusing on BP, medications and weight and spent more time on food. This may be attributed to diet management being more complicated and difficult to implement and record [51], which proposes a new challenge of how to improve the convenience of hypertension diet management. Samiul et al [52] proposed a network for automatic dietary monitoring that can gather food intake information through image, audio, and accelerometer sensors. By analyzing these data, the system can measure the type, volume, and nutrition of food, as well as the eating behavior of a person. Compared with weight history, patients spent more time reading BP history. In KL-T, patients usually chose different health knowledge contents from knowledge list to learn how to improve self-management ability. This knowledge may help patients broadly learn health concepts and behaviors [53]. In RO-T, patients read health reports to understand all aspects of their self-management, and knew the ranking of their self-management outcomes in an online hypertension community from the leaderboard. An interesting finding was that patients spent more energy on the rankings compared to reading health reports. This may suggest that gamification and competitive methods can effectively increase patient motivation and attention to their own self-management states [19,21,54]. In SA-T, patients sought help from health care providers through appointments, and completed the questionnaire to provide their own information and support research on hypertension management. These findings helped us dive into the daily self-management of actual patients and deeply understand patient behavioral characteristics through a longitudinal and multidimensional method. This provides an opportunity to apply some health behavior intervention techniques to promote patients to change their behavior and improve the effectiveness of chronic management results [50,55].

### Patient Behavior Preference Patterns

The behavior preference patterns of patients represent the individual inherent habits of patients in daily hypertension self-management [34-36], which is an essential factor to the design of management strategies. In this study, we found 3 behavior preference patterns of patients engaging in the pathway-based mHealth hypertension management, reflecting which patient groups preferred which type of behavior to achieve better self-management results. All of 3 patterns preferred PT-T, 1 preferred KL-T, and the other preferred SA-T, but there was no obvious preference for RO-T.

Patients with these 3 behavior preference patterns had significantly different demographic (gender and education level) and physiological (BP) characteristics. Compared with other patterns, patients with the preference for PT-T and KL-T had a lower education level, higher BP, and were much less likely to be male. First of all, there were more male patients than female patients, possibly because males were more inclined to use mHealth services for hypertension management [56,57]. However, we found that female patients were more willing to read knowledge content than male patients. Our finding is consistent with the finding by Al-Ansari et al [58] that oral health knowledge and health behavior were statistically significantly higher among the females than the males. Moreover, in this study, we found that patients with lower education level and poorly controlled BP were more likely to be highly involved in hypertension health education. Some scientific literature demonstrated a strong association between lower levels of education and poor health outcomes [59,60]. Health literacy was a potential pathway between levels of education and health outcomes [61]. Lee et al [62] and Nutbeam et al [63] also found that in comparison with people with higher education, people with lower education level were found to demonstrate lower health literacy and poor health. Naturally, these patients had a strong enthusiasm for learning about hypertension to promote their health literacy and improve health outcomes. The findings of patient preferences and characteristics are useful for designing personalized functions in the mHealth app to improve patient ability and engagement in hypertension self-management [34-36].

### Strengths and Limitations

The strengths of this study were that, first of all, the study used time series data to unveil the longitudinal and multidimensional behavioral characteristics of patient engaging in mHealth hypertension management, which contained the type, quantity, time spent, sequence, and transition probability of patient behavior. The findings help provide more precise timing of mHealth behavior interventions for hypertension management. To the best of our knowledge, there has been no literature that uses longitudinal data to describe the trails of patient multibehavior engagement in hypertension self-management. Secondly, we analyzed the demographic and physiological characteristics of patients with different behavior preferences. Our findings revealed the association between behavior preferences, demographics, and physiology. This provided information for the further design of the appropriate types of interventions for specific patient groups. Finally, we used data from real-world patients, which revealed patient behavior in the natural setting rather than attracting patients more likely to bear the burden of research-related evaluation. This helped to generate practical insights for the behavior of actual patients in daily hypertension management.

This study has limitations. One limitation was that the data were only from an mHealth hypertension app, which implies that the sample is presumably not representative of all patient groups. We also don’t know the behavior of patients who engage in self-management but not use the app, such as older patients who cannot use the mobile devices. Moreover, the behavioral characteristics of patients may be influenced and limited by the design of the app. Second, we used a first-order Markov chain, which simplified our analysis. Some other machine learning methods such as higher-order Markov chains should be applied to better represent global patient behavior. Finally, the behavior preferences can be caused by various demographic and social psychological characteristics of patients (such as marital status,

profession, anxiety, depression, etc), so these factors need to be considered in the future studies.

## Conclusion

The mHealth services are an effective way to conduct out-of-hospital health care services for chronic disease. In this study, we have found 3 types of behavior transitions, 4 trails of patient multibehavior engagement, and 3 patient behavior preference patterns in the pathway-based mHealth hypertension self-management. These findings gained insights of actual behavior of patients in daily hypertension self-management. The behavioral characteristics of patients were dynamic, longitudinal, and multidimension, which may create opportunities to design tailored, personalized interventions within a specific behavior time to change patient's behavior, develop positive behavioral habits, and continuously improve and maintain the effect of hypertension management.

## Acknowledgments

This study was supported by grants 2020YFC2003403 and 2020YFC2006405 from the National Key Research and Development Programs of China, grant 2020BFG02002 from the Key Research and Development Program of Ningxia, China, and grant 2021AB11007 from the Key Research and Development Program of Guangxi, China. This work was also supported by the Alibaba Cloud. The authors would like to thank the General Hospital of Ningxia Medical University.

## Authors' Contributions

DW, XH, and ND were the primary authors of this paper, oversaw all aspects of the study design and execution, and led the writing of the manuscript. YS and JH led all data analysis with oversight from DW and XH. HD and ND provided critical revision of the manuscript for important intellectual content.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed information about behavioral states.
[DOCX File , 15 KB - mhealth_v10i2e33189_app1.docx ]

## References

1. Global Status Report On Noncommunicable Diseases 2014. Geneva: World Health Organization; 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf;jsessionid=3C0C9A267DA218A8836BDCA9EB2C5C0D?sequence=1 [accessed 2022-01-14]
2. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
3. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
4. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
5. Wong A, Bhyat R, Srivastava S, Boissé Lomax L, Appireddy R. Patient care during the COVID-19 pandemic: use of virtual care. J Med Internet Res 2021 Jan 21;23(1):e20621 [FREE Full text] [doi: 10.2196/20621] [Medline: 33326410]
6. Cheung K, Ling W, Karr C, Weingardt K, Schueller S, Mohr D. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. J Am Med Inform Assoc 2018 Aug 01;25(8):955-962 [FREE Full text] [doi: 10.1093/jamia/ocy023] [Medline: 29659857]
7. Chen X, Koskela M, Hyväkkä J. mHealth new horizons for health through mobile technologies. Proc Int Work Content-Based Multimed Index 2010;3:130-135. [doi: 10.1109/CBMI.2010.5529886]
8. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
9. 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 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
10. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]

11. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
12. 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 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
13. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
14. Sakakibara BM, Ross E, Arthur G, Brown-Ganzert L, Petrin S, Sedlak T, et al. Using mobile-health to connect women with cardiovascular disease and improve self-management. Telemed J E Health 2017 Dec;23(3):233-239. [doi: 10.1089/tmj.2016.0133] [Medline: 27623231]
15. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Ann Intern Med 2003 Feb 04;138(3):256-261. [doi: 10.7326/0003-4819-138-3-200302040-00034] [Medline: 12558376]
16. 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 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
17. Rehman H, Kamal AK, Morris PB, Sayani S, Merchant AT, Virani SS. Mobile health (mHealth) technology for the management of hypertension and hyperlipidemia: slow start but loads of potential. Curr Atheroscler Rep 2017 Mar;19(3):12. [doi: 10.1007/s11883-017-0649-y] [Medline: 28210974]
18. Cheikh-Moussa K, Mira JJ, Orozco-Beltran D. Improving engagement among patients with chronic cardiometabolic conditions using mHealth: critical review of reviews. JMIR Mhealth Uhealth 2020 Apr 08;8(4):e15446 [FREE Full text] [doi: 10.2196/15446] [Medline: 32267239]
19. Cechetti NP, Bellei EA, Biduski D, Rodriguez JPM, Roman MK, De Marchi ACB. Developing and implementing a gamification method to improve user engagement: a case study with an m-Health application for hypertension monitoring. Telematics Informatics 2019 Aug;41:126-138. [doi: 10.1016/j.tele.2019.04.007]
20. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. Uptake of a consumer-focused mHealth application for the assessment and prevention of heart disease: the <30 days study. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
21. 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 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
22. James J. Health policy brief: patient engagement. Health Affairs 2013 Feb 14. [doi: 10.1377/hpb20130214.898775]
23. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
24. Tyagi H, Sabharwal M, Dixit N, Pal A, Deo S. Leveraging providers' preferences to customize instructional content in information and communications technology-based training interventions: retrospective analysis of a mobile phone-based intervention in India. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e15998 [FREE Full text] [doi: 10.2196/15998] [Medline: 32130191]
25. 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 Dec 7;5(1):4 [FREE Full text] [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]
26. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
27. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82. [doi: 10.2196/mhealth.7336] [Medline: 28630037]
28. 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 Dec;31(12):795-800 [FREE Full text] [doi: 10.1038/jhh.2017.69] [Medline: 28972573]
29. Sanders I, Short CE, Bogomolova S, Stanford T, Plotnikoff R, Vandelanotte C, et al. Characteristics of adopters of an online social networking physical activity mobile phone app: cluster analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e12484 [FREE Full text] [doi: 10.2196/12484] [Medline: 31162130]
30. 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]

31. Wu D, An J, Yu P, Lin H, Ma L, Duan H, et al. Patterns for patient engagement with the hypertension management and effects of electronic health care provider follow-up on these patterns: cluster analysis. J Med Internet Res 2021 Sep 28;23(9):e25630 [FREE Full text] [doi: 10.2196/25630] [Medline: 34581680]
32. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, Manage My Pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
33. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
34. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
35. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
36. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
37. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, Manage My Pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
38. Stragier J, Vandewiele G, Coppens P, Ongenae F, Van den Broeck W, De Turck F, et al. Data mining in the development of mobile health apps: assessing in-app navigation through Markov chain analysis. J Med Internet Res 2019 Jun 07;21(6):e11934 [FREE Full text] [doi: 10.2196/11934] [Medline: 31237838]
39. Borges J, Levene M. Evaluating variable-length Markov chain models for analysis of user web navigation sessions. IEEE Trans Knowl Data Eng 2007 Apr;19(4):441-452. [doi: 10.1109/TKDE.2007.1012]
40. Singer P, Helic D, Taraghi B, Strohmaier M. Detecting memory and structure in human navigation patterns using Markov chain models of varying order. PLoS One 2014 Jul 11;9(7):e102070 [FREE Full text] [doi: 10.1371/journal.pone.0102070] [Medline: 25013937]
41. Puspitasari I, Moriyama K, Fukui K, Numao M. Effects of individual health topic familiarity on activity patterns during health information searches. JMIR Med Inform 2015 Mar 17;3(1):e16 [FREE Full text] [doi: 10.2196/medinform.3803] [Medline: 25783222]
42. Fackrell M. Modelling healthcare systems with phase-type distributions. Health Care Manag Sci 2009 Mar 7;12(1):11-26. [doi: 10.1007/s10729-008-9070-y] [Medline: 19938439]
43. Carreras M, Ibern P, Coderch J, Sánchez I, Inoriza JM. Estimating lifetime healthcare costs with morbidity data. BMC Health Serv Res 2013 Oct 25;13:440 [FREE Full text] [doi: 10.1186/1472-6963-13-440] [Medline: 24156613]
44. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
45. Wang Z, Li C, Huang W, Chen Y, Li Y, Huang L, et al. Effectiveness of a pathway-driven eHealth-based integrated care model (PEICM) for community-based hypertension management in China: study protocol for a randomized controlled trial. Trials 2021 Jan 22;22(1):81 [FREE Full text] [doi: 10.1186/s13063-021-05020-2] [Medline: 33482896]
46. Heyse J, Vandewiele G, Ongenae F, De Backere F, De Turck F. How do users interact with mobile health apps? A Markov chain analysis. Proc 14th EAI Int Conf Pervasive Comput Technol Healthc 2020:255-258. [doi: 10.1145/3421937.3421967]
47. Stragier J, Vandewiele G, Coppens P, Ongenae F, Van den Broeck W, De Turck F, et al. Data mining in the development of mobile health apps: assessing in-app navigation through Markov chain analysis. J Med Internet Res 2019 Jun 07;21(6):e11934 [FREE Full text] [doi: 10.2196/11934] [Medline: 31237838]
48. Tao D, Xie L, Wang T, Wang T. A meta-analysis of the use of electronic reminders for patient adherence to medication in chronic disease care. J Telemed Telecare 2015 Jan;21(1):3-13. [doi: 10.1177/1357633X14541041] [Medline: 25147178]
49. Dell-Kuster S, Sanjuan E, Todorov A, Weber H, Heberer M, Rosenthal R. Designing questionnaires: healthcare survey to compare two different response scales. BMC Med Res Methodol 2014 Aug 03;14:96 [FREE Full text] [doi: 10.1186/1471-2288-14-96] [Medline: 25086869]
50. 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 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
51. Hever J, Cronise R. Plant-based nutrition for healthcare professionals: implementing diet as a primary modality in the prevention and treatment of chronic disease. J Geriatr Cardiol 2017 May;14(5):355-368 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2017.05.012] [Medline: 28630615]

52. Mamud S, Bandyopadhyay S, Chatterjee P, Bhandari S, Chakraborty N. A Body Sensor Network for Automatic Dietary Monitoring System. Singapore: Springer; 2021.
53. Wang Z, Huang H, Cui L, Chen J, An J, Duan H, et al. Using natural language processing techniques to provide personalized educational materials for chronic disease patients in China: development and assessment of a knowledge-based health recommender system. JMIR Med Inform 2020 Apr 23;8(4):e17642 [FREE Full text] [doi: 10.2196/17642] [Medline: 32324148]
54. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
55. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
56. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
57. Khatun F, Heywood AE, Hanifi SMA, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
58. Al-Ansari JM, Honkala S. Gender differences in oral health knowledge and behavior of the health science college students in Kuwait. J Allied Health 2007;36(1):41-46. [Medline: 17425190]
59. Mackenbach JP, Stirbu I, Roskam AR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008 Jun 5;358(23):2468-2481. [doi: 10.1056/NEJMsa0707519] [Medline: 18525043]
60. Kunst AE, Bos V, Lahelma E, Bartley M, Lissau I, Regidor E, et al. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. Int J Epidemiol 2005 Apr;34(2):295-305. [doi: 10.1093/ije/dyh342] [Medline: 15563586]
61. van Der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18 Suppl 1:172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]
62. Lee SD, Tsai T, Tsai Y, Kuo KN. Health literacy, health status, and healthcare utilization of Taiwanese adults: results from a national survey. BMC Public Health 2010 Oct 16;10:614 [FREE Full text] [doi: 10.1186/1471-2458-10-614] [Medline: 20950479]
63. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]


## Abbreviations

**ANOVA:** analysis of variance
**BP:** blood pressure
**DBP:** diastolic blood pressure
**HR:** heart rate
**KL-T:** knowledge learning trail
**mHealth:** mobile health
**PT-T:** perform task trail
**RO-T:** result-oriented trail
**SA-T:** support acquisition trail
**SBP:** systolic blood pressure
**TP value:** transition probability value


*Edited by L Buis; submitted 27.08.21; peer-reviewed by L Min, S Nan; comments to author 11.10.21; revised version received 21.10.21; accepted 20.12.21; published 03.02.22.*

*Please cite as:*
*Wu D, Huyan X, She Y, Hu J, Duan H, Deng N*
*Exploring and Characterizing Patient Multibehavior Engagement Trails and Patient Behavior Preference Patterns in Pathway-Based mHealth Hypertension Self-Management: Analysis of Use Data*
*JMIR Mhealth Uhealth 2022;10(2):e33189*
*URL: https://mhealth.jmir.org/2022/2/e33189*
*doi:10.2196/33189*
*PMID:35113032*

Original Paper

# Validity and Feasibility of the Monitoring and Modeling Family Eating Dynamics System to Automatically Detect In-field Family Eating Behavior: Observational Study

Brooke Marie Bell[1,2], PhD; Ridwan Alam[3,4], MSc, PhD; Abu Sayeed Mondol[5], MSc, PhD; Meiyi Ma[5,6], MSc, PhD; Ifat Afrin Emi[5], MSc, PhD; Sarah Masud Preum[5,7], MSc, PhD; Kayla de la Haye[2], PhD; John A Stankovic[5], MSc, PhD; John Lach[4,8], MSc, PhD; Donna Spruijt-Metz[2,9,10], MFA, PhD

[1]Department of Chronic Disease Epidemiology, School of Public Health, Yale University, New Haven, CT, United States

[2]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[3]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]Department of Electrical and Computer Engineering, School of Engineering and Applied Science, University of Virginia, Charlottesville, VA, United States

[5]Department of Computer Science, School of Engineering and Applied Science, University of Virginia, Charlottesville, VA, United States

[6]Department of Computer Science, School of Engineering, Vanderbilt University, Nashville, TN, United States

[7]Department of Computer Science, Dartmouth College, Hanover, NH, United States

[8]School of Engineering and Applied Science, The George Washington University, Washington, DC, United States

[9]Center for Economic and Social Research, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA, United States

[10]Department of Psychology, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA, United States

**Corresponding Author:**
Brooke Marie Bell, PhD
Department of Chronic Disease Epidemiology
School of Public Health
Yale University
60 College St
8th Floor
New Haven, CT, 06520
United States
Phone: 1 475 235 0643
Email: brooke.bell@yale.edu

## Abstract

**Background:** The field of dietary assessment has a long history, marked by both controversies and advances. Emerging technologies may be a potential solution to address the limitations of self-report dietary assessment methods. The Monitoring and Modeling Family Eating Dynamics (M2FED) study uses wrist-worn smartwatches to automatically detect real-time eating activity in the field. The ecological momentary assessment (EMA) methodology was also used to confirm whether eating occurred (ie, ground truth) and to measure other contextual information, including positive and negative affect, hunger, satiety, mindful eating, and social context.

**Objective:** This study aims to report on participant compliance (feasibility) to the 2 distinct EMA protocols of the M2FED study (hourly time-triggered and eating event–triggered assessments) and on the performance (validity) of the smartwatch algorithm in automatically detecting eating events in a family-based study.

**Methods:** In all, 20 families (58 participants) participated in the 2-week, observational, M2FED study. All participants wore a smartwatch on their dominant hand and responded to time-triggered and eating event–triggered mobile questionnaires via EMA while at home. Compliance to EMA was calculated overall, for hourly time-triggered mobile questionnaires, and for eating event–triggered mobile questionnaires. The predictors of compliance were determined using a logistic regression model. The number of true and false positive eating events was calculated, as well as the precision of the smartwatch algorithm. The Mann-Whitney *U* test, Kruskal-Wallis test, and Spearman rank correlation were used to determine whether there were differences in the detection of eating events by participant age, gender, family role, and height.

**Results:** The overall compliance rate across the 20 deployments was 89.26% (3723/4171) for all EMAs, 89.7% (3328/3710) for time-triggered EMAs, and 85.7% (395/461) for eating event–triggered EMAs. Time of day (afternoon odds ratio [OR] 0.60, 95% CI 0.42-0.85; evening OR 0.53, 95% CI 0.38-0.74) and whether other family members had also answered an EMA (OR 2.07, 95% CI 1.66-2.58) were significant predictors of compliance to time-triggered EMAs. Weekend status (OR 2.40, 95% CI 1.25-4.91) and deployment day (OR 0.92, 95% CI 0.86-0.97) were significant predictors of compliance to eating event–triggered EMAs. Participants confirmed that 76.5% (302/395) of the detected events were true eating events (ie, true positives), and the precision was 0.77. The proportion of correctly detected eating events did not significantly differ by participant age, gender, family role, or height ($P$>.05).

**Conclusions:** This study demonstrates that EMA is a feasible tool to collect ground-truth eating activity and thus evaluate the performance of wearable sensors in the field. The combination of a wrist-worn smartwatch to automatically detect eating and a mobile device to capture ground-truth eating activity offers key advantages for the user and makes mobile health technologies more accessible to nonengineering behavioral researchers.

***(JMIR Mhealth Uhealth 2022;10(2):e30211)*** doi:10.2196/30211

**KEYWORDS**

ecological momentary assessment; wearable sensors; automatic dietary assessment; eating behavior; eating context; smartwatch; mobile phone

## Introduction

### Challenges to Dietary Assessment

A prevailing challenge in dietary and eating research is the ability to accurately measure dietary intake. Historically, the assessment of dietary intake and eating behaviors uses self-reporting tools [1,2], such as food diaries, food frequency questionnaires, and 24-hour dietary recalls [3,4]. All dietary assessment self-report methods have some level of measurement error (difference between measured and true values) [5,6]. Dietary data collected via self-report methods may be misreported because of biases, such as recall or memory bias (when a respondent erroneously recalls their dietary intake) and social desirability bias (when a respondent desires to present oneself positively) [7-9]. Studies have also found that those with certain characteristics (eg, obese weight status and body image dissatisfaction) are more likely to underreport their energy intake [10,11].

### Shifting Focus From Dietary Intake to Eating Behavior and Context

The field of nutritional epidemiology has produced an abundance of studies that have examined the role of *dietary intake* (ie, what and how much is consumed) in human health and disease—specifically, macronutrients (eg, fats and carbohydrates), types of food, quality of food, dietary patterns, and more [12]. Decades of laboratory-based and observational research indicate that dietary intake is a critical component of chronic disease prevention [13]. However, the measurement of diet in *free-living* populations remains a significant challenge in the field. In addition, even if public health researchers can easily and accurately track free-living dietary intake, dietary intake patterns are notoriously difficult to change long-term [14].

*Eating behaviors and patterns* (ie, food choices and motives and feeding practices) and *context* (who is eating, when, where, with whom, etc) also play a significant role in the development of obesity and other chronic diseases, including type 2 diabetes and heart disease [15-20]. These findings indicate that the patterns and features of eating events may be key contexts that shape dietary intake, and thus could be more malleable features of eating behavior that could be intervened. However, the field lacks appropriate behavioral theories that provide a richer understanding of how eating behaviors vary across contexts and across time [21,22].

### Technology-Assisted Dietary Assessment

Emerging technologies offer a potential solution for the accurate assessment of dietary intake by addressing the limitations of self-reported dietary assessment methods. The incorporation of technologies into dietary assessment can improve the quality and validity of dietary data by passively measuring eating in naturalistic settings over long periods with minimal user interaction [23]. Two emerging technological advances in dietary assessment tools include the following:

1. Ecological momentary assessment (EMA): a data collection technique in which one's behavior is repeatedly sampled in real time and in context [24-26].
2. Wearable devices and sensors: allow for the passive collection of various data streams from the physical environment (eg, acoustic, visual, and inertial) [27].

EMA and wearable sensors are able to measure behavior near or just in time, thereby reducing or eliminating the recall bias that can affect retrospective self-report measures. In addition to improving the validity of data, these technologies offer the opportunity to measure eating behavior frequently and over long periods, allowing researchers to examine how it varies over multiple timescales (varies over the day, over the week, etc).

### Monitoring and Modeling Family Eating Dynamics Study

To address the limitations of traditional dietary assessment methods and theories, the Monitoring and Modeling Family Eating Dynamics (M2FED) study developed a sensor system that used smartphones as well as deployable and wearable sensors to collect synchronized real-time data on family eating behavior [28]. This study used the following: (1) wrist-worn

smartwatches containing inertial sensors (accelerometer and gyroscope) to automatically detect arm movements and hand gestures associated with eating; (2) EMA via smartphone to confirm whether the eating occurred and to measure other contextual information, such as who was present during the eating event and the current mood of the respondent; and (3) Bluetooth proximity beacons to determine the approximate location of the smartwatches.

Rather than focusing on dietary intake (caloric intake, portion sizes, etc), this study took a novel approach by measuring eating behaviors (ie, food choices and motives and feeding practices) and context (who is eating, when, where, with whom, etc). Family eating dynamics have yet to be measured and modeled dynamically to better contextualize our understanding of social influence processes within family systems. This paper begins the first step toward producing new models that develop behavioral theory, and it may enable the identification of temporally specific processes and events within the family system that can be targeted for personalized, context-specific, real-time feedback.

### Assessing Validity of Wearable Sensors

The validity of using wearable sensors to automatically assess eating behavior and context has been tested in both laboratory and field settings [27,29-31], indicating that the performance of the wearable sensors decreases in naturalistic settings (compared with controlled laboratory settings). Studies have used a variety of sensors (eg, microphones, cameras, smartwatches, and electromyography electrodes) to measure various dietary outcomes, including bites, chewing, swallowing, and duration of eating occasions [27,29-33]. A review by Bell et al [27] indicates that there is still a strong reliance on retrospective self-report methods (eg, end-of-day food diaries) to determine ground-truth eating activity to evaluate wearable sensors in the field. Given the aforementioned limitations of retrospective self-report methods to accurately assess diet, the M2FED study used event-contingent EMA to determine ground-truth eating activity in families. The use of EMA offers unique methodological advantages, such as the following:

- The ability to measure behavior near or just in time, thereby reducing recall bias and reducing participant burden.
- The ability to measure behavior at the location in which it actually occurs, thereby maximizing ecological validity [24].

The validity of this method has been tested in a few in-field studies [34,35]; however, it has not yet been tested in a family-based study.

### Assessing Feasibility of EMA

One disadvantage of using technologies for data collection is the potential for participant noncompliance. A recent systematic review and meta-analysis by Wen et al [36] found that compliance rates among EMA studies in youth samples were suboptimal; the weighted average compliance rate was 78.3%, falling under the recommended 80% compliance rate [24]. Many studies have explored EMA compliance for various behaviors in various populations [36-40], but the compliance rate for a family-based EMA study is underexplored. A recent EMA study involving mothers and their children found that mothers' presence may enhance children's compliance with EMA questionnaires [41], suggesting that family members and other social relations may be leveraged to increase compliance in future EMA studies.

### Study Aims

Therefore, the overall purpose of this study is to report on participant compliance (feasibility) to the 2 distinct EMA protocols of the study (hourly time-triggered and eating event–triggered assessments) and on the performance (validity) of the wearable sensor in automatically detecting eating events in a family-based study. Specifically, the primary aims of this study include the following:

- Aim 1A—evaluate participant compliance with the EMA protocol, (1) overall, (2) for hourly time-triggered survey assessments, and (3) for eating event–triggered survey assessments—and aim 1B—evaluate the impact of time (time of day, day of week, and deployment day), age, gender, family role, and compliance of other family members (whether another participating family member j had answered a survey that had been received within 15 minutes of focal person i's survey) on compliance.
- Aim 2A—evaluate the performance of the wrist-worn smartwatch to automatically detect eating events of participants at home—and aim 2B—determine whether there are systematic differences in the detection of eating events by age, gender, family role, and height.

## Methods

### Participants and Recruitment

#### *Eligibility*

The research team recruited families that contained at least two members (including at least one adult parent and one child between the ages of 11 and 18 years) living in Los Angeles County. Families with children aged <11 years were eligible to participate; however, children aged <11 years were not permitted to participate in the study. Families were not eligible to participate if one or more family members living at home did not primarily speak English. There were no demographic or disease-related exclusion criteria.

#### *Method of Recruitment*

Families were recruited in public spaces and at public events in Los Angeles County from May 2017 to August 2019. Snowball sampling was also used, such that participating families were offered an additional US $20 if they referred other eligible families that were successfully enrolled in the study.

All families that expressed interest and met the eligibility requirements were invited to participate in the study. An intake screening tool was administered over the phone by recruitment coordination staff to confirm eligibility before enrolling in the study.

This study was approved by the Institutional Review Board of the University of Southern California (UP-16-00227). All

parents provided informed written consent, and all children provided assent.

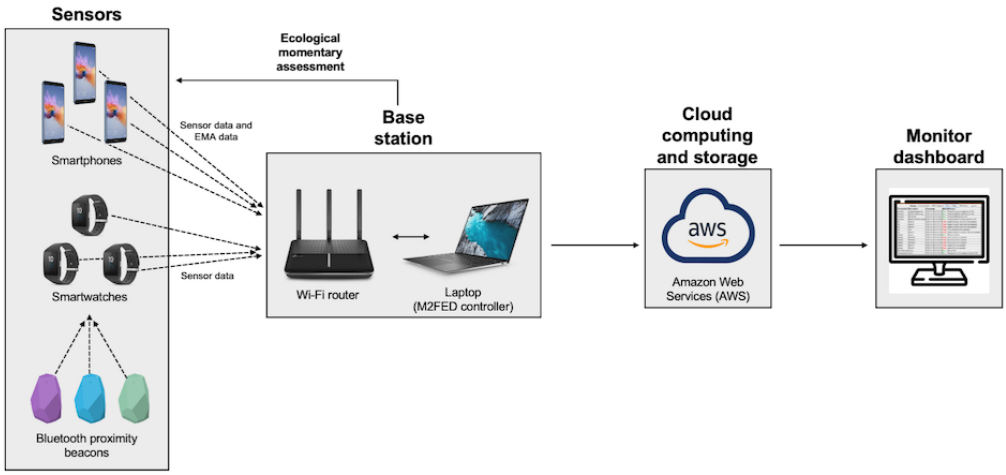
## M2FED System

### Overview

The primary objective of the M2FED study is to develop and deploy the M2FED cyberphysical system (Figure 1) in the homes of families. Cyberphysical systems can be defined as "physical and engineered systems whose operations are monitored, coordinated, controlled, and integrated by a computing and communication core" [42]. This novel system monitored *in-home* family eating behaviors in all participants. This system contained four primary components (1) sensors (including smartwatches, smartphones, and Bluetooth proximity beacons), (2) a base station, (3) an EMA subsystem, and (4) a remote monitoring subsystem, all of which were connected through a Wi-Fi router (Figure 1).

For the scope of this study, all data collected by the system were measured in the home (ie, no data were collected outside of the home).

**Figure 1.** Overview of the Monitoring and Modeling Family Eating Dynamics cyberphysical system. EMA: ecological momentary assessment.




### Sensors

Participants were instructed to wear a Sony Smartwatch 3 (Android Wear operating system) on their dominant hand during all waking hours that they were in their home. The smartwatches were used to *automatically* detect eating-related hand-to-mouth (H-t-M) gestures for each participant at home and in real time. Arm movements and H-t-M gestures were detected via an algorithm that used motion data from the inertial sensors inside the smartwatch (accelerometer and gyroscope) [43]. If a cluster of at least two H-t-M gestures were detected within a 1-minute time frame, then the motion data were processed with a more sophisticated algorithm, and these *clusters* were then characterized as an *eating event*. An eating event can be defined as a set of H-t-M gestures, representing phenomena such as consuming a meal, snack, drink, or a combination of these consumption behaviors in which H-t-M gestures are clustered temporally. The technical details of the eating event detection algorithm are provided in detail elsewhere [43]. Participants were instructed to wear the smartwatch only at home and to not take it outside or wear it outside of the home. Consequently, data on H-t-M gestures and eating events that were determined by the proximity beacons that occurred outside of the home were discarded.

Participants were each provided with a Samsung Galaxy S7 smartphone (Android operating system) preprogramed with limited functioning. The smartphone app in which they responded to mobile questionnaires was pinned to the screen so that they could not access other apps on the smartphone. This smartphone was only intended for use as a data collection tool. Participants were instructed to keep their smartphones at home and not take it outside of the home. If a smartphone left home and was not within the range of the Wi-Fi router, the phone did not receive any mobile questionnaires. Consequently, data on participants' states and behaviors outside of the home were not collected.

Estimote Bluetooth Low Energy proximity beacons were used to determine the approximate location of smartwatches of participants (including approximately which room the watches were in and whether they were still at home) during the study period. The beacons continuously broadcasted *packets* that included the unique media access control address of the Bluetooth interface, whereas the smartwatches periodically scanned for these *packets*. The smartwatches then recorded the received signal strength indicator (signal from the beacons), which indicated the proximity of the smartwatches to the beacons.

Typically, 1 to 2 beacons were placed on a wall in each living space at home (excluding bathrooms and bedrooms), and they required no further action by the participants during the study.

### Base Station

A base station is a radio receiver and transmitter and a computing platform that serves as the hub of a local wireless network (the M2FED system). The base station for the M2FED system was a Lenovo ThinkPad laptop, which was placed in the home of the family for the duration of the study. The laptop was placed in a locked cage so that it could not be tampered

with. The base station collected and processed the data received from the smartphones and smartwatches through the Wi-Fi router, and managed the EMA subsystem that ran on the laptop as well.

### *EMA Subsystem*

EMA is a data collection technique in which one's behavior is repeatedly sampled in a natural environment [24]. In this study, participants were assessed on several individual behaviors and states via mobile questionnaires sent to their smartphone approximately every hour during waking hours. Each smartphone had an app developed by the members of our research team installed on it. The app acted as a mobile questionnaire platform (ie, participants answered the questionnaires within the app interface).

The two types of EMAs that the participants received are as follows: (1) time-triggered mobile questionnaires and (2) eating event–triggered mobile questionnaires.

A *time-triggered* mobile questionnaire was sent to the participants' smartphones every hour at the top of the hour (eg, 10 AM, 11 AM, 12 PM, etc; Figure 2A). The questionnaire included a brief validated positive affect and negative affect survey [44-47] (see Table 1 for the full list of questions).

Shortly after an eating event was detected for any given participant, an *eating event–triggered* mobile questionnaire was sent to the corresponding participant's smartphone asking to confirm whether they had just eaten **(**Figure 2B**)**. If they confirmed that they had just eaten, then following this first question, they were asked a battery of survey items including previously validated measures of hunger and satiety [48], mindful eating [49], positive and negative affect [44-47], and with whom they were eating, if anyone (see Table 1 for the full list of questions). If the participant had not finished eating, they were given the option to request more time before filling out the questionnaire.

If they responded to the first question indicating that they had not just eaten, then they were asked to report what activity they had just completed. They were then asked to respond to validated measures of positive and negative affect [44-47].

Figure 3 illustrates the full eating event–triggered EMA question logic. The full list of questions for the time-triggered and the eating event–triggered mobile questionnaires can be found in Table 1.

**Figure 2.** Examples of a time-triggered and eating event–triggered mobile questionnaire received on the phone of a participant. Figure 2A is an example of a time-triggered mobile questionnaire that the participants received on their phone during the study. It contains the first 4 questions of the questionnaire that measure negative affect. Figure 2B is an example of an eating event–triggered mobile questionnaire that the participants received on their phone during the study. It contains the first question of the questionnaire that measures whether the participant had just eaten or drank. EMA: ecological momentary assessment.

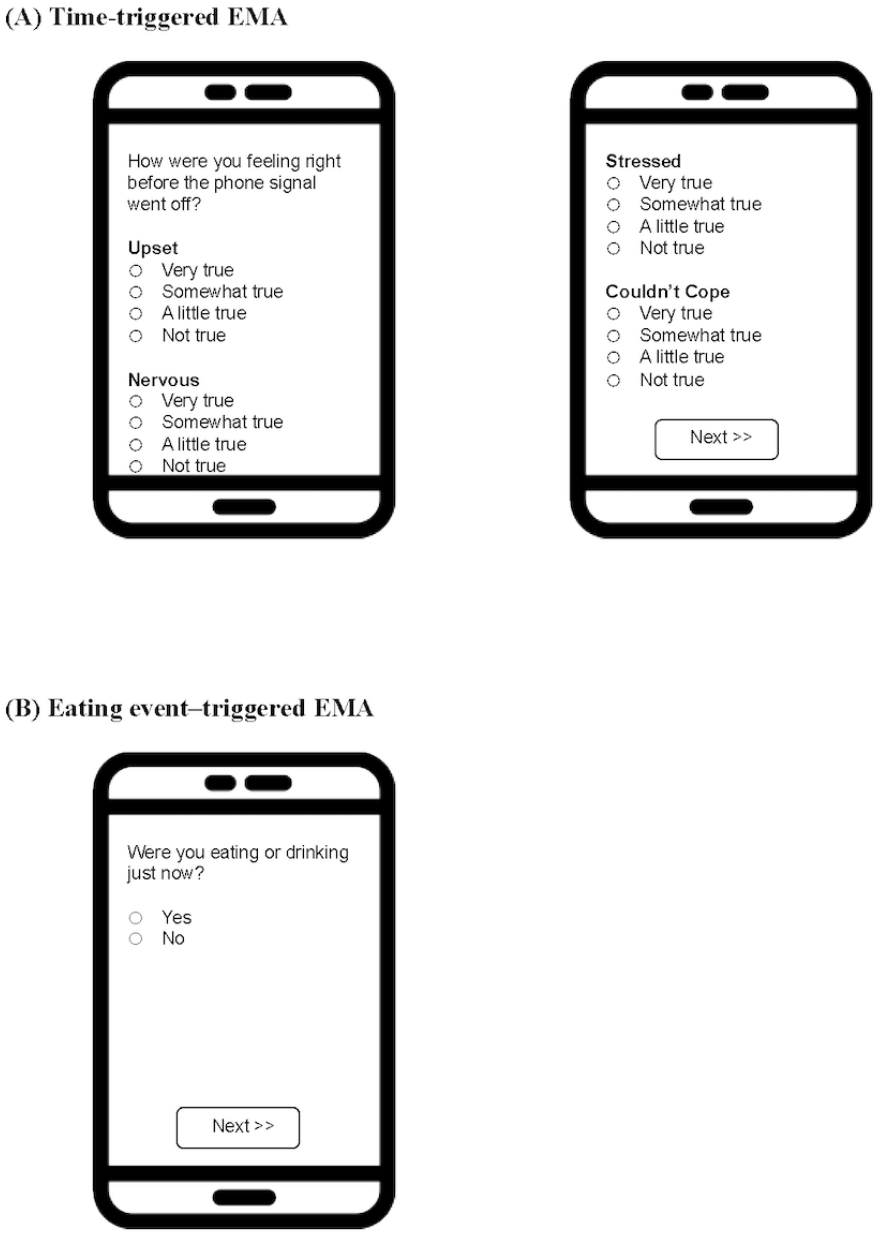
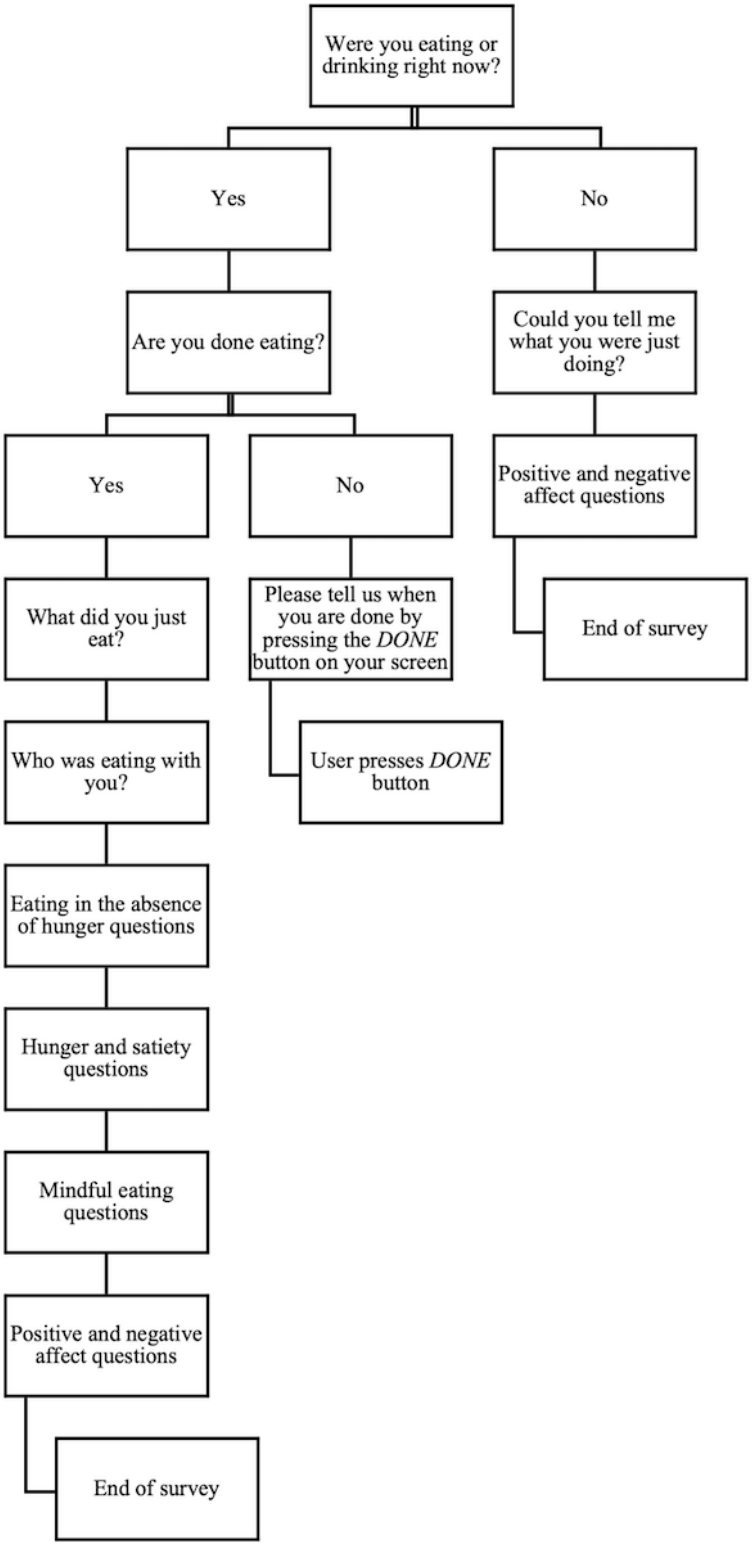


**Table 1.** Ecological momentary assessment (EMA) items.

| Variable (subscale) | Items | Response options | Format |
|---|---|---|---|
| **Time-triggered EMA** | | | |
| Positive and negative affect | How were you feeling right before the phone signal went off? (upset, nervous, stressed, could not cope, happy, great, cheerful, joyful) | • Not at all<br>• A little<br>• Some<br>• Very | Separate screen for each of the 8 items |
| **Eating event–triggered EMA** | | | |
| Eating confirmation | Were you eating or drinking just now? | • Yes<br>• No | —[a] |
| Eating type | What did you just eat? | • Meal<br>• Snack<br>• Drink only | — |
| Social context | Who was eating with you? (check all that apply) | • Nobody<br>• Spouse or partner<br>• Child(ren)<br>• Mother<br>• Father<br>• Sister(s)<br>• Brother(s)<br>• Grandparent<br>• Other family<br>• Friend(s)<br>• Other people | — |
| Eating in the absence of hunger—started eating | I *started* eating because (food looked, tasted, or smelled so good; others were eating; feeling sad or depressed; feeling bored; feeling angry or frustrated; feeling tired; feeling anxious or nervous; my family or parents wanted me to eat). | • Not at all<br>• A little<br>• Some<br>• Very | Separate screen for each of the 8 items |
| Eating in the absence of hunger—kept eating | I *kept* eating because (food looked, tasted, or smelled so good; others were eating; feeling sad or depressed; feeling bored; feeling angry or frustrated; feeling tired; feeling anxious or nervous; I wanted to finish the food on my plate). | • Not at all<br>• A little<br>• Some<br>• Very | Separate screen for each of the 8 items |
| Hunger level before eating | How hungry were you right before you ate? | • 0=Not at all hungry<br>• 100=Greatest imaginable hunger | Sliding scale 0 to 100 |
| Satiation level after eating | How full were you right after you ate? | • 0=Not at all full<br>• 100=Greatest imaginable fullness | Sliding scale 0 to 100 |
| Mindful eating | Before the beep, while I was eating (My thoughts were wandering while I ate; I was thinking about things I need to do while I ate; I ate so quickly that I did not taste anything). | • Very true<br>• Somewhat true<br>• A little true<br>• Not true | Separate screen for each of the 3 items |
| Positive and negative affect | How were you feeling right before the phone signal went off? (upset, nervous, stressed, could not cope, happy, great, cheerful, joyful) | • Not at all<br>• A little<br>• Some<br>• Very | Separate screen for each of the 8 items |

[a]No additional formatting notes.

**Figure 3.** Eating event–triggered ecological momentary assessment question logic.




### *Participation Windows*

Before a family's deployment started, all participants were individually asked about the time at which they normally woke up and the time at which they normally went to bed. The participants were limited to only 1 *personalized participation window* for the study. Therefore, they could not have different windows for Monday versus Tuesday and weekday versus weekend. If the times at which they woke up or went to bed varied extensively among days, then they were asked to provide a time frame that generally worked for all days. The purpose was to create *personalized participation windows* to account for variations in the daily routines and sleeping patterns of the participants. For the duration of the study, the participants only received EMAs during their personalized participation window. For example, if the window of a participant was from 6:30 AM to 11:00 PM, then they only received EMAs during that period.

### *Remote Monitoring Subsystem*

The monitoring subsystem was used to monitor the status of the M2FED system in real time [50]. The subsystem monitored several things, including the battery status and network connection of the smartwatches, smartphones, and base station; the processes running on the base station; the detected eating events; and whether participants responded to any given EMA sent to their smartphones. When the monitoring system detected an issue (eg, the base station was no longer connected to the router), an email was sent to the research team to alert them of the issue.

## Procedures

Following enrollment, 2 members of the research team visited the home of each recruited family 2 separate times.

### *Visit 1*

During the first home visit, the team went to the participants' home to obtain consent from all participating family members, take body measurements of the participants using a research-grade Tanita scale (Model TBF 300) and stadiometer, administer baseline surveys, and install the components of the cyberphysical system around the home (all *living spaces*, not including bedrooms or bathrooms).

The base station, Wi-Fi router, and Bluetooth beacons were placed in a discrete location in the home of the family, so they could run without interference for the duration of the study. Samsung smartphones and Sony smartwatches were provided to all participating family members for the duration of the study (all features except answering questionnaires were turned off). Each phone and watch was designated to a specific participant and labeled with their name so that they knew which devices were their own. The team instructed the family on how to properly wear, charge, and care for the smartwatches and how to answer an EMA on the smartphones. The family was instructed to wear the watch at all times when they were at home and to answer all EMA questionnaires they received when they were at home. They were also instructed to leave their designed phone and watch at home when they left home to prevent the devices from getting damaged or lost while outside of the home.

Upon leaving the visit, family members underwent approximately 14 consecutive days of (1) use of a smartphone to complete hourly time-triggered and eating event–triggered mobile questionnaires, up to once every hour during waking hours; and (2) eating event monitoring, in the form of a wrist-worn smartwatch during waking hours.

### *Visit 2*

At the final home visit, approximately 2 weeks following the first home visit, the research team terminated data collection, and all equipment was uninstalled and removed from the homes. Each participant received US $100 in a Visa gift card format as compensation for the 2-week study.

## Measures

### *Eating Events*

During the 2-week assessment period, participants were asked to wear their dedicated smartwatch on their dominant wrist at all times while they were home during waking hours. Automatic eating event detection software on the smartwatches developed by our research team [43] collected the timestamps (approximate start and end times in the format mm/dd/yyyy, hh:mm:ss) for all detected eating events that occurred while the watch was worn. After an eating event was detected, participants received a brief mobile questionnaire on their study phones to confirm whether the detected eating event was a true event. The first question on the questionnaire was "Were you eating or drinking just now?" If the participant responded "No," they were asked to report what they were doing. Options included *using my phone, smoking, fixing my hair, putting on sunscreen or lotion, or other* with an open text field. If the participant responded "Yes," they were asked to report on a range of momentary measures, such as hunger level before the eating event and with whom they were eating. The full list of questions for the time-triggered and the eating event–triggered mobile questionnaires can be found in Table 1.

### *EMA Questionnaires*

Timestamps (format: mm/dd/yyyy, hh:mm:ss) when the hourly time-triggered and eating event–triggered mobile questionnaires were sent to and received by the smartphones of participants were obtained from the monitoring system. In addition, the responses of the participants to the questionnaires were obtained from the monitoring system.

### *Timing*

*Time of day* at which and *day of week* on which an eating event occurred was calculated using the timestamp of the detected eating events. The time of day at which the eating event occurred was stored in hh:mm:ss format. The *lubridate* R package [51] was used to convert the date on which the eating event occurred (format: mm/dd/yyyy) to the day of corresponding week (Monday, Tuesday, etc), which was then converted to weekday (Monday, Tuesday, etc) and weekend (Saturday or Sunday).

### *Anthropometrics*

During home visit 1, height (cm), weight (kg), and body fat percentage (%) were measured in all participants in a private section of the home, using a portable stadiometer and a research-grade Tanita scale (model TBF 300).

### *Demographics*

During home visit 1, participants were asked to provide basic demographic information via a paper-based questionnaire, including their current age (years), gender (female or male), race (Hispanic or Latino, Asian or Pacific Islander, White, Black or African American, American Indian or Native American, Mixed, or other), Hispanic or Latino ethnicity (Yes, No, or Do not know), and family role (mother, father, child, grandparent, aunt, uncle, and others).

## Analytic Approach

### *Data Processing*

A limitation of the EMA sampling protocol of the M2FED study was that the study phones of participants (which were instructed to be kept at home at all times) received hourly, time-triggered surveys regardless of whether the participants themselves were

at home or not (at school or work, running errands, etc). This means that the time frame in which any given participant was at home and participating in the study was not necessarily continuous. Although we do not possess the ground truth for presence of the participants at home (eg, no cameras and no self-report diaries), our research team generated a *participation algorithm* using the EMA system, proximity sensors, and accelerometer in the watch to identify time intervals in which we were confident that the participants were both at home and actively participating in the study (ie, answering EMAs or wearing the smartwatch; Figure 4).

If participants had answered an EMA at time t, then we assigned their status as *participating* for the 30-minute interval surrounding time t (ie, from t −15 to t +15 minutes). For times outside the EMA interaction windows, we used data from the sensors (smartwatch accelerometer and Bluetooth beacons) to determine the status of the participants. For every minute, if the accelerometer data of the smartwatch was both available (ie, not missing for that minute) and indicated movement (ie, the frequency and instantaneous changes of the sensor signal was above a threshold, representing change in the signal because of movement) and beacon data were available, then they were classified as *participating* for that 1-minute interval. Contiguous minute intervals with *participating* status were merged to acquire larger time intervals. For each participant, these *participation* time intervals were calculated, and the union of all intervals (Figure 5) was used as the valid time interval in the analyses.

**Figure 4.** Decision tree to determine when study participants were participating at home. EMA: ecological momentary assessment.

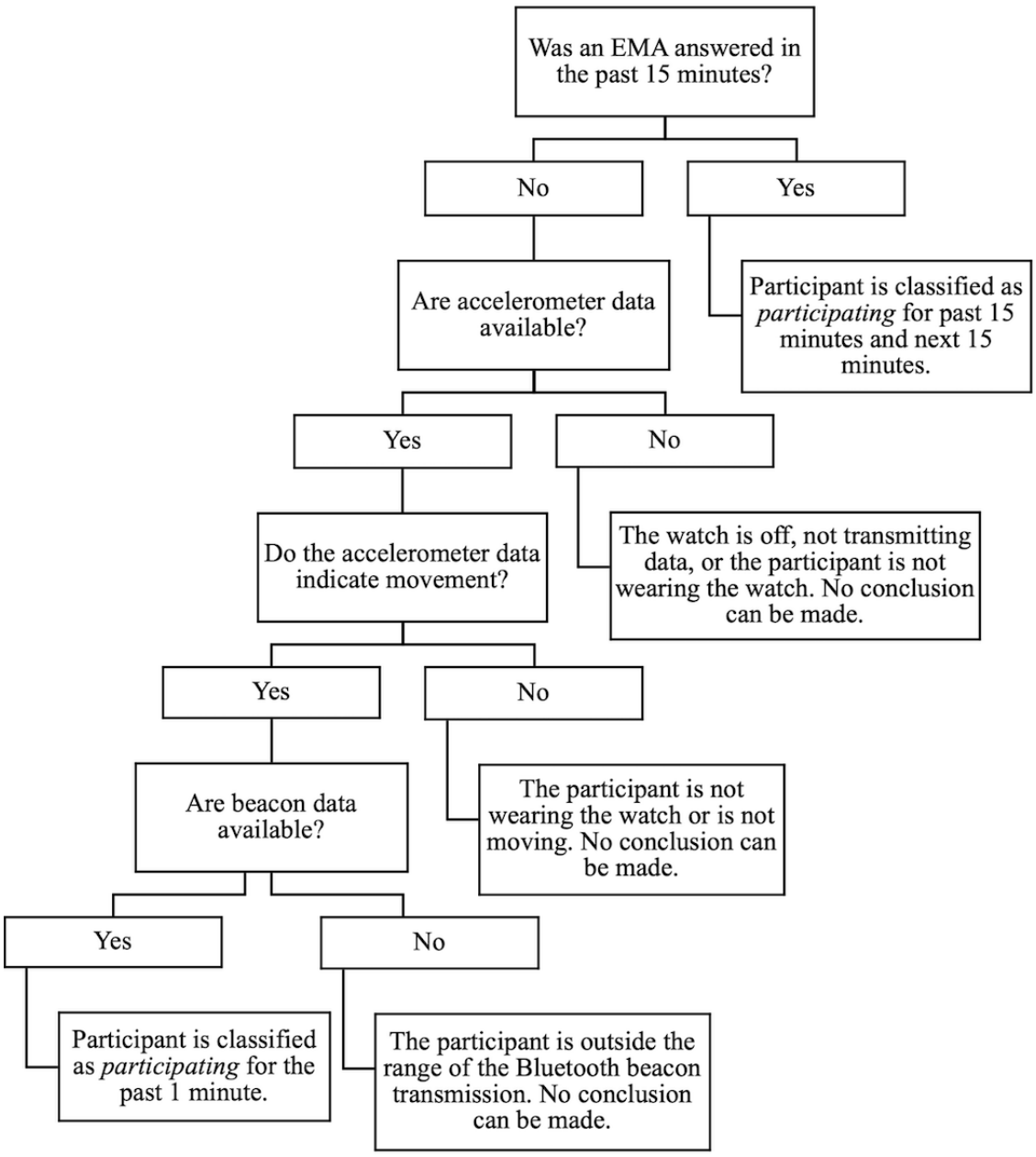
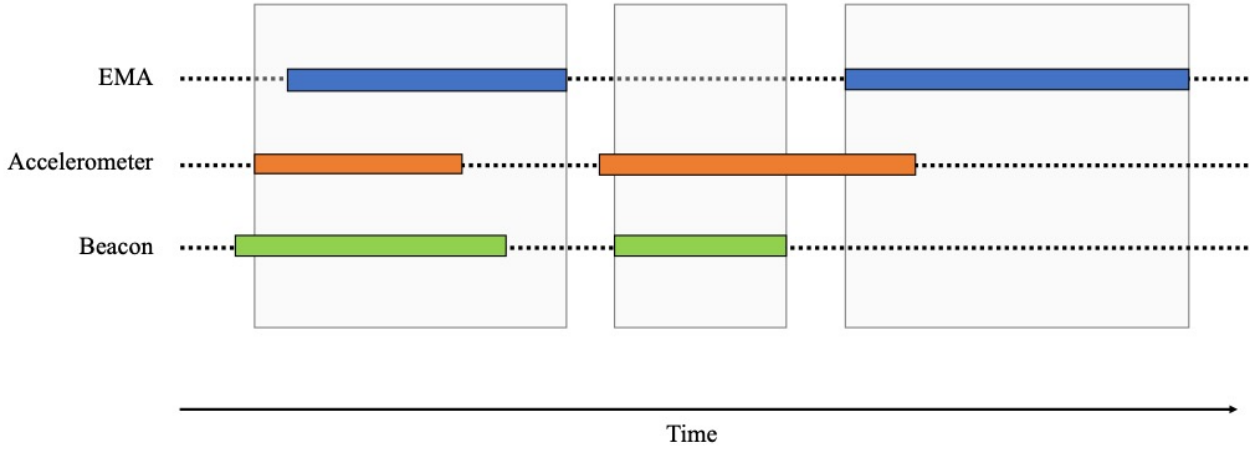


**Figure 5.** Example of participation time intervals for a participant. In this example, the shaded gray regions indicate the valid participation time intervals for this participant. In the first interval, we see that the participant answered an ecological momentary assessment (EMA), and there were available data from the accelerometer and beacon. In the second interval, the participant did not answer an EMA, but there were available data from the accelerometer and beacon. In the third interval, the participant answered an EMA and there were some available data from the accelerometer.




### *Data Analyses*

#### *Individual- and Family-Level Characteristics*

The mean and SD or the count and proportion of the analytic sample's age, BMI, gender, race, and ethnicity were calculated and reported by family role (child or parent). At the family level, the count and proportion of the type of household of the family (1- or 2-parent household), number of children living at home, and average length of family deployment were reported.

#### *EMA Characteristics*

The mean and SD of EMAs received per family, received per person, and received per person per day were calculated after applying the participation algorithm to the EMA data. The frequency distribution of EMAs by family role and time of day was calculated.

#### *Primary Analyses*

To test study aim 1A, EMA compliance was calculated as follows (*i* can be values from 1 to *n*, where *n* represents the number of participants in the study):

> Overall compliance to EMAs for participant$_i$ = total number of EMAs answered by participant$_i$ / total number of EMAs received at home by participant$_i$ **(1)**

> Compliance to *time-triggered* EMAs for participant$_i$ = total number of time-triggered EMAs answered / total number of time-triggered EMAs received at home by participant$_i$ **(2)**

> Compliance to *eating event–triggered* EMAs for participant$_i$ = total number of eating event–triggered EMAs answered / total number of eating event–triggered EMAs received at home by participant$_i$ **(3)**

Means and SDs of overall compliance to EMAs, compliance to time-triggered EMAs, and compliance to eating event–triggered EMAs were also calculated across all participants.

To test study aim 1B, the unit of analysis was every EMA that was sent to and received by the smartphones of the participants throughout the span of the 2-week data collection period. Compliance (dependent variable) was calculated as *1* if the questionnaire was answered and as *0* if the survey was not answered. A logistic regression model was fitted with the following independent variables: type of EMA (time-triggered and eating event–triggered), time of day (morning, defined as midnight to 11:59:59 AM; afternoon, defined as noon to 16:59:59 PM; and evening, defined as 17:00:00 PM to 23:59:59 PM), day of week (weekday, defined as Monday through Friday; and weekend, defined as Saturday and Sunday), gender (male or female), family role (parent, child, or other), and social factors (whether another participating family member *j* had answered a survey that had been received within 15 minutes of the focal person *i*'s questionnaire).

To test study aim 2A, we evaluated the performance of the smartwatch by computing the following metrics for all eating events automatically detected during deployments:

True positives = cases in which an eating event actually *occurred*, and that eating event was *correctly* detected by the smartwatch algorithm

False positives = cases in which an eating event actually *did not* occur, but an eating event was *erroneously* detected by the smartwatch algorithm.

> Precision = true positives / (true positives + false positives) **(4)**

To test study aim 2B, nonparametric methods were used to determine whether there were differences in the detection of eating events by participant age, gender, family role, and height. The metric we used to compare across demographic groups was the following:

> Proportion of correctly detected eating events for participant$_i$ = true positives for participant$_i$ / total number of detected eating events for participant$_i$ **(5)**

If any participant had received fewer than 3 eating event–triggered EMAs, their data were excluded from this analysis.

For categorical variables with 2 groups (ie, gender), the appropriate assumptions were tested, and then the Mann–Whitney *U* test was used to test for equality of central tendency of the 2 distributions; for categorical variables with 3 or more categories (ie, family role), the Kruskal-Wallis test was used. Finally, for continuous variables (ie, height [cm] and age [years]), the appropriate assumptions were tested, and Spearman rank correlation was used to measure the strength and direction of the relationship between the continuous variable and the proportion of correctly detected events.

### *Missing Data*

There were no missing anthropometric or demographic data. Similarly, there were no missing data on detected eating events and corresponding variables, including time of eating event and day of eating event; however, there were missing data for time-triggered and eating event–triggered EMAs.

#### Missingness Attributed to Technical Issues

Preliminary analyses indicated that not all EMAs that were *sent* to the study phones of the participants by the M2FED system were *received* by the phone. The M2FED system ran independently on the base station regardless of the network connection, and therefore *sent* EMAs regardless of network connection. However, a network connection was needed for the phone to successfully *receive* the EMA.

Although we do not have data that explain why this happened at every instance, we know from in-the-field troubleshooting and from accounts given by participants that at least a portion of the nonreceived EMAs resulted from (1) network connection issues at home (ie, the router was not working and the EMAs could not be received on the phone) and (2) EMA app failure (ie, the EMA app on the phone failed to work properly).

For these analyses, we removed any EMAs that were sent by the system but were not received by the phone.

#### Missingness Attributed to Participant Nonresponse or Partial Response

The different types of missing data that we encountered were because of participant nonresponse (ie, participants did not respond to any EMA questions) or partial responses (ie, participants did not respond to all EMA questions).

For aim 1 analyses, if participants did not respond to *any* questions on a given mobile questionnaire, then this EMA was labeled as *received but not answered*. If participants did not respond to *all* questions, then this EMA was labeled as *received and partially answered*. These EMA observations were kept in the data set to calculate EMA compliance.

For aim 2 analyses, if participants did not respond to at least the *first* question on a given eating event–triggered EMA ("Were you eating or drinking just now?"), then this EMA observation was removed from the data set.

*Statistical software* R (version 4.0.2) was used to perform these analyses.

## Results

### Individual- and Family-Level Characteristics

A total of 74 participants from 20 families were enrolled in the M2FED study. In all, 18% (13/74) of participants dropped out of the study or were removed from the data set if their participation (as determined by the participation algorithm) was 0% (ie, they did not answer any EMAs and never wore the smartwatch; Figure 6).

In addition, the data from 4% (3/74) nonparent adult participants made up approximately 1.44% (61/4232) of the EMAs received, so these participants were removed from the analytic sample as well. The remaining 78% (58/74) of participants included in the analytic sample did not significantly differ from the enrolled sample (N=74) by age, gender, or parent role ($P$>.05; Table 2).

Of the 58 participants, 43% (n=25) were parents and 57% (n=33) were children. On average, children were aged 15.12 years (SD 3.97 years) and parents were aged 43.72 years (SD 6.71 years). There were 39% (13/33) female children and 68% (17/25) female parents. In all, 61% (20/33) of children and 68% (16/25) of parents identified as Hispanic or Latino (Table 3).

Of the 20 enrolled families, most (17/20, 85%) were 2-parent households, 15% (3/20) of the families had 1 child living at home, 75% (15/20) of the families had 2 children, 5% (1/20) of the families had 3 children, and 5% (1/20) of the families had 4 children (Table 4). On average, family deployments lasted 14.90 days (SD 3.13 days).

**Figure 6.** Flow of participants in the Monitoring and Modeling Family Eating Dynamics study. Participants may not have received an eating event–triggered ecological momentary assessment (EMA) as no eating event was detected by the system or technical issues prevented the EMA from sending.

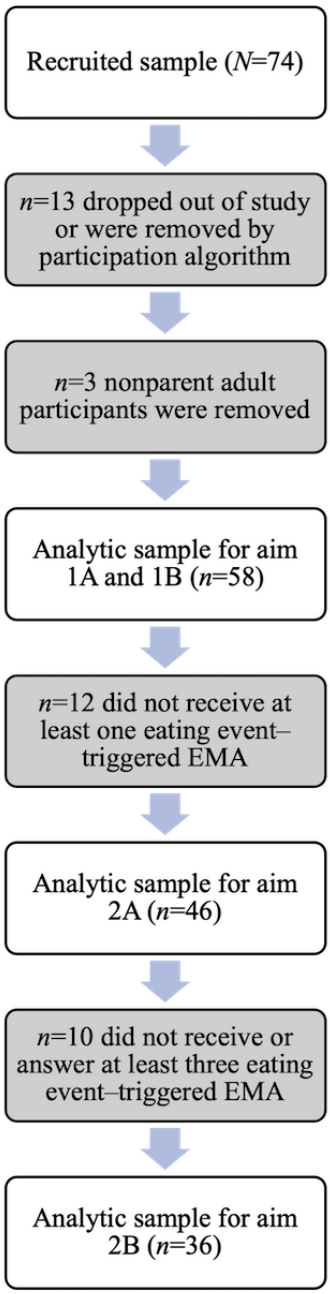



**Table 2.** Comparison of recruited sample and analytic samples.

| Characteristics | Values | | | | | | |
|---|---|---|---|---|---|---|---|
| | Recruited sample (N=74) | Analytic sample for aim 1A and 1B (n=58) | *P* value[a] | Analytic sample for aim 2A (n=46) | *P* value[a] | Analytic sample for aim 2B (n=36) | *P* value[a] |
| Age (years), mean (SD) | 28.91 (15.79) | 27.45 (15.23) | .59 | 28.76 (15.51) | .96 | 26.67 (14.83) | .47 |
| Sex (female), n (%) | 37 (50) | 30 (52) | .78 | 24 (52) | .81 | 20 (56) | .49 |
| Parent (yes), n (%) | 32 (43) | 25 (43) | .99 | 21 (46) | .77 | 15 (42) | .97 |

[a]*P* values were calculated by comparing the analytic sample to the recruited sample. Welch 2 independent sample 2-tailed *t* test was used for continuous variables (ie, age), and Pearson chi-square test was used for categorical variables (ie, sex and parent).

**Table 3.** Individual-level characteristics of the Monitoring and Modeling Family Eating Dynamics analytic sample (N=58), by family member role.

| Characteristics | Child (n=33)[a] | Parent (n=25)[a] |
|---|---|---|
| Age (years), mean (SD) | 15.12 (3.97) | 43.72 (6.71) |
| Sex (female), n (%) | 13 (39) | 17 (68) |
| **Race and ethnicity, n (%)** | | |
| Asian or Pacific Islander | 1 (3) | 1 (4) |
| Black or African American | 2 (6) | 1 (4) |
| Hispanic or Latino | 20 (61) | 16 (68) |
| White | 4 (12) | 4 (16) |
| Mixed | 6 (18) | 1 (4) |
| Other | 0 (0) | 1 (4) |
| BMI[b] percentile (n=53), mean (SD) | 22.36 (4.66) | 32.90 (7.38) |

[a]The percentages presented are column percentages.

[b]BMI: body mass index.

**Table 4.** Family-level and deployment-level characteristics of the Monitoring and Modeling Family Eating Dynamics study families (N=20).

| Characteristics | Values |
|---|---|
| **Number of parents living at home, n (%)[a]** | |
| 1-parent household | 3 (15) |
| 2-parent household | 17 (85) |
| **Number of children living in the home, n (%)[a]** | |
| 1 child | 3 (15) |
| 2 children | 15 (75) |
| 3 children | 1 (5) |
| 4 children | 1 (5) |
| Deployment length (days), mean (SD) | 14.90 (3.13) |

[a]The percentages presented are column percentages.

## EMA Characteristics

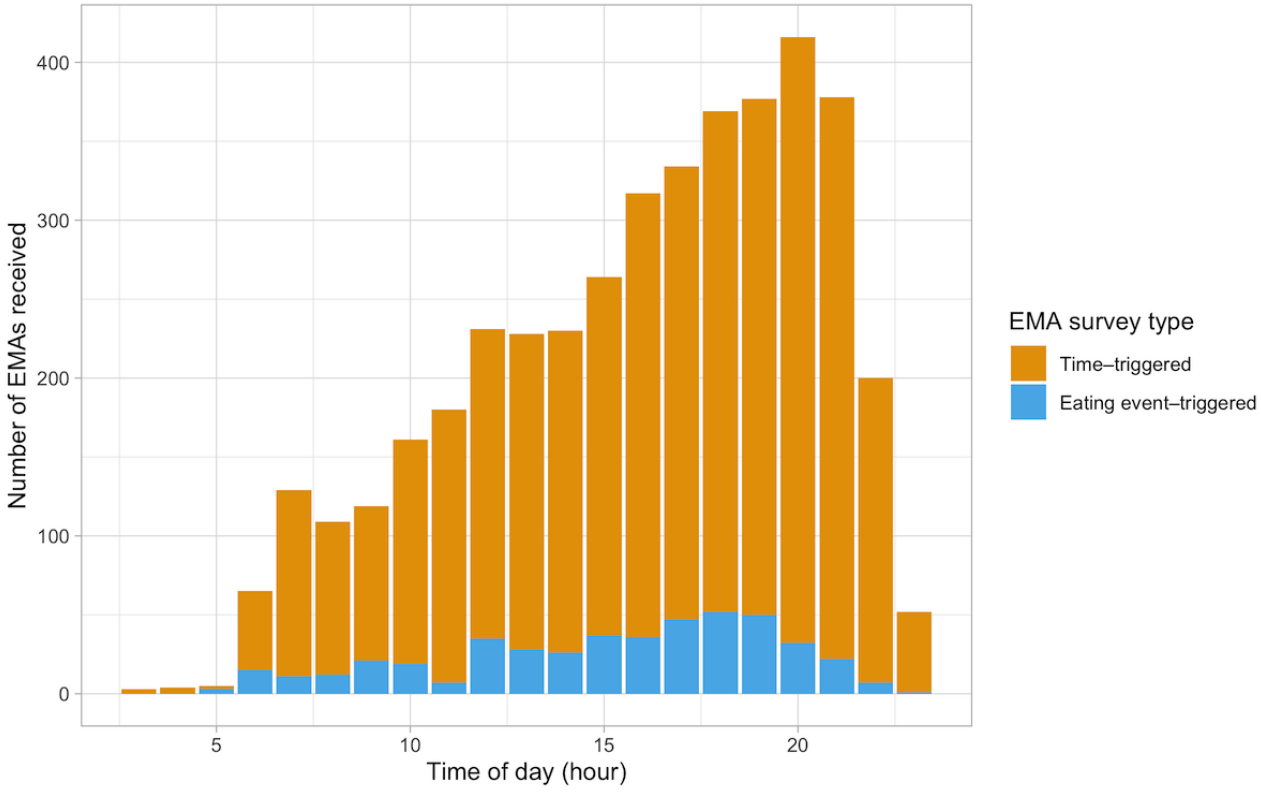
In total, 15,010 EMAs (14,348/15,010, 95.59% time-triggered and 662/15,010, 4.41% eating event–triggered) were sent by the M2FED system and received by study phones of the participants. After filtering the data through the participation algorithm, 27.78% (4171/15,010) EMAs remained in the data set: 88.95% (3710/4171) of which were time-triggered and 11.05% (461/4171) were eating event–triggered (Table 5).

On average, families received 209.0 EMAs (SD 89.4; range 86-391), and individuals received 71.9 EMAs (SD 34.3; range 8-176) each. Participants received, on average, 64.0 time-triggered EMAs (SD 31.3; range 8-147) and 8.0 eating event–triggered EMAs (SD 8.9; range 0-40) across the deployment. The daily average number of EMAs received per person was 5.2 (SD 2.7; range 0.6-11.7) for all EMAs, 4.7 (SD 2.4; range 0.3-10.2) for time-triggered EMAs, and 0.6 (SD 0.6; range 0-2.7) for eating event–triggered EMAs (Table 5). Of the 4171 total EMAs, 18.58% (775/4171) were received in the morning, 30.46% (1270/4171) in the afternoon, and 50.97% (2126/4171) in the evening. Of the 461 eating event–triggered EMAs, most, 45.8% (211/461), were sent in the evening (Figure 7). Children received 57.52% (2399/4171), fathers received 10.72% (447/4171), and mothers received 31.77% (1325/4171) of the total EMAs. Of the 461 eating event–triggered EMAs, 49.9% (n=230) were received by children, 7.4% (n=34) by fathers, and 42.7% (n=197) by mothers.

**Table 5.** Ecological momentary assessment (EMA) summary statistics after applying participation algorithm, by prompt type.

| Type of EMA | Total EMAs received, N | EMAs received per family, mean (SD; range) | EMAs received per person, mean (SD; range) | EMAs received per person per day, mean (SD; range) |
|---|---|---|---|---|
| All EMA | 4171 | 209.0 (89.4; 86-391) | 71.9 (34.3; 8-176) | 5.2 (2.7; 0.6-11.7) |
| Time-triggered EMA | 3710 | 186.0 (84.3; 77-356) | 64.0 (31.3; 8-147) | 4.7 (2.4; 0.3-10.2) |
| Eating event–triggered EMA | 461 | 23.0 (17.2; 3-69) | 8.0 (8.9; 0-40) | 0.6 (0.6; 0-2.7) |

**Figure 7.** Distribution of ecological momentary assessments (EMAs) received across the time of day (hour), by EMA survey type.



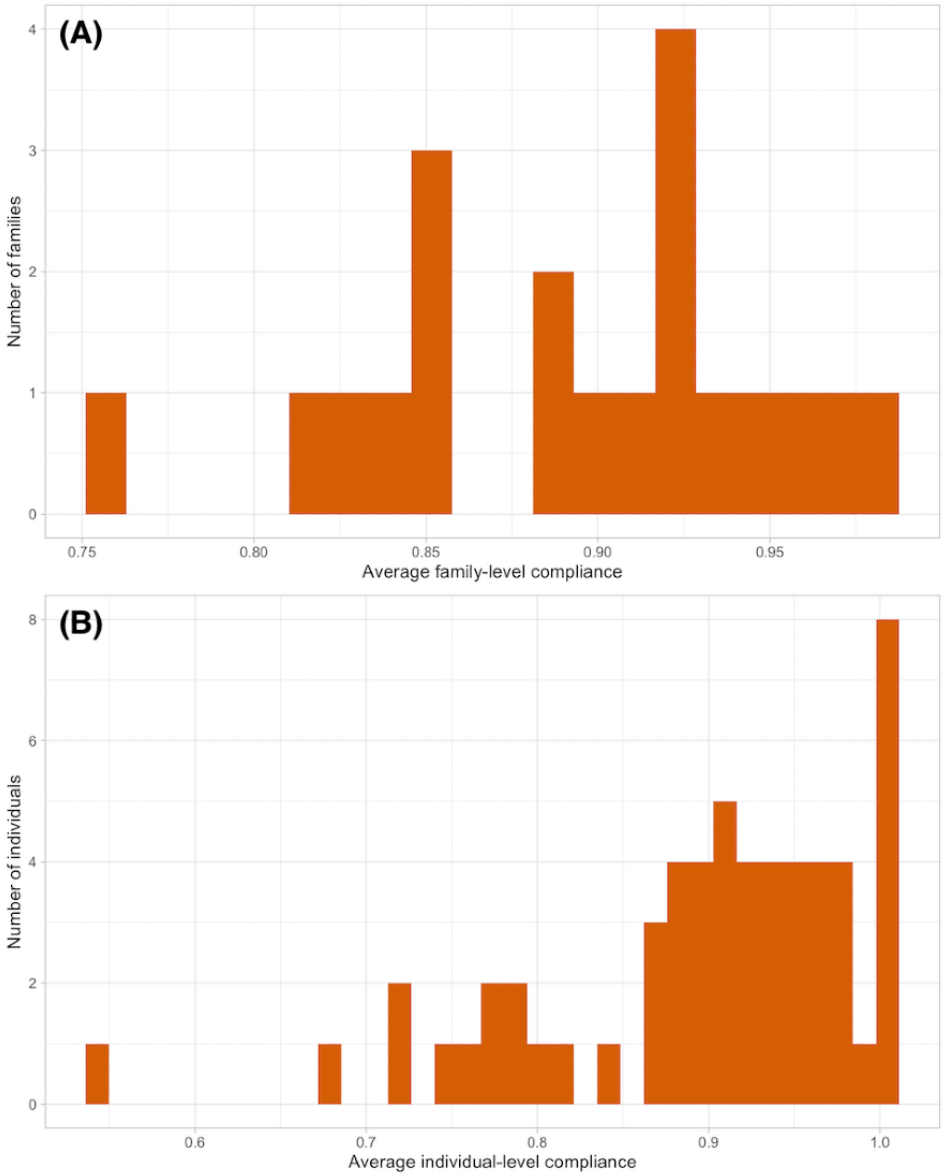

## Participant Compliance

The overall compliance rate across the 20 deployments was 89.26% (3723/4171) for all EMAs, 89.7% (3328/3710) for time-triggered EMAs, and 85.7% (395/461) for eating event–triggered EMAs (Table 6). The average family-level compliance was 89.4% (SD 5.74%; range 75.7%-98.1%) for all EMAs, 89.8% (SD 5.84%; range 75.8%-98.7%) for time-triggered EMAs, and 85.9% (SD 14.3%; range 55.6%-100%) for eating event–triggered EMAs. At the individual-level, the average compliance for all EMAs was 89.6% (SD 9.5%; range 53.8%-100%), for time-triggered EMAs was 89.5% (SD 10.1%; range 50%-100%), and for eating event–triggered EMAs was 88% (SD 17.5%; range 28.6%-100%). The distributions of individual- (Figure 8A) and family-level compliance (Figure 8B) are shown in Figure 8.

**Table 6.** Ecological momentary assessment (EMA) compliance rates after applying participation algorithm, by prompt type.

| Type of EMA | Total EMAs received, N | Total EMAs answered (compliance), n (%) | Family-level compliance (%), mean (SD; range) | Individual-level compliance (%), mean (SD; range) |
|---|---|---|---|---|
| All EMA | 4171 | 3723 (89.3) | 89.4 (5.74; 75.7-98.1) | 89.6 (9.5; 53.8-100) |
| Time-triggered EMA | 3710 | 3328 (89.7) | 89.8 (5.8; 75.8-98.7) | 89.5 (10.1; 50-100) |
| Eating event–triggered EMA | 461 | 395 (85.7) | 85.9 (14.3; 55.6-100) | 88.0 (17.5; 28.6-100) |

**Figure 8.** Distribution of (A) family-level and (B) individual-level compliance.




## Predictors of Compliance

Three separate logistic regression models were fitted with the following data sets: (1) all EMAs, (2) time-triggered EMAs, and (3) eating event–triggered EMAs.

Results from the first model indicate that time of day and whether other family members had also answered an EMA were significant predictors of compliance to all EMAs (Table 7). Participants were 37% less likely (odds ratio [OR] 0.63, 95% CI 0.46-0.86) to respond to an EMA in the afternoon and 39% less likely (OR 0.61, 95% CI 0.45-0.81) to respond to an EMA in the evening compared with the morning (reference group). Participants were 91% more likely (OR 1.91, CI 1.56-2.34) to respond to an EMA if another family member had responded to an EMA in the surrounding 30-minute time interval.

The results from the second model indicate that time of day and whether other family members had also answered an EMA were significant predictors of compliance to time-triggered EMAs (Table 7). Participants were 40% less likely (OR 0.60, 95% CI 0.42-0.85) to respond to a time-triggered EMA in the afternoon and 47% less likely (OR 0.53, 95% CI 0.38-0.74) to respond to a time-triggered EMA in the evening than in the morning (reference group). Participants were approximately 2 times as likely (OR 2.07, 95% CI 1.66-2.58) to respond to a time-triggered EMA if another family member had responded to any EMA in the surrounding 30-minute time interval.

Results from the third model indicate that weekend status and deployment day were significant predictors of compliance to eating event–triggered EMAs (Table 7). Participants were 2.4 times as likely (OR 2.40, 95% CI 1.25-4.91) to respond to an eating event–triggered EMA on the weekend, than on a weekday. Participants were 8% less likely (OR 0.92, 95% CI 0.86-0.97) to respond to an eating event–triggered EMA for every 1-day increase in deployment day.

**Table 7.** Logistic regression model results, examining predictors of compliance[a].

| Characteristics | Model 1: all EMAs[b] | | Model 2: time-triggered EMAs | | Model 3: eating event–triggered EMAs | |
|---|---|---|---|---|---|---|
| | β (SE) | OR[c] (95% CI) | β (SE) | OR (95% CI) | β (SE) | OR (95% CI) |
| Intercept | *2.17*[d,e] *(0.27)* | 8.75 (5.20-14.82) | *2.22*[e] *(0.29)* | 9.22 (5.24-16.36) | *2.41*[f] *(0.74)* | 11.15 (2.65-48.64) |
| Age (years) | .00 (0.01) | 1.00 (0.98-1.03) | .00 (0.01) | 1.00 (0.98-1.02) | .02 (0.03) | 1.02 (0.96-1.08) |
| Afternoon | *−.47*[f] *(0.16)* | 0.63 (0.46-0.86) | *−.51*[f] *(0.18)* | 0.60 (0.42-0.85) | −.35 (0.38) | 0.71 (0.33-1.46) |
| Evening | *−.50*[f] *(0.15)* | 0.61 (0.45-0.81) | *−.63*[e] *(0.17)* | 0.53 (0.38-0.74) | .28 (0.38) | 1.32 (0.62-2.75) |
| Weekend, yes | .06 (0.11) | 1.06 (0.86-1.31) | −.06 (0.12) | 0.95 (0.75-1.19) | *.87*[g] *(0.35)* | 2.40 (1.25-4.91) |
| Deployment day | −.02 (0.01) | 0.98 (0.96-1.01) | −.01 (0.01) | 0.99 (0.97-1.01) | *−.09*[f] *(0.03)* | 0.92 (0.86-0.97) |
| Female, yes | .19 (0.15) | 1.21 (0.90-1.65) | .31 (0.17) | 1.37 (0.98-1.92) | −.65 (0.43) | 0.52 (0.22-1.22) |
| Mother | −.01 (0.34) | 0.99 (0.51-1.93) | .06 (0.36) | 1.06 (0.53-2.16) | −.65 (1.07) | 0.52 (0.06-4.56) |
| Father | −.42 (0.35) | 0.66 (0.33-1.30) | −.37 (0.38) | 0.69 (0.33-1.47) | −.64 (0.93) | 0.53 (0.08-3.26) |
| Others answered, yes | *.65*[e] *(0.10)* | 1.91 (1.56-2.34) | *.73*[e] *(0.11)* | 2.07 (1.66-2.58) | −.02 (0.30) | 0.99 (0.54-1.76) |

[a]Akaike information criteria is 2805.16, 2417.32, and 375.57 for models 1-3, respectively. Bayesian information criteria is 2868.52, 2479.50, and 416.91 for models 1-3, respectively.

[b]EMA: ecological momentary assessment.

[c]OR: odds ratio.

[d]Values indicate significant estimates.

[e]*P*<.001.

[f]*P*<.01.

[g]*P*<.05.

### Smartwatch Algorithm Evaluation

At least one eating event was automatically detected during the deployment for 46 participants. This subsample (ie, the analytic sample for aim 2A) did not significantly differ from the enrolled sample (N=74) by age, gender, or parent role (*P*>.05; Table 2).

A total of 461 eating events were automatically detected using the smartwatch algorithm across these 46 participants. Participants responded to 85.7% (395/461) of the corresponding eating event–triggered EMAs. Participants confirmed that 76.5% (302/395) of the detected events were true eating events (ie, true positives) and 23.5% (93/395) were not true eating events (ie, false positives). For approximately one-third of these false positives, participants reported that they were using their phones at the time. The calculated precision measure, that is, the number of true positives divided by the sum of true positives and false negatives, was 0.77.

### Differences in Eating Event Detection

At least three eating event–triggered EMAs were received by 36 participants. This subsample (ie, the analytic sample for aim 2B) did not significantly differ from the enrolled sample (N=74) by age, gender, or parent role (*P*>.05; Table 2). For this subsample, the average individual-level proportion of correctly detected eating events (true positives / total number of detected eating events) was 78.5% (SD 19%; range 30%-100%). In all, 72% (26/36) of the analytic sample had at least one falsely detected eating event (false positive).

Neither age (years) nor height (inches) was significantly correlated with the proportion of correctly detected eating events ($r_s$=0.24, *P*=.17 and $r_s$=−0.12, *P*=.52, respectively). The average individual-level proportion of correctly detected eating events for women was 82.1% (SD 20.4%; range 30%-100%) and was 74% (SD 16.6%; range 50%-100%) for men. The difference between the 2 groups was not significant (W=112; *P*=.13). The average individual-level proportion of correctly detected eating events for children was 74.3% (SD 19.3%; range 30%-100%), for fathers was 76.1% (SD 21.5%; range 58.3%-100%), and for mothers was 86.5% (SD 16.8%; range 54.5%-100%). The differences among these 3 groups were not significant (Kruskal–Wallis $\chi^2_2$=2.998; *P*=.22).

## Discussion

The M2FED study sought a dramatically different mobile health (mHealth) approach to obesity prevention and intervention by not focusing directly on diet and activity, but rather on family eating dynamics. An in-home sensor system was developed and deployed to monitor family eating dynamics in real time and context.

### Evaluating EMA Compliance

After applying our customized participation algorithm, we found that both individual- and family-level compliance rates to the EMA protocols of the study were relatively high (both greater than the recommended 80%) [24]. Compliance was significantly higher in the mornings overall and higher on the weekends for eating event–triggered EMAs, which supported the informal

feedback we received from participants that they were more likely to *participate* (ie, respond to EMAs and wear the smartwatch) when they did not need to go to work or school (typically the weekend days). We also saw that overall compliance decreased as the 2-week study went on, most likely attributable to participant fatigue.

One particularly interesting finding was that participants were significantly more likely to answer an EMA if another family member had answered an EMA in a similar time frame. A similar finding was reported by Dzubur et al [41], in which mother-child dyads were more likely to comply with prompts when they were together. Although the overarching aims of the M2FED study were to measure the social influence of family members on eating behavior, this finding also indicates that social influence came into play in other parts of the study as well. Drawing from the social psychology field, several social mechanisms could partially explain these findings. For instance, an expectation could have been set early on in particular families to answer the EMA prompts, thus establishing a social norm for EMA compliance [52,53]. Similarly, some individuals may have been inclined to answer EMA prompts to conform to the behavior of other family members around the same time [52,53], especially considering that family members received their time-triggered EMAs at approximately the same time as each other.

Studies have used EMA to measure various dietary outcomes, including frequency of food intake, intake of specific types of foods (eg, low glycemic index foods), and energy intake [25]. It has been suggested in a recent systematic review of mobile ecological momentary diet assessment methods that EMA has the potential to be a novel dietary assessment method, both on its own and as a supplement to other mHealth technologies [25]. The use of EMA to assess dietary intake and eating behavior provides some key advantages, namely, the reduction of participant burden and recall bias and the maximization of ecological validity [25]. Taken together with the findings from Dzubur et al [41] and Schembre et al [25], our findings suggest that EMA can be used to sufficiently supplement automatic dietary assessment (ADA) approaches and may be a particularly useful approach for leveraging social relations and maintaining compliance in dyad- and group-based EMA studies.

## Evaluating ADA

Various technologies have been used to passively measure eating activity in naturalistic settings over long periods with minimal user interaction. One of the most popular technologies for assessing eating activity in the field is the wrist-worn smartwatch or accelerometer [23,27]. The performance of automatic, wearable-based, in-field eating detection approaches to date has been reviewed by Bell et al [27]. The smartwatch used in the M2FED study performed on par with other in-field devices, although comparability is difficult owing to the wide and varying metrics used by other papers [27]. Although some wearable devices included in this review performed very well, the duration of the free-living deployment was 1 day (approximately 24 hours) or shorter for more than half of the studies, and another one-third were 1 week in length or shorter [27].

Overall, 3 studies had durations that lasted at least two weeks or longer [34,54,55], 66% (n=2) of which had sample sizes of only 1 participant each. Therefore, the M2FED study is one of the first studies to extensively test the feasibility of deploying an ADA approach for a considerable amount of time (2 weeks) and with a relatively large same size (>50 participants). Part of this success stems from the combined use of mobile devices (for EMA) and smartwatches, which were selected for the M2FED study to maximize long-term usability. Although other technologies have been able to perform better in the field, the usability of these technologies (electromyography electrodes, ear and neck sensors, wearable video cameras, etc) may be lower compared with wrist-worn devices because of the inconvenient location of sensor placement, the potential to interfere with the behavior of participants in real life [56], and the potential intrusiveness or discomfort caused by the sensor [57].

This study also demonstrates that EMA is a feasible tool for collecting ground-truth eating activity and thus evaluating the performance of wearable sensors in the field. Only 2 studies [34,35] included in the review by Bell et al [27] used a novel method for obtaining ground-truth eating activity in the wild, similar to the way EMA was used in the M2FED study. In a study by Ye et al [34], when an eating gesture was automatically detected via a wrist-worn sensor, participants were sent a short message on their smartwatch to confirm or reject in real time whether they were eating. Similarly, in a study by Gomes and Sousa [35], when drinking activity was detected via a wearable sensor, participants were sent an alert on their smartphone and could then confirm or reject whether they were drinking via EMA. Although EMA and similar self-report methods have their own limitations [23,58], they offer the ability to capture and validate ground-truth eating activity near the time of eating, thus improving research scalability and participant acceptability [25].

Another key feature of the M2FED study was the ability to capture intrapersonal (individual) and interpersonal (social) contexts with our combined event- and signal-contingent protocols. A systematic review noted that <7% of EMA studies assessing diet use a combined approach [59]. EMA is a powerful tool that can be used to validate automatically detected eating behavior in the field and to easily collect information about meaningful contexts; however, few studies have used this approach and still rely on paper–pen questionnaires to validate their findings [27].

## Limitations and Strengths

The M2FED study design had notable limitations. First, our method of obtaining ground-truth eating was only deployed via eating event–triggered EMA after an eating event was detected by the smartwatch. Thus, we could only verify true positive and false positive eating events. The M2FED system was not designed to verify true negative or false negative eating events, which limited our ability to calculate common evaluation metrics (ie, accuracy and $F_1$-score) and compare our results to other in-field studies described in the literature. Future research can build upon our study by implementing a verification of true negative and false negative eating events, via time-triggered

EMA or other methods, to gain a better understanding of the strengths and weaknesses of such an event detection algorithm.

Second, the false positive eating events were self-reported validation, which might be subject to social desirability in underreporting an eating event. This could potentially bias the validity of the results. Third, we encountered various difficulties with the deployed technologies, including smartwatches (ie, limited battery), mobile phones (ie, limited battery and app crashes), and the Wi-Fi router (ie, wireless connection dropped). Although these challenges were anticipated and were addressed in a timely manner on all occasions, some data were lost during the data collection process.

Finally, as the scope of this study only covered in-home eating behavior, we observed relatively few eating event–triggered EMAs per person across the 2-week study (approximately 8 per person). However, the range was very wide, indicating that some participants consumed more meals inside their homes compared with others. Reasons often provided informally by participants included eating all or most meals at school or work, working early or late, traveling for work, and participating in after-school extracurricular activities.

On the other hand, this study also possesses several strengths. First, we recruited a large and ethnically diverse sample of families from Los Angeles. It has been previously noted that the lack of diverse samples in eating-related mHealth and EMA studies is a major limitation of past research [60]. Second, as noted above, the M2FED study facilitated one of the longest in-field deployments found in the literature so far. Most ADA research has been conducted in the laboratory. By deploying in the field, we are able to better understand real-life eating behavior (vs eating behavior in a laboratory) and gain a better understanding of the challenges that arise when deploying wearable sensors outside of the laboratory. Third, as the deployment process was across a 2-year period, we were able to iteratively improve our automatic eating event detection algorithm and then use the newest version in the following deployments. Finally, this study produced momentary measures of theoretical constructs as well as momentary measures of eating behaviors. The theoretical work that we can now contribute would be to understand which constructs influence behavior, which behaviors influence various constructs, and which constructs play no role at all. We can also begin to understand the role of timing in these influences.

## Future Directions

The mHealth field is converging toward the use of a combination of user-friendly devices to assess eating behavior in the wild (eg, mobile phones and wrist-worn devices) [27,31]. Implementing user-friendly technologies for in-field dietary assessment or eating behavior interventions offers at least two substantial advantages—people are generally familiar with them [31] and may be willing to use them for longer periods compared with more intrusive devices. Although early studies experimented with less familiar, often not off-the-shelf technologies (eg, piezoelectric strain gauge sensors), most recent studies have opted for accelerometers and gyroscopes that are embedded within a wrist-worn smartwatch [27]. Furthermore, the combination of a wrist-worn smartwatch to automatically detect eating and a mobile or wearable device to capture ground-truth eating has been featured in a few studies published in the past year [61-63]. This approach is becoming more common, and these types of devices offer advantages for the user (participant) and make the use of mHealth technologies more accessible to nonengineering behavioral researchers. However, a number of related challenges have emerged. Future research will need to address comparability between newer technology-assisted measures and more traditional self-report measures of eating [64] versus other similar technology-assisted measures [27].

These user-friendly technologies also allow for passive measurement or low-effort reporting of various contexts and environments with relative ease. For example, fine-grained real-time GPS data can be scraped from both mobile devices and smartwatches to determine an individual's location and potentially assess the external influences on behavior [65,66]. Similarly, the social environment can be gleaned from wearable cameras [67], self-report EMA [68], or proximity Bluetooth sensors [69].

The ability to determine one's context or environment is a necessary component of ecological momentary interventions [70] or just-in-time interventions [71]. These types of intervention designs aim to provide the right amount of support at the right time and in the right context to promote behavior change [71-73]. These types of designs are well-suited for and offer unique opportunities for family-based settings [74]. They offer the ability to intervene in children and adolescents and can be designed to target the behavior of multiple family members at once [74]. As family members share genetic, environmental, and behavioral risks, family units are especially important targets for intervention and prevention [75] and have the potential to halt the intergenerational transmission of obesity and other chronic diseases.

## Conclusions

This paper demonstrates that EMA is a feasible tool to collect ground-truth eating activity and thus evaluate the performance of wearable sensors in the field. The combination of a wrist-worn smartwatch to automatically detect eating and a mobile or wearable device to capture ground-truth eating activity offers key advantages for the user (participant) and makes the use of mHealth technologies more accessible to nonengineering behavioral researchers.

## Acknowledgments

This work was supported in part by the National Science Foundation grants SCH-1521740 and SCH-1521722 and the National Institutes of Health grant U01CA229445. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the paper.

## Conflicts of Interest

None declared.

## References


1. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
2. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
3. Magarey A, Watson J, Golley RK, Burrows T, Sutherland R, McNaughton SA, et al. Assessing dietary intake in children and adolescents: considerations and recommendations for obesity research. Int J Pediatr Obes 2011;6(1):2-11. [doi: 10.3109/17477161003728469] [Medline: 20874449]
4. Willett W. 24-Hour Dietary Recall and Food Record Methods. Nutritional Epidemiology: Oxford University Press; 1998.
5. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Am J Clin Nutr 1997;65(4 Suppl):1100-1107. [doi: 10.1093/ajcn/65.4.1100S] [Medline: 9094905]
6. Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TE, Wilson MM, et al. The National Cancer Institute's Dietary Assessment Primer: a resource for diet research. J Acad Nutr Diet 2015;115(12):1986-1995 [FREE Full text] [doi: 10.1016/j.jand.2015.08.016] [Medline: 26422452]
7. Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. Br J Nutr 2004;92 Suppl 2:213-222. [doi: 10.1079/bjn20041169] [Medline: 15522159]
8. Westerterp KR, Goris AH. Validity of the assessment of dietary intake: problems of misreporting. Curr Opin Clin Nutr Metab Care 2002;5(5):489-493. [doi: 10.1097/00075197-200209000-00006] [Medline: 12172471]
9. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
10. Wehling H, Lusher J. People with a body mass index  30 under-report their dietary intake: a systematic review. J Health Psychol 2019;24(14):2042-2059. [doi: 10.1177/1359105317714318] [Medline: 28810493]
11. Novotny JA, Rumpler WV, Riddick H, Hebert JR, Rhodes D, Judd JT, et al. Personality characteristics as predictors of underreporting of energy intake on 24-hour dietary recall interviews. J Am Diet Assoc 2003;103(9):1146-1151 [FREE Full text] [doi: 10.1016/S0002-8223(03)00975-1] [Medline: 25587207]
12. Willett W. Nutritional Epidemiology. Oxford: Oxford University Press; 2013.
13. Nishida C, Uauy R, Kumanyika S, Shetty P. The Joint WHO/FAO Expert Consultation on diet, nutrition and the prevention of chronic diseases: process, product and policy implications. Public Health Nutr 2004;7(1a):245-250. [doi: 10.1079/PHN2003592]
14. Wood W, Neal DT. Healthy through habit: interventions for initiating and maintaining health behavior change. Behav Sci Policy 2016;2(1):71-83. [doi: 10.1353/bsp.2016.0008]
15. Neuhouser ML. The importance of healthy dietary patterns in chronic disease prevention. Nutr Res 2019;70(6):3-6 [FREE Full text] [doi: 10.1016/j.nutres.2018.06.002] [Medline: 30077352]
16. Jannasch F, Kröger J, Schulze MB. Dietary patterns and type 2 diabetes: a systematic literature review and meta-analysis of prospective studies. J Nutr 2017;147(6):1174-1182. [doi: 10.3945/jn.116.242552] [Medline: 28424256]
17. Higgs S, Thomas J. Social influences on eating. Curr Opin Behav Sci 2016;9:1-6. [doi: 10.1016/j.cobeha.2015.10.005]
18. Tourlouki E, Matalas AL, Panagiotakos DB. Dietary habits and cardiovascular disease risk in middle-aged and elderly populations: a review of evidence. Clin Interv Aging 2009;4:319-330 [FREE Full text] [doi: 10.2147/cia.s5697] [Medline: 19696896]
19. Robinson E, Thomas J, Aveyard P, Higgs S. What everyone else is eating: a systematic review and meta-analysis of the effect of informational eating norms on eating behavior. J Acad Nutr Diet 2014;114(3):414-429 [FREE Full text] [doi: 10.1016/j.jand.2013.11.009] [Medline: 24388484]
20. Reicks M, Banna J, Cluskey M, Gunther C, Hongu N, Richards R, et al. Influence of parenting practices on eating behaviors of early adolescents during independent eating occasions: implications for obesity prevention. Nutrients 2015;7(10):8783-8801 [FREE Full text] [doi: 10.3390/nu7105431] [Medline: 26506384]
21. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
22. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015;5(3):335-346 [FREE Full text] [doi: 10.1007/s13142-015-0324-1] [Medline: 26327939]
23. Spruijt-Metz D, Wen CK, Bell BM, Intille S, Huang JS, Baranowski T. Advances and controversies in diet and physical activity measurement in youth. Am J Prev Med 2018;55(4):81-91 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.012] [Medline: 30135037]

24. 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]
25. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen S, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
26. Engel SG, Crosby RD, Thomas G, Bond D, Lavender JM, Mason T, et al. Ecological momentary assessment in eating disorder and obesity research: a review of the recent literature. Curr Psychiatry Rep 2016;18(4):37. [doi: 10.1007/s11920-016-0672-7] [Medline: 26893235]
27. Bell BM, Alam R, Alshurafa N, Thomaz E, Mondol AS, de la Haye K, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. NPJ Digit Med 2020;3:38 [FREE Full text] [doi: 10.1038/s41746-020-0246-2] [Medline: 32195373]
28. Spruijt-Metz D, de la Haye K, Lach J, Stankovic J. M2FED: Monitoring and Modeling Family Eating Dynamics. In: Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM. 2016 Presented at: SenSys '16: The 14th ACM Conference on Embedded Network Sensor Systems; November 14 - 16, 2016; Stanford, CA USA p. 352-353. [doi: 10.1145/2994551.2996702]
29. Vu T, Lin F, Alshurafa N, Xu W. Wearable food intake monitoring technologies: a comprehensive review. Computers 2017;6(1):4. [doi: 10.3390/computers6010004]
30. Hassannejad H, Matrella G, Ciampolini P, De Munari I, Mordonini M, Cagnoni S. Automatic diet monitoring: a review of computer vision and wearable sensor-based methods. Int J Food Sci Nutr 2017;68(6):656-670. [doi: 10.1080/09637486.2017.1283683] [Medline: 28139173]
31. Heydarian H, Adam M, Burrows T, Collins C, Rollo ME. Assessing eating behaviour using upper limb mounted motion sensors: a systematic review. Nutrients 2019;11(5):1168 [FREE Full text] [doi: 10.3390/nu11051168] [Medline: 31137677]
32. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2016;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
33. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. J Acad Nutr Diet 2015;115(1):64-77. [doi: 10.1016/j.jand.2014.09.015] [Medline: 25441955]
34. Ye X, Chen G, Gao Y, Wang H, Cao Y. Assisting food journaling with automatic eating detection. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose, CA USA p. 3255-3262. [doi: 10.1145/2851581.2892426]
35. Gomes D, Sousa I. Real-time drink trigger detection in free-living conditions using inertial sensors. Sensors 2019;19(9):2145 [FREE Full text] [doi: 10.3390/s19092145] [Medline: 31075843]
36. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
37. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
38. Maher JP, Rebar AL, Dunton GF. Ecological momentary assessment is a feasible and valid methodological tool to measure older adults' physical activity and sedentary behavior. Front Psychol 2018;9:1485 [FREE Full text] [doi: 10.3389/fpsyg.2018.01485] [Medline: 30158891]
39. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
40. Nam S, Whittemore R, Vlahov D, Dunton G. 824-P: ecological momentary assessment of diabetes self-management: a systematic review of methods and procedures. Diabetes 2019;68(Supplement 1):824–P [FREE Full text] [doi: 10.2337/db19-824-P]
41. Dzubur E, Huh J, Maher JP, Intille SS, Dunton GF. Response patterns and intra-dyadic factors related to compliance with ecological momentary assessment among mothers and children. Transl Behav Med 2018;8(2):233-242 [FREE Full text] [doi: 10.1093/tbm/ibx002] [Medline: 29381785]
42. Rajkumar R, Lee I, Sha L, Stankovic J. Cyber-physical systems: the next computing revolution. In: Proceedings of the 47th Design Automation Conference. 2010 Presented at: DAC '10: The 47th Annual Design Automation Conference 2010; June 13 - 18, 2010; Anaheim, CA USA p. 731-736. [doi: 10.1145/1837274.1837461]
43. Mondol MA, Bell B, Ma M, Alam R, Emi I, Preum S, et al. MFED: a system for monitoring family eating dynamics. arXiv 2020 [FREE Full text]
44. Laurent J, Catanzaro SJ, Joiner TE, Rudolph KD, Potter KI, Lambert S, et al. A measure of positive and negative affect for children: scale development and preliminary validation. Psychol Assess 1999;11(3):326-338. [doi: 10.1037/1040-3590.11.3.326]

45. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]
46. Forrest CB, Ravens-Sieberer U, Devine J, Becker BD, Teneralli RE, Moon J, et al. Development and evaluation of the PROMIS pediatric positive affect item bank, child-report and parent-proxy editions. J Happiness Stud 2018;19(3):699-718 [FREE Full text] [doi: 10.1007/s10902-016-9843-9] [Medline: 29760578]
47. Terry PC, Lane AM, Lane HJ, Keohane L. Development and validation of a mood measure for adolescents. J Sports Sci 1999;17(11):861-872. [doi: 10.1080/026404199365425] [Medline: 10585166]
48. Cardello AV, Schutz HG, Lesher LL, Merrill E. Development and testing of a labeled magnitude scale of perceived satiety. Appetite 2005;44(1):1-13. [doi: 10.1016/j.appet.2004.05.007] [Medline: 15604029]
49. Framson C, Kristal AR, Schenk JM, Littman AJ, Zeliadt S, Benitez D. Development and validation of the mindful eating questionnaire. J Am Diet Assoc 2009;109(8):1439-1444 [FREE Full text] [doi: 10.1016/j.jada.2009.05.006] [Medline: 19631053]
50. Ma M, Alam R, Bell B, de la Haye K, Spruijt-Metz D, Lach J, et al. M^2G: a monitor of monitoring systems with ground truth validation features for research-oriented residential applications. In: Proceedings of the IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). 2017 Presented at: IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS); October 22 - 25, 2017; Orlando, FL USA. [doi: 10.1109/mass.2017.107]
51. Grolemund G, Wickham H. Dates and times made easy with lubridate. J Stat Soft 2011;40(3):1-25 [FREE Full text] [doi: 10.18637/jss.v040.i03]
52. Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol 2004;55:591-621. [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]
53. Cialdini RB, Trost MR. Social influence: social norms, conformity and compliance. In: Gilbert GT, Fiske ST, Lindzey G, editors. The Handbook of Social Psychology. New York: McGraw-Hill; 1998:151-192.
54. Navarathna P, Bequette B, Cameron F. Wearable device based activity recognition and prediction for improved feedforward control. In: Proceedings of the Annual American Control Conference (ACC). 2018 Presented at: Annual American Control Conference (ACC); June 27 - 29, 2018; Milwaukee, WI USA p. 27-29. [doi: 10.23919/acc.2018.8430775]
55. Thomaz E, Essa I, Abowd GD. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. 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; September 7 - 11, 2015; Osaka, Japan p. 1029-1040 URL: https://doi.org/10.1145/2750858.2807545 [doi: 10.1145/2750858.2807545]
56. Rast FM, Labruyère R. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments. J Neuroeng Rehabil 2020;17(1):148 [FREE Full text] [doi: 10.1186/s12984-020-00779-y] [Medline: 33148315]
57. Fontana JM, Farooq M, Sazonov E. Chapter 20 - Detection and characterization of food intake by wearable sensors. In: Sazonov E, editor. Wearable Sensors (Second Edition). Oxford: Academic Press; 2021:571-574.
58. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary intake and physical activity assessment: current tools, techniques, and technologies for use in adult populations. Am J Prev Med 2018;55(4):93-104 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.011] [Medline: 30241622]
59. Maugeri A, Barchitta M. A systematic review of ecological momentary assessment of diet: implications and perspectives for nutritional epidemiology. Nutrients 2019;11(11):2696 [FREE Full text] [doi: 10.3390/nu11112696] [Medline: 31703374]
60. Smith KE, Juarascio A. From Ecological Momentary Assessment (EMA) to Ecological Momentary Intervention (EMI): past and future directions for ambulatory assessment and interventions in eating disorders. Curr Psychiatry Rep 2019;21(7):53. [doi: 10.1007/s11920-019-1046-8] [Medline: 31161276]
61. Morshed MB, Kulkarni SS, Li R, Saha K, Roper LG, Nachman L, et al. A real-time eating detection system for capturing eating moments and triggering ecological momentary assessments to obtain further context: system development and validation study. JMIR Mhealth Uhealth 2020;8(12):e20625 [FREE Full text] [doi: 10.2196/20625] [Medline: 33337336]
62. Goldstein SP, Hoover A, Evans EW, Thomas JG. Combining ecological momentary assessment, wrist-based eating detection, and dietary assessment to characterize dietary lapse: A multi-method study protocol. Digit Health 2021;7:2055207620988212 [FREE Full text] [doi: 10.1177/2055207620988212] [Medline: 33598309]
63. Sen S, Subbaraju V, Misra A, Balan R, Lee Y. Annapurna: an automated smartwatch-based eating detection and food journaling system. Perv Mobile Comput 2020;68:101259 [FREE Full text] [doi: 10.1016/j.pmcj.2020.101259]
64. Fowler LA, Grammer AC, Staiano AE, Fitzsimmons-Craft EE, Chen L, Yaeger LH, et al. Harnessing technological solutions for childhood obesity prevention and treatment: a systematic review and meta-analysis of current applications. Int J Obes (Lond) 2021;45(5):957-981 [FREE Full text] [doi: 10.1038/s41366-021-00765-x] [Medline: 33627775]
65. Yang J, Wang J, Nakandala S, Kumar A, Jankowska M. Predicting eating events in free living individuals. In: Proceedings of the 15th International Conference on eScience (eScience). 2019 Presented at: 15th International Conference on eScience (eScience); September 24 - 27, 2019; San Diego, CA USA p. 627-629. [doi: 10.1109/escience.2019.00090]
66. Cetateanu A, Jones A. How can GPS technology help us better understand exposure to the food environment? A systematic review. SSM Popul Health 2016;2(6):196-205 [FREE Full text] [doi: 10.1016/j.ssmph.2016.04.001] [Medline: 28018957]

67. Gemming L, Doherty A, Utter J, Shields E, Ni Mhurchu C. The use of a wearable camera to capture and categorise the environmental and social context of self-identified eating episodes. Appetite 2015;92:118-125. [doi: 10.1016/j.appet.2015.05.019] [Medline: 26002278]
68. O'Connor SG, Habre R, Bastain TM, Toledo-Corral CM, Gilliland FD, Eckel SP, et al. Within-subject effects of environmental and social stressors on pre- and post-partum obesity-related biobehavioral responses in low-income Hispanic women: protocol of an intensive longitudinal study. BMC Public Health 2019;19(1):253 [FREE Full text] [doi: 10.1186/s12889-019-6583-x] [Medline: 30819155]
69. Mundnich K, Booth BM, L'Hommedieu M, Feng T, Girault B, L'Hommedieu J, et al. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. Sci Data 2020;7(1):354 [FREE Full text] [doi: 10.1038/s41597-020-00655-3] [Medline: 33067468]
70. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
71. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014;13(3):13-17. [doi: 10.1109/MPRV.2014.46]
72. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015;34(Suppl):1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
73. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
74. Heron KE, Miadich SA, Everhart RS, Smyth JM. Ecological momentary assessment and related intensive longitudinal designs in family and couples research. In: Fiese BH, Celano M, Deater-Deckard K, Jouriles EN, Whisman MA, editors. APA Handbook of Contemporary Family Psychology: Foundations, Methods, and Contemporary Issues Across the Lifespan. Washington, DC: American Psychological Association; 2019:369-385.
75. Kral TV, Rauh EM. Eating behaviors of children in the context of their family environment. Physiol Behav 2010;100(5):567-573 [FREE Full text] [doi: 10.1016/j.physbeh.2010.04.031] [Medline: 20457172]


## Abbreviations

**ADA:** automatic dietary assessment
**EMA:** ecological momentary assessment
**H-t-M:** hand-to-mouth
**M2FED:** Monitoring and Modeling Family Eating Dynamics
**mHealth:** mobile health
**OR:** odds ratio


*Edited by L Buis; submitted 10.05.21; peer-reviewed by L König, E Evans; comments to author 03.08.21; revised version received 28.09.21; accepted 03.12.21; published 18.02.22.*

*Please cite as:*
*Bell BM, Alam R, Mondol AS, Ma M, Emi IA, Preum SM, de la Haye K, Stankovic JA, Lach J, Spruijt-Metz D*
*Validity and Feasibility of the Monitoring and Modeling Family Eating Dynamics System to Automatically Detect In-field Family Eating Behavior: Observational Study*
*JMIR Mhealth Uhealth 2022;10(2):e30211*
*URL: https://mhealth.jmir.org/2022/2/e30211*
*doi:10.2196/30211*
*PMID:35179508*

Original Paper

# Continuous Noninvasive Remote Automated Blood Pressure Monitoring With Novel Wearable Technology: A Preliminary Validation Study

Michael H McGillion[1,2*], BScN, PhD; Nazari Dvirnik[2*], MD, MSc; Stephen Yang[3*], MD, CM, FRCP; Emilie Belley-Côté[2,4], MD, PhD, FRCP; Andre Lamy[2,4], MD, MHSc, FRCS; Richard Whitlock[2,4], MD, PhD; Maura Marcucci[2,4], MD, MSc; Flavia K Borges[2,4], MD, PhD; Emmanuelle Duceppe[5], MD, PhD, FRCP; Carley Ouellette[1,2], BScN, MSc; Marissa Bird[1], BA, BSN; Sandra L Carroll[1,2], PhD; David Conen[2,4], MD, MPH; Jean-Eric Tarride[4], PhD; Prathiba Harsha[6], MBBS, MSc, CCRA; Ted Scott[6], MAppSc, PhD; Amber Good[2], MSc; Krysten Gregus[6], BScN; Karla Sanchez[2], BSc; Pamela Benoit[7], BN; Julian Owen[7], BHSc, MD; Valerie Harvey[2], BSc; Elizabeth Peter[8], PhD; Jeremy Petch[2,4,9,10], PhD; Jessica Vincent[2], MClinEpi, CCRP; Michelle Graham[11], MD; P J Devereaux[2,4], MD, PhD

[1]School of Nursing, McMaster University, Hamilton, ON, Canada
[2]Population Health Research Institute, Hamilton, ON, Canada
[3]McGill University, Montreal, QC, Canada
[4]Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[5]Department of Medicine, Faculty of Nedicine, Université de Montréal, Montréal, QC, Canada
[6]Research and Innovation, Hamilton Health Sciences, Hamilton, ON, Canada
[7]Cardiac and Vascular Intensive Care Unit, Hamilton Health Sciences, Hamilton, ON, Canada
[8]University of Toronto, Toronto, ON, Canada
[9]Centre for Data Science and Digital Health, Hamilton Health Sciences, Hamilton, ON, Canada
[10]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
[11]University of Alberta, Edmonton, AB, Canada
[*]these authors contributed equally

**Corresponding Author:**
Michael H McGillion, BScN, PhD
School of Nursing
McMaster University
1280 Main St W
Hamilton, ON, L8S 4L8
Canada
Phone: 1 905 525 9140 ext 26140
Email: mmcgill@mcmaster.ca

## Abstract

**Background:** Wearable continuous monitoring biosensor technologies have the potential to transform postoperative care with early detection of impending clinical deterioration.

**Objective:** Our aim was to validate the accuracy of Cloud DX Vitaliti continuous vital signs monitor (CVSM) continuous noninvasive blood pressure (cNIBP) measurements in postsurgical patients. A secondary aim was to examine user acceptance of the Vitaliti CVSM with respect to comfort, ease of application, sustainability of positioning, and aesthetics.

**Methods:** Included participants were ≥18 years old and recovering from surgery in a cardiac intensive care unit (ICU). We targeted a maximum recruitment of 80 participants for verification and acceptance testing. We also oversampled to minimize the effect of unforeseen interruptions and other challenges to the study. Validation procedures were according to the International Standards Organization (ISO) 81060-2:2018 standards for wearable, cuffless blood pressure (BP) measuring devices. Baseline BP was determined from the gold-standard ICU arterial catheter. The Vitaliti CVSM was calibrated against the reference arterial catheter. In static (seated in bed) and supine positions, 3 cNIBP measurements, each 30 seconds, were taken for each patient with the Vitaliti CVSM and an invasive arterial catheter. At the conclusion of each test session, captured cNIBP measurements were

extracted using MediCollector BEDSIDE data extraction software, and Vitaliti CVSM measurements were extracted to a secure laptop through a cable connection. The errors of these determinations were calculated. Participants were interviewed about device acceptability.

**Results:** The validation analysis included data for 20 patients. The average times from calibration to first measurement in the static position and to first measurement in the supine position were 133.85 seconds (2 minutes 14 seconds) and 535.15 seconds (8 minutes 55 seconds), respectively. The overall mean errors of determination for the static position were –0.621 (SD 4.640) mm Hg for systolic blood pressure (SBP) and 0.457 (SD 1.675) mm Hg for diastolic blood pressure (DBP). Errors of determination were slightly higher for the supine position, at 2.722 (SD 5.207) mm Hg for SBP and 2.650 (SD 3.221) mm Hg for DBP. The majority rated the Vitaliti CVSM as comfortable. This study was limited to evaluation of the device during a very short validation period after calibration (ie, that commenced within 2 minutes after calibration and lasted for a short duration of time).

**Conclusions:** We found that the Cloud DX's Vitaliti CVSM demonstrated cNIBP measurement in compliance with ISO 81060-2:2018 standards in the context of evaluation that commenced within 2 minutes of device calibration; this device was also well-received by patients in a postsurgical ICU setting. Future studies will examine the accuracy of the Vitaliti CVSM in ambulatory contexts, with attention to assessment over a longer duration and the impact of excessive patient motion on data artifacts and signal quality.

**Trial Registration:** ClinicalTrials.gov NCT03493867; https://clinicaltrials.gov/ct2/show/NCT03493867

***(JMIR Mhealth Uhealth 2022;10(2):e24916)*** doi:10.2196/24916

**KEYWORDS**

validation study; continuous vital signs monitor; continuous non-invasive blood pressure monitoring; wearable; blood pressure; monitoring; validation; mHealth; vital sign; biosensor; accuracy; usability

## Introduction

### Background

Intraoperatively, continuous hemodynamic monitoring is the standard of care for patients undergoing surgery [1]. Continuous monitoring of patients' vital signs in the operating room (ie, blood pressure [BP], heart rate, respiratory rate, blood oxygen saturation [$SpO_2$], core body temperature, and electrocardiogram [ECG]) facilitates immediate recognition of hemodynamic instability and patient deterioration [1]. In contrast, once patients are transferred to surgical wards, their vital signs are assessed only periodically [2]. Hospital policies typically dictate that nursing staff assess patients' vital signs every 4 hours to 12 hours on surgical wards [2-6]. Patients are then discharged home routinely without surveillance [2]. Such infrequent in-hospital monitoring, followed by no monitoring at home, presents a danger to surgical patients. Cumulative published data support that deteriorations in patients' physiologic status in hospital, for example, often go undetected [7,8], conferring risk for hemodynamic compromise and serious postoperative adverse events (eg, hypotension leading to myocardial ischemia, stroke, and death).

New remote automated monitoring (RAM) technologies that enable continuous acquisition of physiologic data from biosensors; transmission, integration, and syntheses of multiple data sources to indicate patient status; as well as real-time alerts to clinicians have the potential to revolutionize the science of RAM [2]. Major developments in the field over the last decade include (1) the evolution of RAM systems capacity for semi-automatic (ie, clinician-promoted) discrete measurement of vital signs to fully automatic continuous measurement of vital signs; (2) the development of ultra-lightweight, unobtrusive sensors that facilitate unencumbered patient ambulation; and (3) the incorporation of more powerful microprocessors that enable higher sampling frequencies and, ultimately, higher fidelity signal inputs for increased precision of early adverse event detection [2,9,10]. These advancements are now seeing the commercial availability of a few noninvasive systems [11,12] that are capable of incorporating combinations of a number of vital signs parameters and related metrics, including heart rate, respiratory rate, skin temperature, $SpO_2$, BP, and movement.

Although significant progress has been made, continuous RAM systems are not yet in routine use in clinical care. A number of tactical and feasibility-related barriers remain, related to signal transmission, range, and speed; duration of power supply; as well as cybersecurity concerns [2,10]. At the clinical care level, a key barrier to advancing RAM has been the need to rely on systems that employ traditional methods for measuring BP noninvasively [2]. Such methods include the use of a sphygmomanometer with manual measurements by auscultation of Korotkoff sounds [13] or palpatory methods [14] and the derivation of automatic measurements through oscillometry [13]. These methods provide discrete or interval-based measurements with a pneumatic cuff typically situated on the brachial or radial arteries.

A challenge with systems that feature intermittent, pneumatic cuffs for the measurement of noninvasive BP is that they can be uncomfortable for patients and infeasible for longer-term patient monitoring [2]. Moreover, reliance on pneumatic cuffs does not help to overcome the problem of episodic vital sign measurement on surgical wards [2]. It is crucial that reliable, continuous, noninvasive blood pressure (cNIBP) measurement be achieved—while clinically important hypotension has been shown to have significant population-attributable risk for postoperative death and stroke, prolonged episodes of hypotension (and hypertension) are often missed in the context of intermittent BP monitoring [6-8].

Recent technologies for cNIBP measurement have emerged that utilize volume-clamp and arterial applanation tonometry methods [15]. Although these cNIBP methods have resulted in clinically accurate medical devices, they are limited in terms of portability and ambulatory use, shorter durations of application due to patient discomfort, and high cost [15]. The pursuit of cNIBP methods that provide seamless integration into a patient's daily activities and that offer a low-cost alternative while delivering clinical-grade BP metrics is a current focus for the biomedical engineering and RAM communities [10]. Cloud DX has developed one such device called the Vitaliti continuous vital signs monitor (CVSM), which supports the derivation of cNIBP through fundamental principles of biomechanics and pulse wave velocity [2].

**Figure 1.** The Vitaliti continuous vital signs monitor and user interface.


## Objectives

In accordance with standards set forth by the International Organization for Standardization (ISO 81060-2:2018) [16], we sought to establish the accuracy of Vitaliti CVSM cNIBP measurements versus gold standard invasive continuous arterial BP measurements in postsurgical patients. A secondary objective was to examine the usability of the Vitaliti CVSM with respect to perceived patient acceptance.

## Vitaliti Continuous Vital Signs Monitor

The Vitaliti CVSM [2,17] (Figure 1) is a wearable CVSM that can continuously and noninvasively measure 5-lead ECG, heart rate and heart rate variability, respiration rate, temperature (infrared sensor applied to the ear), $SpO_2$, and cNIBP [2]. See Multimedia Appendix 1 for details on Vitaliti CVSM donning, device configuration and features, and clinical workflow including calibration procedure.

# Methods

## Testing Authorization and Measurement Standards Requirements

The verification testing portion of this study received an investigational testing authorization (STP-VIT-002) for Class II medical devices from Health Canada. Study setting, inclusion criteria, and methods were in compliance with ISO 81060-2:2018 requirements [16], as described in the following sections.

### *ISO 81060-2-2018 Requirements*

For the demographic requirements, ISO 81060-2:2018 [16] stipulates that cNIBP testing must include a minimum of 15 patients and that 30% of the sample are male and 30% are female. In addition, those included for verification testing were to meet the following required proportions for baseline BP ranges [16]:

- At least 10% shall have a reference systolic blood pressure (SBP) ≤100 mm Hg (13.33 kPa).
- At least 10% shall have a reference SBP ≥160 mm Hg (21.33 kPa).
- At least 10% shall have a reference diastolic blood pressure (DBP) ≤70 mm Hg (9.33 kPa).
- At least 10% shall have a reference DBP ≥85 mm Hg (11.33 kPa).

In keeping with ISO restrictions for special populations [16], patients who were pregnant or experiencing cardiac arrhythmias were to be excluded.

For the accuracy requirements, according to the ISO standard [16], one determination of cNIBP measurement represents the average of one 30-second interval for a given patient position. To ensure equal weighting of BP measurements across participants, the ISO standard also requires that *no more than 10* BP measurements be included per patient. Thus, for each test session, 3 separate 30-second determinations were calculated per patient for each position for both the arterial catheter reference and the Vitaliti CVSM. Errors of each measurement determination were calculated. If the determination of the Vitaliti CVSM was within 1 (±) SD of the determination of the arterial

catheter, the error of that determination equaled 0. If any SBP or DBP determination from the Vitaliti CVSM was outside of 1 (±) SD of the corresponding arterial catheter determination, then the error for that determination equaled the upper or lower limit of the arterial catheter reference measurement minus the Vitaliti CVSM determination [16].

All errors of valid, paired BP determinations (included participants only) were then used to calculate the experimental mean and SD of errors for SBP and DBP. If the mean of the errors of determination was not greater than 5 mmHg and the SD of the error was not greater than 8 mmHg, then the Vitaliti CVSM device was determined to be compliant with ISO guidelines [16].

Bland-Altman plots [18] were generated to visualize agreement between arterial catheter and Vitaliti CVSM mean BP measurements and inspect the bias (ie, mean error) and distribution of errors of determination within 95% limits of agreement (ie, ±1.96 SD).

## Setting and Participants

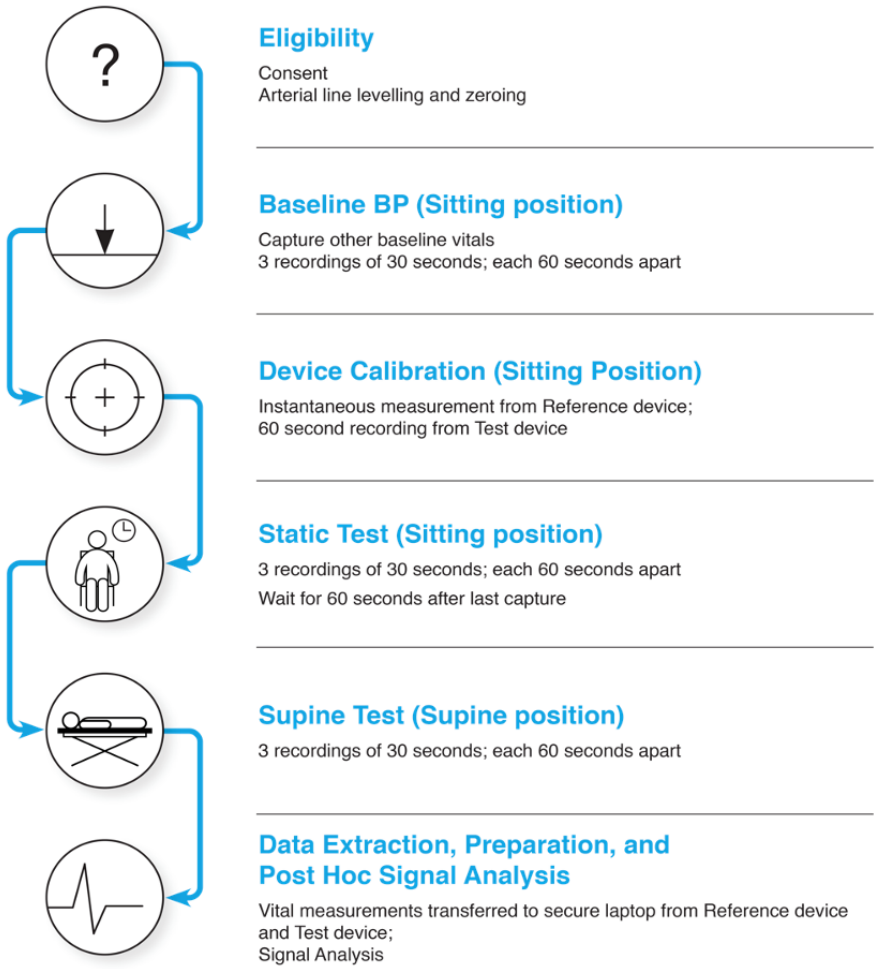
This study required comparison of the Vitaliti CVSM to a gold-standard comparator for continuous BP measurement. We therefore required access to patients with an invasive arterial catheter for hemodynamic monitoring. Recruited participants provided written, informed consent and included patients 18 years of age or older who underwent cardiac surgery and were admitted for immediate postoperative recovery in the Hamilton Health Sciences (Hamilton General Hospital site) Cardiac Surgical Intensive Care Unit (ICU) with an arterial line in situ. The ICU setting was chosen given that arterial lines for the continuous measurement of SBP, DBP, and mean arterial pressure are the standard of care in this setting. Moreover, based on operating room schedules, the cardiac surgical unit had predictable patient flows, allowing for planning and efficient execution of study procedures. The study coordinating center was the Population Health Research Institute (PHRI) in Hamilton, Ontario.

Given minimal ISO requirements [16] for participant numbers and prespecified baseline BP ranges, we targeted a maximum recruitment of 80 participants for verification and acceptance testing. We intentionally oversampled given the high likelihood of labile hemodynamic status in postoperative cardiac surgery patients. Based on clinical experience, we anticipated that abrupt changes in baseline BP, or other aspects of physiologic status, would preclude moving forward with testing procedures for some participants. Given the complexity of the clinical setting, we also oversampled in anticipation of interruptions to study procedures (eg, immediate patient care needs, emergencies) and technical challenges with respect to data downloads and intersystem comparisons in the context of real-time cNIBP monitoring.

## Procedures and Data Collection

Study flow is depicted in Figure 2. Patients expected to fulfill eligibility criteria were first approached and invited to participate by the ICU nurse educator. Those interested in hearing more were then approached by study personnel to obtain written, informed consent and collect baseline demographic information.

**Figure 2.** Study flow diagram.




## Continuous Blood Pressure Measurement Requirements and Data Collection

### *ISO Guidance*

ISO [16] stipulates that, in the context of cNIBP testing where gold-standard comparator devices and test devices are cuffless, cNIBP determinations measured during 30-second intervals are considered equivalent to a single determination with a traditional cuff-based sphygmomanometer. This guidance was used to capture baseline BP recordings as well as all ISO [16] test recordings, as described in the following sections.

### *Baseline Blood Pressure Recording*

Baseline BP was determined from the gold-standard ICU arterial catheter. The patients' ICU nurse first levelled and zeroed the ICU arterial catheter transducer (TruWave disposable pressure transducer; Edwards Lifesciences, Irvine, CA) to achieve consistent reference measurements and to negate the influence of external atmospheric pressure on BP recordings. Per ISO requirements [16], 3 BP recordings were then taken to establish each patient's baseline BP category. Patients were assisted by the ICU nurse into a seated position in bed and asked to sit quietly. Three BP recordings, lasting 30 seconds each, were taken by the research assistant; each of these readings was taken 60 seconds apart. The mean value of these readings was defined as the patient's baseline BP, and this value was logged according to the appropriate ISO category [16], as applicable. Those patients who did not meet one of the prespecified baseline BP category requirements were immediately excluded, and their participation was discontinued.

### *Vitaliti CVSM Donning Process and Setup*

Per manufacturer instructions, the research assistant placed the Vitaliti CVSM around the patient's neck and positioned the collar to be flush with the neck and shoulders. The flexibility of the device allowed for positioning of the collar and contact electrodes on the chest, such that surgical site dressings or ICU equipment and tubing were unencumbered. A disposable sheath (for infection control purposes) was placed on the tip of the earpiece, which was then positioned in the patient's ear. The research assistant then used a tablet to access Vitaliti companion software, in order to conduct a systems check. This check included ensuring proper positioning and contact of all sensors on the patient, as well as real-time visualization of the capture of all biometrics and physiologic wave forms.

### *Vitaliti CVSM Calibration*

Following baseline BP assessment and equipment setup, the Vitaliti CVSM was calibrated against the reference arterial catheter. Patients were again asked to sit still and refrain from movement or talking during this step. The research assistant first recorded an instantaneous reference BP reading from the arterial catheter and registered this value into the Vitaliti

application on the tablet. Next, the Vitaliti system captured and analyzed the patients' vital metrics and physiological signals for a period of 60 seconds, in order to calibrate against the reference measurement. At the conclusion of this calibration step, the Vitaliti system analyzed the recorded data to ensure consistent signal quality and that there were limited movement artifacts. If the Vitaliti software indicated to the research assistant that the calibration was unsuccessful, the procedure was repeated.

### *Test Blood Pressure Recordings: Static and Supine Positions*

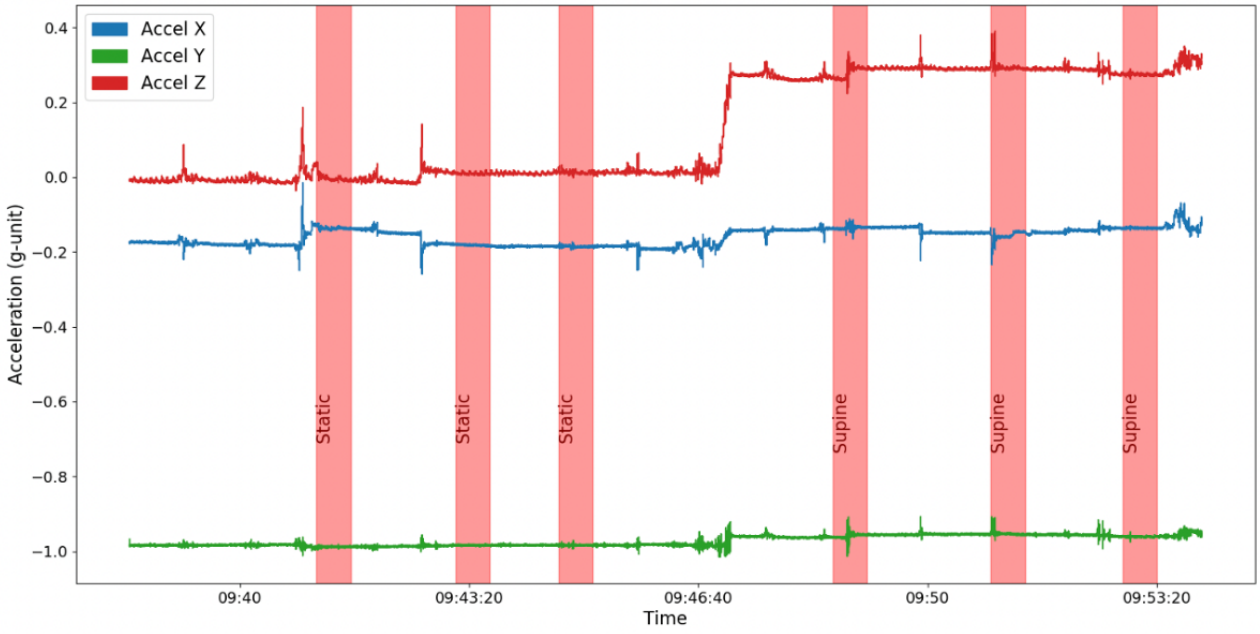
Simultaneous cNIBP readings from the arterial catheter and the Vitaliti CVSM were first captured with the patient in the seated, static position (in bed). These simultaneous measurements were captured for a period of 10 minutes, without interruption. Following this procedure, the patient was assisted by the ICU nurse into the supine position, in order to achieve a change in posture for continued measurement. Another 10 minutes of simultaneous cNIBP recordings were then captured.

## Data Extraction and Preparation for Analysis

At the conclusion of each test session, captured cNIBP measurements were extracted from the arterial catheter ICU monitor device using MediCollector BEDSIDE [19] data extraction software. Vitaliti CVSM measurements were extracted by connecting the device USB Type C port to a secure study laptop with a cable. Custom Python scripts were provided by Cloud DX to extract measurements from the Vitaliti CVSM at 1-second time-stamped intervals, for comparison against the arterial catheter data. Both devices were synchronized to ensure time alignment in post-signal processing.

The 10-minute intervals of cNIBP recordings with patients in seated (static) and supine position alignments were verified with Vitaliti CVSM accelerometer and gyroscope data collected during the test period (Figure 3). Per ISO requirements [16], for each patient, we isolated 3 separate 30-second intervals of cNIBP measurements for seated and supine positions; each of these determinations was at least 1 minute apart. Each 30-second interval also had to feature uninterrupted cNIBP measurements without any measurement loss from either the arterial catheter or the Vitaliti CVSM. The 30-second intervals selected for analysis were extracted to allow for 2-minute transition periods between patient positions in order to ensure stable measurements.

**Figure 3.** Tri-axial accelerometer data showing static and then supine patient positions with overlaid 30-second measurement intervals.




## Vitaliti Device—Perceived User Acceptance

Human factors testing is utilized to evaluate if a medical device can support users in the intended environment for all critical tasks [20]. To provide an assessment of elements of human factors and usability of the Vitaliti CVSM from the patient perspective, an exit interview was conducted with patients by the clinical team, and an online survey was completed to capture their responses. A customized questionnaire was developed in keeping with regulatory guidance provided by the Food and Drug Administration (FDA) [20] for applying human factors and usability engineering to medical devices. This questionnaire consisted of 13 questions to establish user acceptance of the Vitaliti CVSM with respect to comfort, ease of application, sustainability of positioning, and aesthetics; possible responses to each item ranged from "strongly disagree" to "strongly agree." See Multimedia Appendix 2 for the questionnaire.

## Data Analyses

### *Demographic Characteristics and User Acceptance Ratings*

Descriptive statistics were used to summarize participants' demographic characteristics and user acceptance ratings. The distribution of patients across baseline Association for the Advancement of Medical Instrumentation (AAMI)-ISO BP

categories [16] was summarized. Analyses were performed using SAS/STAT (release 9.4) statistical analysis software. Prior to analyses, Python was used to prepare BP recordings captured by the reference arterial catheter and the Vitaliti CVSM for post hoc signal analyses.

### *Post Hoc Signal Analyses*

In accordance with AAMI-ISO guidelines [16], all cNIBP measurements, in either the static or supine position and for any given participant, were excluded if (1) the invasive reference SBP had a range ≥20 mm Hg (2.67 kPa) or (2) the invasive reference DBP had a range ≥12 mm Hg (1.6 kPa). All cNIBP recordings were evaluated against these criteria. All participants with data meeting these constraints were removed from validation analyses.

Two additional assessments of signal quality were performed on the physiological data captured by the Vitaliti CVSM. First, tri-axial accelerometer and gyroscope data were reviewed to identify any test sessions with an excessive amount of movement that impacted the quality of photoplethysmography and ECG signals (metrics vital to the derivation of cNIBP). An activity index based on accelerometer data [21], defined as the time derivative of acceleration, was used to evaluate the amount of movement by each patient during test sessions. This index reflects the combined impact of the rate at which a patient's acceleration measurements change with respect to time in 3 perpendicular planes of movement. An average value of the activity index was calculated for all activity throughout the static and supine positions. An equivalent activity index has been used for mobile application–based activity monitoring and a wellness motivation system for senior adults [21]. Vitaliti CVSM activity intensities were empirically derived during data collection; a critical threshold of 2.4 gravities per second (g/s) represented "vigorous patient activity," equivalent to patient movement on a treadmill with a speed of 4.5 miles to 5 miles per hour [21]. Activity levels above this threshold were deemed to have negatively impacted signal quality for the purposes of our validation analysis. Data from participants with an index score ≥2.4 g/s were removed from the study [21].

Second, all ECG recordings were examined for noise, given that poor signal quality would introduce false positive R peaks in the QRS complex, which could affect the performance of the BP algorithm [22]. Cases were identified where the ECG signal quality was low, such that R peaks could not be reliably determined in the resultant signal; these cases were subsequently removed prior to validation analysis [22].

### *Validation Analyses*

According to the ISO standard [16], one determination of cNIBP measurement represents the average of one 30-second interval for a given patient position. Thus, for each test session, 3 determinations were calculated for each position, for both the arterial catheter reference and the Vitaliti CVSM. Errors of each measurement determination were calculated. If the determination of the Vitaliti CVSM was within 1 (±) SD of the determination of the arterial catheter, the error of that determination equaled 0. If any SBP or DBP determination from the Vitaliti CVSM was outside of 1 (±) SD of the corresponding arterial catheter determination, then the error for that determination equaled the upper or lower limit of the arterial catheter reference measurement minus the Vitaliti CVSM determination [16].

All errors of valid, paired BP determinations (included participants only) were then used to calculate the experimental mean and SD of errors for SBP and DBP. If the mean of the errors of determination was not greater than 5 mm Hg and the SD of the error was not greater than 8 mm Hg, then the Vitaliti CVSM device was determined to be compliant with ISO guidelines [16]. Bland-Altman plots [23] were generated to visualize agreement between arterial catheter and Vitaliti CVSM mean BP measurements and inspect the bias (ie, mean error) and distribution of errors of determination within 95% limits of agreement (ie, ±1.96 SD).

## Results

### Demographics

Derivation of the sample is presented in Figure 4. In total, 202 patients were screened for inclusion in the cardiac ICU between June 2018 and October 2019. Of these, 118 were ineligible due to baseline BPs outside of ISO requirements [16], current arrhythmia (ie, atrial fibrillation), or pregnancy; 7 patients declined participation; and 77 eligible patients consented to participate. Of the 77 eligible patients who consented to participate, an additional 22 were excluded due to technical challenges that precluded completion of the test sessions, shift in BP outside of study requirements, and development of a new arrhythmia (ie, atrial fibrillation) prior to the start of testing procedures. Technical challenges included wireless communication problems, data extraction software failures, reference device data transfer problems, and sensor disconnections. A total of 55 patients were included for validation testing procedures. Of these, 35 patients (64%) were male, and 20 patients (36%) were female; the mean age was 64 (SD 11.5) years (Table 1). Most patients had undergone cardiac surgery (33/55, 60%) including coronary artery bypass grafting or aortic valve repair.

**Figure 4.** Derivation of the study sample. BP: blood pressure; ECG: electrocardiogram; ISO: International Organization for Standardization.

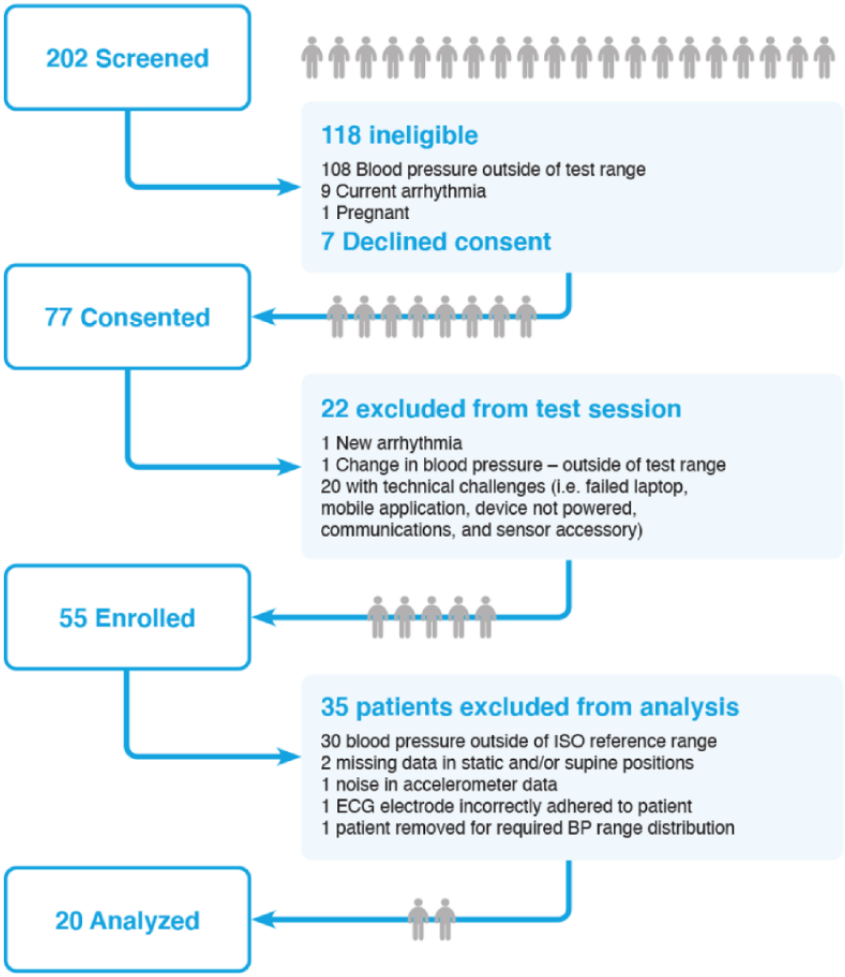


**Table 1.** Patient characteristics.

| Patient characteristics | Values for entire sample (n=55) | Values for subgroup of patients included in the validation (n=20) |
|---|---|---|
| Age (years), mean (SD) | 64 (11.5) | 64.7 (10.9) |
| **Sex, n (%)** | | |
| Male | 35 (64) | 11 (55) |
| Female | 20 (36) | 9 (45) |
| **BMI (kg/m$^2$), n (%)** | | |
| <18.5 (underweight) | 2 (4) | 0 (0) |
| 18.5-24.9 (healthy) | 12 (22) | 5 (25) |
| 25.0-29.9 (overweight) | 18 (33) | 8 (40) |
| 30.0-34.9 (obese I) | 9 (16) | 3 (15) |
| 35.0-39.9 (obese II) | 7 (13) | 1 (5) |
| ≥40 (obese III) | 6 (11) | 2 (10) |
| Unavailable[a] | 1 (2) | 1 (5) |
| **Cardiac surgery, n (%)** | | |
| Total | 33 (60) | 15 (75) |
| Coronary artery bypass grafting (CABG) | 25 (76)[b] | 13 (87)[c] |
| Aortic valve repair or replacement (AVR) | 6 (18)[b] | 1 (7)[c] |
| Other | 2 (6)[b] | 1 (7)[c] |
| **Vascular surgery, n (%)** | | |
| Total | 11 (20) | 3 (15) |
| Open abdominal aortic aneurysm repair | 1 (9)[d] | 1 (33)[e] |
| Aorto-femoral bypass | 2 (18)[d] | 0 (0)[e] |
| Axillo-femoral bypass | 1 (9)[d] | 0 (0)[e] |
| Femoral-femoral bypass | 1 (9)[d] | 0 (0)[e] |
| Other | 6 (55)[d] | 2 (67)[e] |
| Other type of surgery, n (%) | 11 (20) | 2 (8) |

[a]Patient height, weight, or BMI data unavailable from clinical record.

[b]n=33.

[c]n=15.

[d]n=11.

[e]n=3.

## Post Hoc Signal Analysis

Based on signal analysis, the data for an additional 30 patients were excluded due to arterial catheter reference SBP (≥20 mm Hg) or DBP (≥12 mm Hg) ranges falling outside of ISO requirements during each 30-second measurement interval in both static and supine testing positions. The data for 2 additional patients were excluded due to missing data segments (for either the arterial catheter reference or Vitaliti CVSM) that precluded analysis. Missing data were caused by unforeseen interruptions in data transmission related to excessive movement of the arterial catheter transducer or accidental disconnection or displacement of the Vitaliti CVSM earpiece or ECG electrodes.

Finally, the data for the last patient recruited with baseline BP within the normal range were excluded, as this group would have been over-represented for the required ISO BP range distributions. Following signal analysis, the cNIBP data of 20 patients were included for validation analyses. Table 1 presents the demographic characteristics of all 55 patients enrolled, as well as the demographic characteristics of the 20-patient subgroup (of the 55 enrolled patients) who was included in validation analyses.

In accordance with the AAMI-ISO guidelines [16], of the 20 patients whose data were included for validation analyses, a minimum of 30% were male, and a minimum of 30% were female. Baseline arterial catheter cNIBP measurements also

spanned high and low systolic and diastolic ranges, with at least 10% of readings falling into each AAMI-ISO prespecified category (See Table 2).

**Table 2.** Blood pressure distribution of patients included in data analysis (n=20).

| Characteristic | Results, n (%) |
|---|---|
| **Entry ISO[a] BP[b] range** | |
| Normal: SBP[c] >100 mm Hg and <160 mm Hg and DBP[d] >70 mm Hg and <85 mm Hg | 11 (55) |
| SBP ≤100 mm Hg | 2 (10) |
| DBP ≤70 mm Hg | 6 (30) |
| SBP ≥160 mm Hg | 3 (15) |
| DBP ≥85 mm Hg | 2 (10) |
| **Sex** | |
| Male | 11 (55) |
| Female | 9 (45) |

[a]ISO: International Organization for Standardization.

[b]BP: blood pressure.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

## Validation of Continuous Noninvasive Blood Pressure Measurements

For each of the 20 patients included in the final analysis, 3 determinations were calculated for both the reference arterial catheter and Vitaliti CVSM within each position (static and supine), resulting in a total of 60 average SBP and DBP measurements. The average time elapsed from calibration to first measurement in the static position was 133.85 seconds (2 minutes 14 seconds). The average time elapsed from calibration to first measurement in the supine position was 535.15 seconds (8 minutes 55 seconds). With respect to delimitation of the total validated time frame across patient positions, (1) the minimum and maximum times elapsed from calibration to first measurement in the static position were 14.0 seconds and 1392.0 seconds, respectively, and (2) the minimum and maximum times elapsed from calibration to last measurement in the supine position were 575.0 seconds and 2274.0 seconds, respectively.

The errors of determination between the 2 devices were calculated, as described in the Methods section. Bland-Altman plots [18], illustrating agreement between the arterial catheter reference and the Vitaliti CVSM for each of these errors of determination by patient position (static and supine), are presented in Figures 5-8. The mean (horizontal axis) and errors (vertical axis) of each determination are presented, along with the mean error and limits of agreement (±1.96 SD). In the static position, Bland-Altman plots illustrated a mean error of determinations of –0.62 mm Hg and 95% limits of agreement of –9.64 mm Hg to 8.40 mm Hg for SBP measurements and a mean error of determinations of 0.46 mm Hg and 95% limits of agreement of –2.80 mm Hg to 3.71 mm Hg for DBP measurements. In the supine position, Bland-Altman plots revealed a greater mean error of determinations (2.72 mm Hg) and 95% limits of agreement of –7.40 mm Hg to 12.84 mm Hg for SBP measurements and mean error of determinations of 2.65 mm Hg and 95% limits of agreement of –3.61 mm Hg to 8.91 mm Hg for DBP measurements.

**Figure 5.** Bland-Altman plot of systolic blood pressure determinants from the Vitaliti continuous vital signs monitor (CVSM) versus the arterial line in the static position.

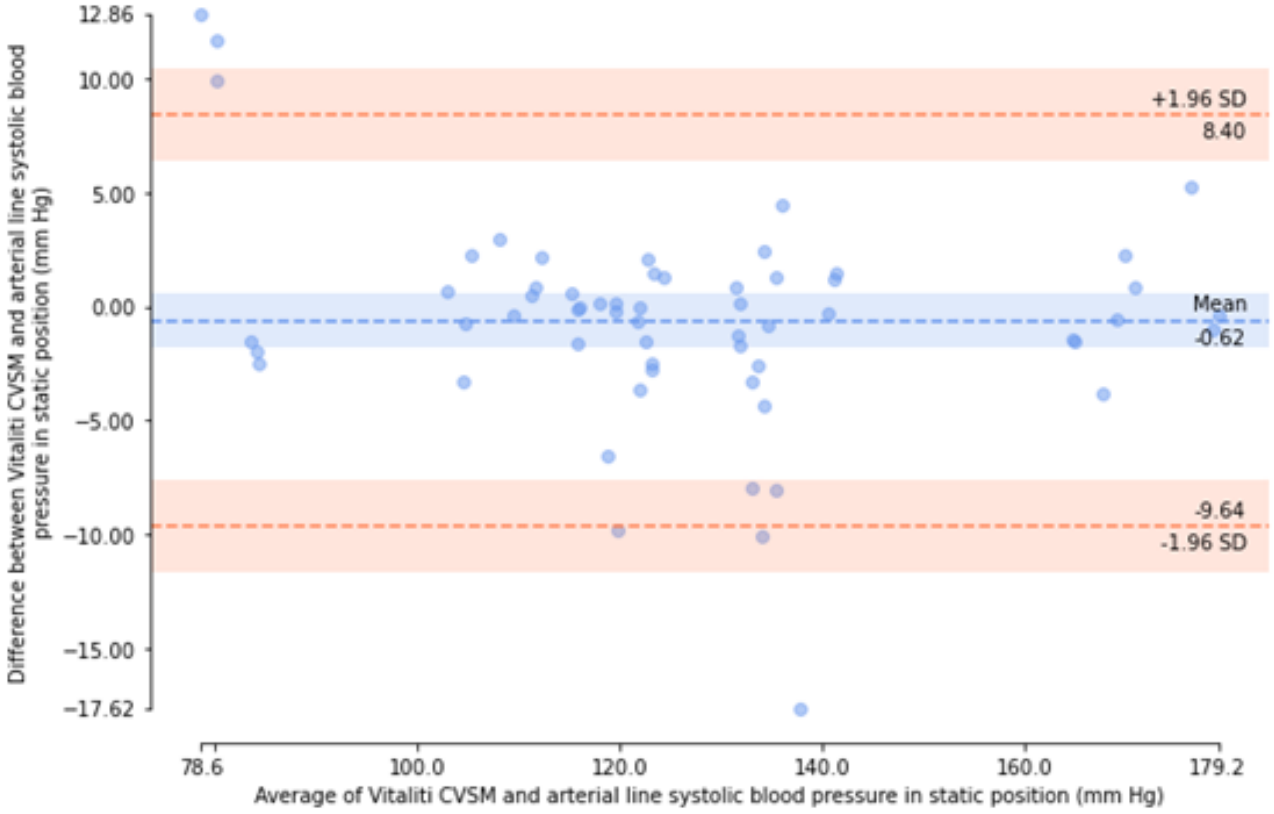



**Figure 6.** Bland-Altman plot of diastolic blood pressure determinants from the Vitaliti continuous vital signs monitor (CVSM) versus the arterial line in the static position.

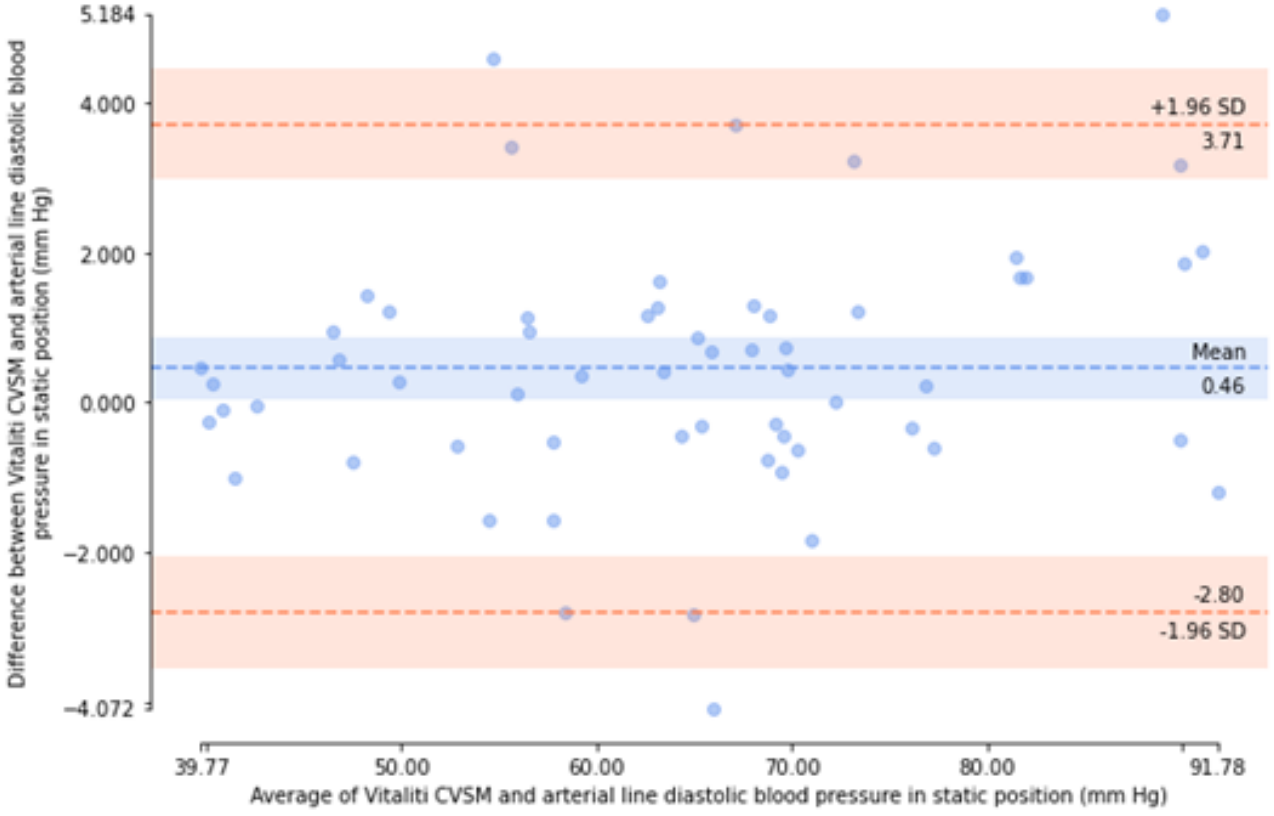


**Figure 7.** Bland-Altman plot of systolic blood pressure determinants from the Vitaliti continuous vital signs monitor (CVSM) versus the arterial line in the supine position.

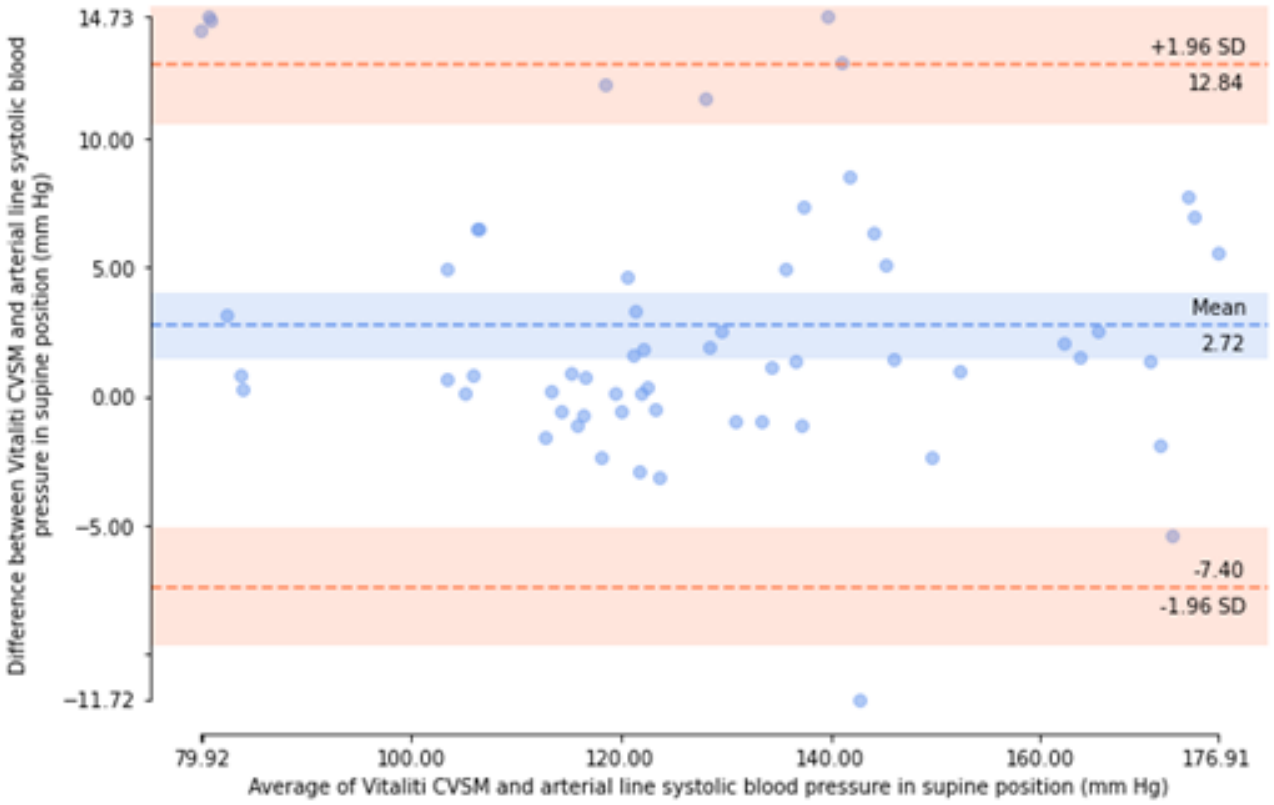
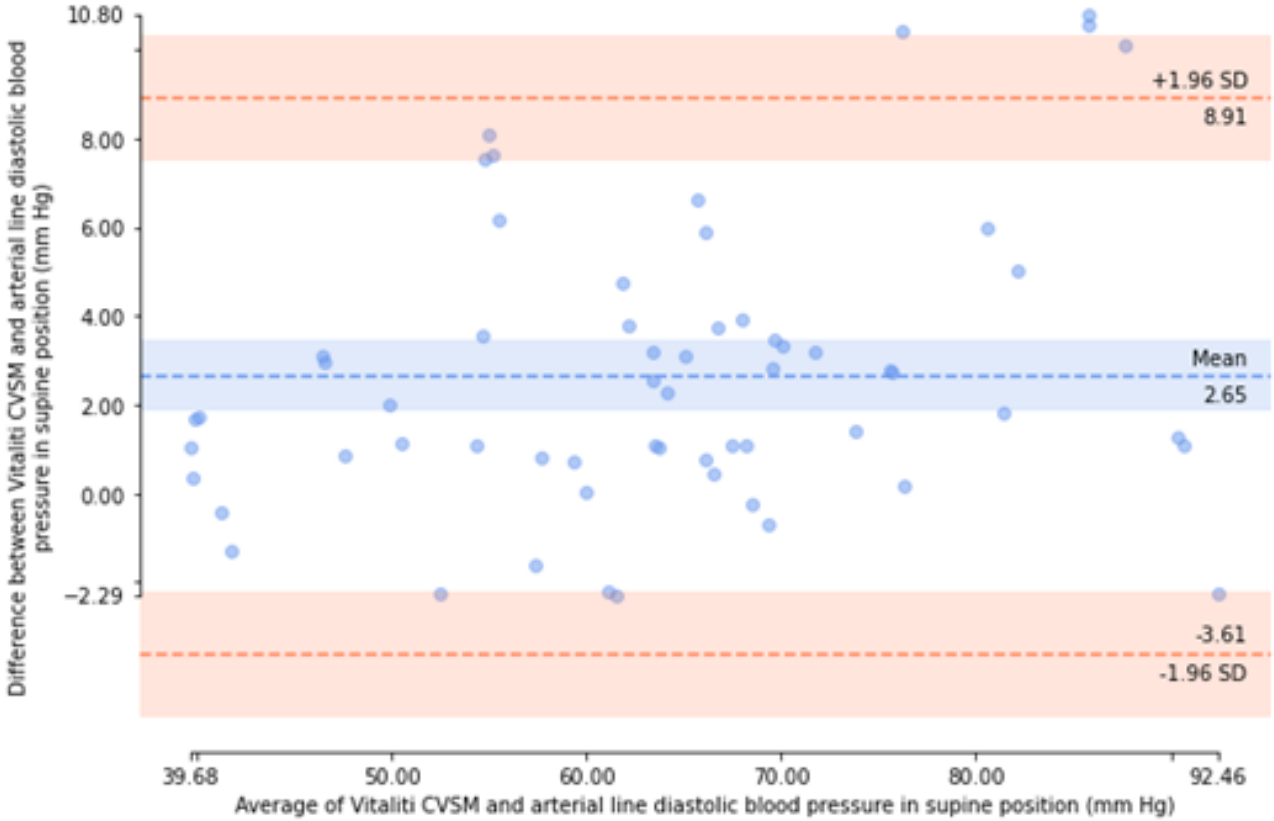



**Figure 8.** Bland-Altman plot of diastolic blood pressure determinants from the Vitaliti continuous vital signs monitor (CVSM) versus the arterial line in the supine position.



Per ISO requirements, Table 3 summarizes the errors of determination for SBP and DBP measurements in the static and supine positions. The overall means of the errors of determination for the static position were –0.621 (SD 4.640) mm Hg for SBP and 0.457 (SD 1.675) mm Hg for DBP. Errors of determination were slightly higher for the supine position, at 2.722 (SD 5.207) mm Hg for SBP and 2.650 (SD 3.221) mm Hg for DBP. These results indicate compliance with the ISO standard [16], which stipulates that errors of determination should not exceed 5 mmHg and that the SD of the error not exceed 8 mm Hg.

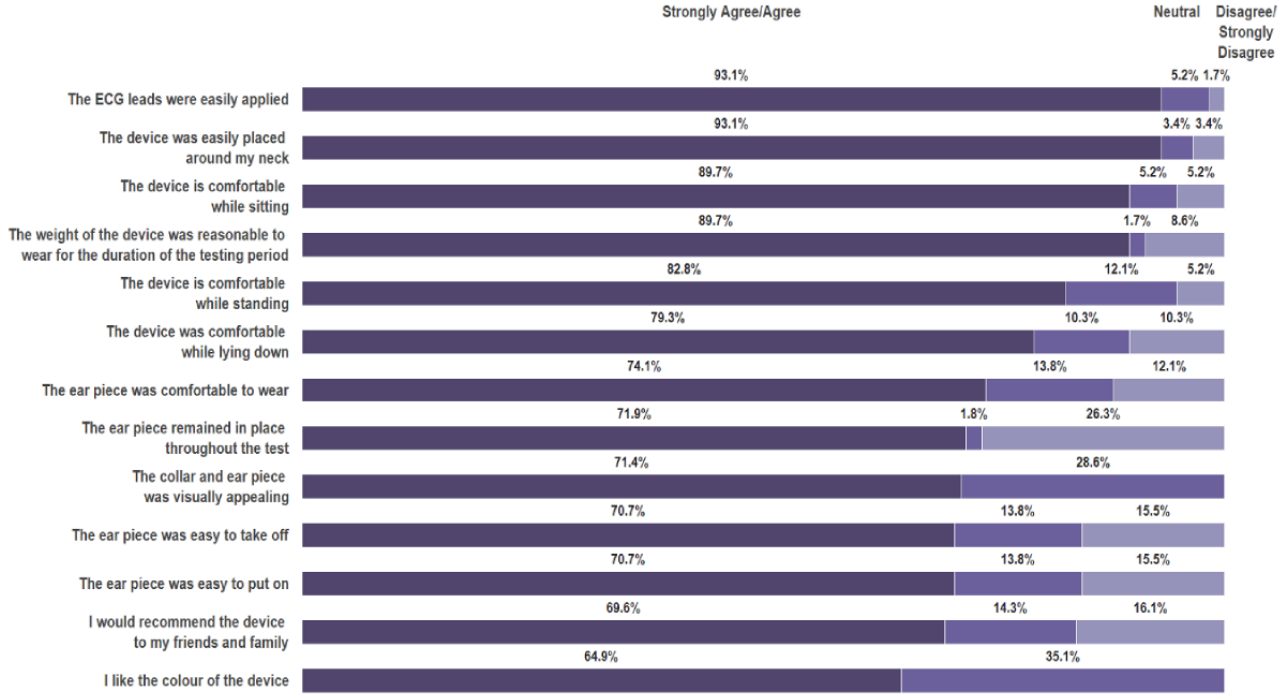
**Table 3.** Summary of the errors of determination for systolic blood pressure (SBP) and diastolic blood pressure (DBP) measurements in static and supine positions.

| Position | SBP | DBP |
|---|---|---|
| **Static position** | | |
| Number of observations | 60 | 60 |
| Mean of the errors (mm Hg) | –0.621 | 0.457 |
| SD of the errors (mm Hg) | 4.640 | 1.675 |
| **Supine position** | | |
| Number of observations | 60 | 60 |
| Mean of the errors (mm Hg) | 2.722 | 2.650 |
| SD of the errors (mm Hg) | 5.207 | 3.221 |

## Patient Usability Feedback

The responses from the 58 patients who responded to the human factors and usability feedback questionnaire are summarized in Figure 9. Questions related to ease of donning Vitaliti and device accessories yielded high acceptance ratings, with 54 (54/58, 93%) agreeing or strongly agreeing. Responses regarding device comfort in multiple positions were also positive, ranging from 52 (52/58, 90%) participants responding with strongly agree/agree to 40 (40/58, 75%) participants responding with strongly agree/agree. Questions regarding aesthetics of the device provided more neutral responses (range: 16/58, 28% to 20/58, 35%), indicating that these aspects were of lesser importance for most. The item regarding the sustainability of the positioning of the earpiece yielded the greatest amount of negative responses (disagree/strongly disagree: 15/58, 26%), indicating that dislodgment of this sensor during testing was an issue for just over 25% of respondents.

**Figure 9.** Usability feedback. ECG: electrocardiogram.



## Discussion

### Principal Findings

This validation study addressed the accuracy and usability of the Vitaliti CVSM for the measurement of cNIBP in an ICU setting. Device performance was evaluated based on the ISO 81060-2:2018 standard [16] for clinical investigation with a reference invasive arterial catheter BP monitor. This standard is accepted by the FDA and Health Canada. Human factors validation testing was also performed to evaluate if the Vitaliti CVSM was acceptable to users. A voluntary exit interview was conducted to capture participants' usability feedback. In total, 55 participants completed the testing procedures and the exit interview. Data for 20 of these participants were included for validation analyses.

When comparing the accuracy of the Vitaliti CVSM to a gold-standard reference, we found errors of determination of –0.621 (SD 4.640) mm Hg for SBP and 0.457 (SD 1.675) mm Hg for DBP in the static (seated) position. Errors of determination were slightly higher when patients were supine, at 2.722 (SD 5.207) mm Hg for SBP and 2.650 (SD 3.221) mm Hg for DBP. These results indicate compliance with the ISO standard [16], which stipulates that errors of determination should not exceed 5 mm Hg and that the SD of those errors should not exceed 8 mm Hg.

This validation study also demonstrated a high degree of usability in terms of perceived patient acceptance of the Vitaliti CVSM throughout the test procedures and positions (ie, static and supine). Displacement of the earpiece was the most negative aspect of patient experience. Earpiece displacement was caused by our use of an off-the-shelf disposable ear sheath to cover the light emitting diode (LED) sensor in order to comply with hospital infection requirements. In future work, the Vitaliti CVSM will require custom earpiece sheathing with improved fit in order to provide more secure mounting of the LED sensor and improved patient comfort.

### Comparison With Prior Work

A few comparable studies have assessed the accuracy of wearable technologies for continuous vital sign measurement in hospital. A pilot study by Weenk et al [11] (n= 20) investigated the use of 2 wearable technologies, ViSi Mobile and HealthPatch, to continuously measure vital signs (ie, heart rate, respiration rate, $SpO_2$, BP) of patients admitted to an internal medicine and surgical ward; patients' vital signs were measured continuously for 2 days to 3 days. A comparison was performed between continuous ViSi Mobile and HealthPatch measurements and nurses' manual vital sign observations entered into an electronic medical record. Results demonstrated that, in general, nurses' manual vital sign observations correlated well with paired instances of continuous vital sign measurements. Data artifacts and data outages were noted concerns and were attributed to Wifi connectivity challenges and signal noise in the context of patient ambulation [11].

More recently, Downey et al [12] investigated the reliability of a wireless wearable patch, SensiumVitals, to monitor vital signs (ie, temperature, heart rate, respiration rate) continuously from patients (n=51) following major elective general surgery. Nurses' manual vital sign observations were compared against paired instances of SensiumVitals biometric measurements. A median of 19 sets of manual measurements was captured for each patient, for a total of 1135 observation sets of paired comparisons for analysis [12]. In contrast to the results of this study, the error between manual and continuous vital sign measurements in the study by Downey et al [12] did not fall within prespecified limits of agreement, as defined through clinical expert consensus. Wide error distributions were again attributed to patient ambulation and related signal artifact, as well as possible human error during various manual vital sign measurements.

Data from these pilot studies support that wearable technologies capture continuous vital sign measurements from hospitalized ambulatory patients with varying degrees of accuracy. Whether prespecified levels of agreement with a reference standard are met, signal artifact and other sources of error, such as human error, pose challenges to validation of wearable sensor technologies in real-world clinical settings. Although our comparison of the Vitaliti CVSM to a continuous invasive reference (in an ICU) met strict ISO prespecified limits of agreement [16], our study patients were necessarily restricted to their hospital bed while undergoing cNIBP monitoring with an arterial catheter. Hence, in this environment, excessive patient motion and human measurement did not pose major challenges. The compliance of the Vitaliti CVSM with rigorous ISO standards [16] in a complex ICU environment is nonetheless promising; further validation testing in ambulatory patients is required.

A few studies have also examined wearable biosensor user acceptance from the patient perspective. In the pilot study of the ViSi Mobile and HealthPatch systems by Weenk et al [11], user experience was captured through semistructured interviews after patients had worn these devices for 2 to 3 consecutive days. Thematic content analysis revealed largely positive experiences, with most patients reporting that the monitoring devices were reassuring for them because nurses could monitor them from a distance [11]. Most also felt that these sensors did not encumber their personal care activities (eg, dressing, bathing). Good device adhesive and small sensor size were also noted as factors that were important to patients with respect to wearability of several device components [11]. Wearability of the Visi Mobile system was reported by several patients to be impacted negatively by the size and weight of the wristwatch component, numerous cables, as well as short battery life [11].

In a reactive post hoc analysis report, Harsha et al [24] examined challenges with implementing continuous oximetry monitoring in the VItal siGns monItoring with continuous puLse oximetry And wireless cliNiCian notification aftEr surgery (VIGILANCE) study (n= 2049), a randomized controlled trial testing the effectiveness of the Nellcor Oxinet III system (Covidien, Mansfield, MA) for continuous pulse oximetry (CPOX) monitoring on the incidence of postoperative respiratory complications among noncardiac surgery patients. VIGILANCE investigators found that 10.68% of patients withdrew from CPOX monitoring before intervention completion. Analysis of trial case report forms found a number of reasons for patient

nonadherence, including obtrusiveness of the CPOX cables, an uncomfortable $SpO_2$ probe, restrictions to ambulation, and device-related agitation of carpal tunnel syndrome [24].

In contrast to the usability assessments by Weenk et al [11] and Harsha et al [24], our examination of patient acceptance of the Vitaliti CVSM was in the context of a controlled measurement study, rather than in the context of live clinical system deployment where patients were ambulating on hospital wards and engaged in personal care activities. Although context differed, our results corroborate that patients value unobtrusive and comfortable sensor components. Our results also corroborate that patients are impacted negatively by technology features that are experienced as cumbersome or that require continual repositioning or reapplication.

### Strengths

Strengths of this study include the validation of the Vitaliti CVSM in the context of ISO standards for cNIBP measurement [16], as well as rigorous methods and planned approaches to post hoc signal analyses in order to ensure high data quality. Conduct of this study in a complex ICU setting also required the interconnection of numerous pieces of technology with operational independence to achieve time-matched cNIBP data sets for comparison of the Vitaliti CVSM with a gold-standard arterial catheter reference. An additional strength, therefore, was our plan to oversample to compensate for anticipated data losses due to technical problems and inevitable changes to patient hemodynamic status in an ICU setting.

### Limitations

Although the ISO standard [16] is rigorous from a measurement perspective, a limitation it imposed was the restriction of study participants to static (seated) and supine positions. Patients were also confined to bed while undergoing invasive cNIBP monitoring, thereby limiting the generalizability of our results to nonambulatory environments. It should also be noted that, in a complex ICU setting, this investigation could not span the full duration of the calibration period for the Vitaliti CVSM, which is 24 hours (Multimedia Appendix 1). Rather, this study was limited to evaluation of the device during a very short validation period after calibration (ie, that commenced within 2 minutes after calibration and that lasted for a short duration of time). For practical reasons within an acute ICU setting, we had to set up our equipment and take our BP measurements on each participant as expediently as possible to minimize disruption to nursing and medical care by virtue of the presence of our study team and related equipment. This timing precluded possible BP variations from calibration values that may otherwise be observed during a full 24-hour calibration period. Hence, results of this study cannot be applied to the BP accuracy of the Vitaliti CVSM over a 24-hour period with BP variations that may be normal, including individual readings that may vary considerably from calibration values. Future research should incorporate evaluation over the full calibration period of the device.

Our results for patients in the static and supine positions are within the clinically allowable tolerances for accuracy according to ISO. As such, claims regarding the accuracy of Vitaliti cNIBP can only be made at this point in the context of accuracy requirements set forth by Health Canada and the FDA. More research will be required to examine the accuracy of Vitaliti cNIBP measurements according to international standards, such as those set forth by the European Society of Hypertension [23] and the British Hypertension Society [25]. Examination of the accuracy of the Vitaliti cNIBP measurement according to these standards was not within the scope of this study.

Finally, patients enrolled in this study were postsurgical cardiac ICU patients given the requirement to compare the Vitalilti CVSM to an invasive gold-standard arterial catheter. The hemodynamic profile of patients in the postsurgical cardiac ICU typically features a greater degree of variability than seen in other populations in the early postoperative period. Moreover, some patients may experience atrial fibrillation following cardiac surgery. It should therefore be recognized that the included sample is not representative of all postsurgical patient populations—particularly those patients who undergo noncardiac and same day surgeries.

### Conclusions

Wearable RAM technologies that enable continuous acquisition of physiologic data from biosensors have the potential to transform postoperative care. This study found that one such wearable technology, Cloud DX's Vitaliti CVSM, demonstrated cNIBP measurement in compliance with ISO 81060-2:2018 standards [16] in the context of evaluation that commenced within 2 minutes of device calibration; this device was also well-received by patients in a postsurgical ICU setting. Future studies will examine the accuracy of the Vitaliti CVSM in ambulatory contexts for both cardiac and noncardiac surgery patients, with attention to assessment of the impact of excessive patient motion on data artifacts and signal quality. The Vitaliti CVSM will also be evaluated longitudinally as part of a postoperative remote patient monitoring solution both in hospital and while patients are recovering at home for up to 30 days following surgery. This work will feature intensive focus on the use of derived vital metrics and high-fidelity physiological data collected with the Vitaliti CVSM in order to develop predictive models with machine learning. The aim of these predictive models will be to identify early signs of postoperative complications in order to facilitate timely clinical interventions.

## Acknowledgments

Funding was received from the 2017/18 Federal Development Fund of Ontario Health Ecosphere competition for verification testing of Vitaliti. This pilot test was approved by the Hamilton Research Ethics Board (#4439). MHM holds the Heart and Stroke Foundation/Michael G DeGroote Endowed Chair in Cardiovascular Nursing.

## Authors' Contributions

MHM is the principal investigator of this study and contributed toward conceptualization, study design, protocol writing, validation testing, data analysis, and manuscript writing. ND, SY, CO, MB, PH, TS, and PJD contributed toward study design, validation testing, data collection, data analysis, and manuscript writing. AG contributed toward study coordination, protocol development, validation testing, and data management. KS, KG, VH, and JV contributed toward protocol development and data management. PB and JO contributed toward clinical site management and study recruitment. EBC, ED, AL, RW, MM, FKB, SLC, DC, and JET contributed toward signal analyses, quantitative validation analyses, and manuscript revisions. EP, JP, and MG contributed toward qualitative data analyses and manuscript revisions.

## Conflicts of Interest

MHM and PJD are members of a research group that does not accept honorariums nor other payments from industry for personal financial gain. They do accept honorariums or payments from industry to support research endeavors and costs to participate in meetings. On the basis of study questions PJD has originated and grants he has written, he has received grants from Abbott Diagnostics, Boehringer Ingelheim, Covidien, Octapharma, Philips Healthcare, Roche Diagnostics, and Stryker. PJD has participated in a consultancy advisory board meeting for Boehringer Ingelheim. EBC has received grants from Roche Diagnostics, BMS-Pfizer, and Bayer. DC has received fees from Roche Diagnostics and BMS/Pfizer.

Multimedia Appendix 1
Vitaliti continuous vital signs monitor (CVSM): device features and clinical workflow.
[DOCX File , 38 KB - mhealth_v10i2e24916_app1.docx ]

Multimedia Appendix 2
Vitaliti Patient Feedback Questionnaire.
[DOCX File , 500 KB - mhealth_v10i2e24916_app2.docx ]

## References


1. Aseni P, Orsenigo S, Storti E, Pulici M, Arlati S. Current concepts of perioperative monitoring in high-risk surgical patients: a review. Patient Saf Surg 2019 Oct;13:32 [FREE Full text] [doi: 10.1186/s13037-019-0213-5] [Medline: 31660064]
2. McGillion MH, Duceppe E, Allan K, Marcucci M, Yang S, Johnson AP, PROTECT Network Investigators. Postoperative remote automated monitoring: need for and state of the science. Can J Cardiol 2018 Jul;34(7):850-862 [FREE Full text] [doi: 10.1016/j.cjca.2018.04.021] [Medline: 29960614]
3. Efthymiou CA, O'Regan DJ. Postdischarge complications: what exactly happens when the patient goes home? Interact Cardiovasc Thorac Surg 2011 Feb;12(2):130-134. [doi: 10.1510/icvts.2010.249474] [Medline: 21123196]
4. Leuvan CHV, Mitchell I. Missed opportunities? An observational study of vital sign measurements. Crit Care Resusc 2008 Jun;10(2):111-115. [Medline: 18522524]
5. McGain F, Cretikos MA, Jones D, Van Dyk S, Buist MD, Opdam H, et al. Documentation of clinical review and vital signs after major surgery. Med J Aust 2008 Oct 06;189(7):380-383. [doi: 10.5694/j.1326-5377.2008.tb02083.x] [Medline: 18837681]
6. Sun Z, Sessler DI, Dalton JE, Devereaux PJ, Shahinyan A, Naylor AJ, et al. Postoperative hypoxemia is common and persistent: a prospective blinded observational study. Anesth Analg 2015 Sep;121(3):709-715 [FREE Full text] [doi: 10.1213/ANE.0000000000000836] [Medline: 26287299]
7. Sessler DI, Saugel B. Beyond 'failure to rescue': the time has come for continuous ward monitoring. Br J Anaesth 2019 Mar;122(3):304-306 [FREE Full text] [doi: 10.1016/j.bja.2018.12.003] [Medline: 30770047]
8. Turan A, Chang C, Cohen B, Saasouh W, Essber H, Yang D, et al. Incidence, severity, and detection of blood pressure perturbations after abdominal surgery: a prospective blinded observational study. Anesthesiology 2019 Apr;130(4):550-559 [FREE Full text] [doi: 10.1097/ALN.0000000000002626] [Medline: 30875354]
9. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, et al. Wearable sensors to improve detection of patient deterioration. Expert Rev Med Devices 2019 Feb;16(2):145-154. [doi: 10.1080/17434440.2019.1563480] [Medline: 30580650]
10. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019 May 30;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
11. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
12. Downey C, Ng S, Jayne D, Wong D. Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: a clinical validation study from the TRaCINg trial. BMJ Open 2019 Aug 15;9(8):e031150 [FREE Full text] [doi: 10.1136/bmjopen-2019-031150] [Medline: 31420399]

13. Ogedegbe G, Pickering T. Principles and techniques of blood pressure measurement. Cardiology Clinics 2010 Nov;28(4):571-586 [FREE Full text] [doi: 10.1016/j.ccl.2010.07.006] [Medline: 20937442]
14. Sahu D, Bhaskaran M. Palpatory method of measuring diastolic blood pressure. J Anaesthesiol Clin Pharmacol 2010 Oct;26(4):528-530 [FREE Full text] [Medline: 21547184]
15. Meidert AS, Saugel B. Techniques for non-invasive monitoring of arterial blood pressure. Front. Med 2018 Jan 8;4:231 [FREE Full text] [doi: 10.3389/fmed.2017.00231] [Medline: 29359130]
16. ISO 81060-2:2018 - Non-invasive sphygmomanometers - Part 2: Clinical investigation of intermittent automated measurement type. International Organization for Standardization. URL: https://www.iso.org/standard/73339.html [accessed 2022-01-04]
17. Vitaliti. CloudDX. URL: https://www.clouddx.com/#/vitaliti [accessed 2021-04-21]
18. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 2016 Jul 02;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 1050165]
19. MediCollector. URL: https://www.medicollector.com/about-medicollector-company-llc.html [accessed 2022-01-04]
20. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff. US Food and Drug Administration. 2016 Feb. URL: http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm259760.pdf [accessed 2022-01-04]
21. Vankipuram M, McMahon S, Fleury J. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. AMIA Annu Symp Proc 2012;2012:931-939 [FREE Full text] [Medline: 23304368]
22. Chen W, Kobayashi T, Ichikawa S, Takeuchi Y, Togawa T. Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. Med. Biol. Eng. Comput 2000 Sep;38(5):569-574 [FREE Full text] [doi: 10.1007/bf02345755]
23. O'Brien E, Atkins N, Stergiou G, Karpettas N, Parati G, Asmar R, Working Group on Blood Pressure Monitoring of the European Society of Hypertension. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. Blood Press Monit 2010 Feb;15(1):23-38 [FREE Full text] [doi: 10.1097/MBP.0b013e3283360e98] [Medline: 20110786]
24. Harsha P, Paul JE, Chong MA, Buckley N, Tidy A, Clarke A, et al. Challenges With Continuous Pulse Oximetry Monitoring and Wireless Clinician Notification Systems After Surgery: Reactive Analysis of a Randomized Controlled Trial. JMIR Med Inform 2019 Oct 28;7(4):e14603 [FREE Full text] [doi: 10.2196/14603] [Medline: 31661079]
25. O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, O'Malley K, et al. The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. J Hypertens 1990 Jul;8(7):607-619. [doi: 10.1097/00004872-199007000-00004] [Medline: 2168451]


## Abbreviations

**AAMI:** Association for the Advancement of Medical Instrumentation
**BP:** blood pressure
**cNIBP:** continuous non-invasive blood pressure
**CPOX:** continuous pulse oximetry
**CVSM:** continuous vital signs monitor
**DBP:** diastolic blood pressure
**ECG:** electrocardiogram
**FDA:** Food and Drug Administration
**ICU:** intensive care unit
**ISO:** International Organization for Standardization
**LED:** light emitting diode
**PHRI:** Population Health Research Institute
**RAM:** remote automated monitoring
**SBP:** systolic blood pressure
**$SpO_2$:** blood oxygen saturation
**VIGILANCE:** VItal siGns monItoring with continuous puLse oximetry And wireless cliNiCian notification aftEr surgery

*Edited by L Buis; submitted 09.10.20; peer-reviewed by S Rowland, H Mehdizadeh, C Son, E Scruth, J Goris; comments to author 22.11.20; revised version received 21.04.21; accepted 12.11.21; published 28.02.22.*

*Please cite as:*
*McGillion MH, Dvirnik N, Yang S, Belley-Côté E, Lamy A, Whitlock R, Marcucci M, Borges FK, Duceppe E, Ouellette C, Bird M, Carroll SL, Conen D, Tarride JE, Harsha P, Scott T, Good A, Gregus K, Sanchez K, Benoit P, Owen J, Harvey V, Peter E, Petch J, Vincent J, Graham M, Devereaux PJ*
*Continuous Noninvasive Remote Automated Blood Pressure Monitoring With Novel Wearable Technology: A Preliminary Validation Study*
*JMIR Mhealth Uhealth 2022;10(2):e24916*
*URL: https://mhealth.jmir.org/2022/2/e24916*
*doi:10.2196/24916*
*PMID:34876396*

Original Paper

# Digital Therapeutic Care Apps With Decision-Support Interventions for People With Low Back Pain in Germany: Cost-Effectiveness Analysis

Daniel Lewkowicz[1], MSc; Attila M Wohlbrandt[1], Dr-Ing; Erwin Bottinger[1,2], Prof Dr

[1]Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

[2]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

**Corresponding Author:**
Daniel Lewkowicz, MSc
Digital Health Center
Hasso Plattner Institute
University of Potsdam
Prof Dr Helmert Str 2-3
Potsdam, 14482
Germany
Phone: 49 (0)331 55094843
Email: daniel.lewkowicz@hpi.de

## Abstract

**Background:** Digital therapeutic care apps provide a new effective and scalable approach for people with nonspecific low back pain (LBP). Digital therapeutic care apps are also driven by personalized decision-support interventions that support the user in self-managing LBP, and may induce prolonged behavior change to reduce the frequency and intensity of pain episodes. However, these therapeutic apps are associated with high attrition rates, and the initial prescription cost is higher than that of face-to-face physiotherapy. In Germany, digital therapeutic care apps are now being reimbursed by statutory health insurance; however, price targets and cost-driving factors for the formation of the reimbursement rate remain unexplored.

**Objective:** The aim of this study was to evaluate the cost-effectiveness of a digital therapeutic care app compared to treatment as usual (TAU) in Germany. We further aimed to explore under which circumstances the reimbursement rate could be modified to consider value-based pricing.

**Methods:** We developed a state-transition Markov model based on a best-practice analysis of prior LBP-related decision-analytic models, and evaluated the cost utility of a digital therapeutic care app compared to TAU in Germany. Based on a 3-year time horizon, we simulated the incremental cost and quality-adjusted life years (QALYs) for people with nonacute LBP from the societal perspective. In the deterministic sensitivity and scenario analyses, we focused on diverging attrition rates and app cost to assess our model's robustness and conditions for changing the reimbursement rate. All costs are reported in Euro (€1=US $1.12).

**Results:** Our base case results indicated that the digital therapeutic care strategy led to an incremental cost of €121.59, but also generated 0.0221 additional QALYs compared to the TAU strategy, with an estimated incremental cost-effectiveness ratio (ICER) of €5486 per QALY. The sensitivity analysis revealed that the reimbursement rate and the capability of digital therapeutic care to prevent reoccurring LBP episodes have a significant impact on the ICER. At the same time, the other parameters remained unaffected and thus supported the robustness of our model. In the scenario analysis, the different model time horizons and attrition rates strongly influenced the economic outcome. Reducing the cost of the app to €99 per 3 months or decreasing the app's attrition rate resulted in digital therapeutic care being significantly less costly with more generated QALYs, and is thus considered to be the dominant strategy over TAU.

**Conclusions:** The current reimbursement rate for a digital therapeutic care app in the statutory health insurance can be considered a cost-effective measure compared to TAU. The app's attrition rate and effect on the patient's prolonged behavior change essentially influence the settlement of an appropriate reimbursement rate. Future value-based pricing targets should focus on additional outcome parameters besides pain intensity and functional disability by including attrition rates and the app's long-term effect on quality of life.

*(JMIR Mhealth Uhealth 2022;10(2):e35042)* doi:10.2196/35042

**KEYWORDS**

cost-utility analysis; low back pain; back pain; cost-effectiveness; Markov model; digital therapy; digital health app; mHealth; orthopedic; eHealth; mobile health; digital health; pain management; health apps

# Introduction

## Background

Low back pain (LBP) is the leading cause for worldwide years lost due to disability, and poses a major societal and economic burden with a point prevalence ranging between 9.4% and 37.1% and a lifetime prevalence of up to 85% [1-3]. In Germany, the cost of illness of LBP, referring to International Classification of Diseases-10 code M54 alone, was quantified to be €4.5 billion (~US $5 billion) in 2015 according to the Federal Office of Statistics, which represents 1.3% of the total national health care expenditure [4]. In various cost-of-illness studies, the overall costs of LBP are estimated to be even higher. For example, one study estimated average costs of €1322 (US $1474) per patient per year from a sample of 5650 LBP patients and extrapolated these costs to €48 billion (~US $53.5 billion) for the whole German population [5]. The major cost driver for health care systems lies in the predominantly high portion of indirect costs, which are estimated to range from 55% up to 87% for LBP patients and result from the vast majority of absences from work, leading to productivity losses [6,7]. Conversely, the minor part concerning direct costs quantifies the amount of health care resource utilization such as the number of primary care consultations or prescribed medications. Patients suffering from chronic LBP, which evolves in approximately 10%-15% of all LBP cases, are prone to occasion more than three-times higher costs than those incurred by patients with acute LBP [6,7]. Moreover, an increasing chronic pain grade (eg, as measured with the Von Korff pain scale) has been identified as the strongest predictor of high LPB costs [8]. According to the German Disease Management Guidelines, current treatment recommendations for nonspecific, nonacute LBP include remaining physically active, exercise and educational therapy, and psychosocial interventions. A recent systematic review investigated 15 distinct clinical practice guidelines for the management of nonspecific LBP in primary care from 15 different countries published between 2010 and 2017 [9]. All included guidelines provided recommendations for exercise therapy [9]. Moreover, pharmacological treatment should be reduced to a minimum and should only be prescribed as part of an overall therapeutic concept [9-11].

Digital therapeutic care apps are innovative new treatment programs with a variety of indication-specific video-based exercises and educational material accessible through a smartphone or a web-based app [12]. Recent research endeavors have shown that this multidisciplinary treatment modality can counteract the rising health care expenditure in multiple dimensions [12-14]. First, digital therapy apps provide a scalable and broadly accessible approach, enabling the treatment of LPB in rural areas and when the availability and workload of physiotherapists are limited [15]. Moreover, although stratified care is not yet implemented in routine care in Germany (eg, using the STarT-Back questionnaire), digital therapy apps enable early and immediate utilization for patients at high risk for developing a worsening or chronic condition [16,17]. Digital therapeutic care apps also support self-management and increase the patient's literacy through in-depth educational information [18]. Hence, these apps can further induce positive reinforcement through personalized decision support that entail motivational automated push notifications or tailored exercise recommendations based on personal preferences [19,20].

By contrast, previous retrospective studies of real-world user-generated data have shown that digital therapeutic care apps may imply low user retention and high early attrition rates [12]. The reasons are not entirely clear, but these drawbacks might mitigate the apps' overall health and economic benefits [18]. An analytical study conducted by the German Bertelsmann Foundation in 2012 estimated potential cost savings of approximately €3 billion per year based on resolving noncompliance and lack of therapy adherence of chronic LBP patients [21]. Therefore, the economic consequences of reimbursing and prescribing digital therapeutic care apps are unclear, and the tradeoff between positive impact and low engagement rates requires investigation within an economic evaluation. Considering the perpetual rise of health care expenditure in middle-to-high-income countries such as in Germany or the United States reaching 11.9% and 17.7% of the total gross domestic product, respectively, the appropriate allocation of health care resources and the management of cost-effective treatment modalities for LBP need to be analyzed and addressed more profoundly from an economic perspective [22,23]. Furthermore, the new Digital Healthcare Act in Germany allows apps with proven scientific evidence to be part of the reimbursement catalog of the statutory health insurance providers once the app is listed in the Digital Health Applications (DiGa) directory [24,25]. This initiative helps to reduce out-of-pocket costs for patients with LBP by covering digital therapeutic app costs and essentially removing one major barrier for upscaling the utilization of digital health apps. Moreover, a regulatory-driven decision process, including the selection of reimbursable therapy apps, will further increase the transparency for patients concerning which apps already have proven scientific evidence of being effective. Overall, these developments underline the urgency of evaluating the initial pricing of digital therapeutic care apps for people with LBP as well as for the statutory health insurance based on a long-term model-based economic evaluation.

## Objectives

In this cost-effectiveness analysis, we investigated the economic impact of a digital therapeutic care app as an alternative treatment approach to current treatment-as-usual (TAU) practices, including face-to-face (F2F) physiotherapy and concomitant pharmacological treatment. We extrapolated short-term evidence of the impact of a therapeutic care app

based on the effectiveness data of a single randomized controlled trial (RCT) [18]. Essentially, we aimed to simulate the divergent tradeoffs between digital therapeutic care and TAU: higher cost per app prescription and higher app attrition rates but also more self-management support to achieve behavior change, versus lower cost for F2F physiotherapy but also fewer possibilities to provide patients with reinforcing education material for coping with LBP. In addition, we simulated the reimbursement of a therapy app based on the current procedure in Germany in a best-practice decision-analytic Markov model and calculated the overall cost utility from a societal perspective. This study provides economic evidence that can inform other researchers and decision-makers, and further addresses the gap in health economic research by performing the first model-based economic evaluation of digital therapeutic care apps for people with nonspecific LBP.

## Methods

### Study Design

We constructed a decision-analytic, discrete-time Markov chain to simulate the long-term effects on treatment and cost outcomes of a digital therapeutic care app compared to TAU for patients with subacute and chronic nonspecific LBP. A Markov chain is a state-transition model that represents a stochastic process in which subsequent events occur with a predefined probability, which is also called the transition probability [26]. In health economic evaluations, these events are defined as health states that represent the patient's disease process over time. In the simulation, after each model cycle, the patient's health condition might change, and thus the patient cohort moves around to one or more subsequent health states given a specific transition probability [26].

Our economic evaluation was based on the effectiveness data from the currently only available RCT performed in Germany by Toelle et al [18], in which the impact of a therapy app was investigated without any additional interventions or tools. The authors investigated the effectiveness of an app over 12 weeks and compared it to six F2F physiotherapy sessions combined with online education material. Hence, our model narrows down treatment modalities and compares a therapeutic care app to TAU (ie, F2F individual physiotherapy), both accompanied by general practitioner (GP) and specialist consultations, concomitant pharmacological treatment, and diagnostic procedures. Particularly, we focus on the recurring pain episodes and implications of various treatment attrition rates on health care resource utilization in the primary care setting. However, we chose to exclude inpatient procedures, rehabilitation care, and injection therapy as further treatment modalities for two reasons. The first reason is that none of these interventions is recommended for our target population of LBP patients, and the second is that we lack data to populate the Markov chain and claim that by adding more complex treatment combinations and assumptions, the usefulness of our model would decrease. Moreover, we used quality-adjusted life year (QALY) utility scores to account for health-related quality of life (QoL) effects and express cost in 2021 Euro (€) values (€1=US $1.12). All utility scores and cost data were discounted with a discount factor of 3%. We derived all information regarding the amount of health care resource utilization from other clinical or cost-of-illness studies in the literature. Model simulation and calculations were performed in R using the "heemod: Health Economic Evaluations MODeling" package [27]. Our economic evaluation adheres to all items of the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [28]; also see Multimedia Appendix 1. More detailed explanations and calculations regarding all model input parameters can be found in Multimedia Appendix 2.

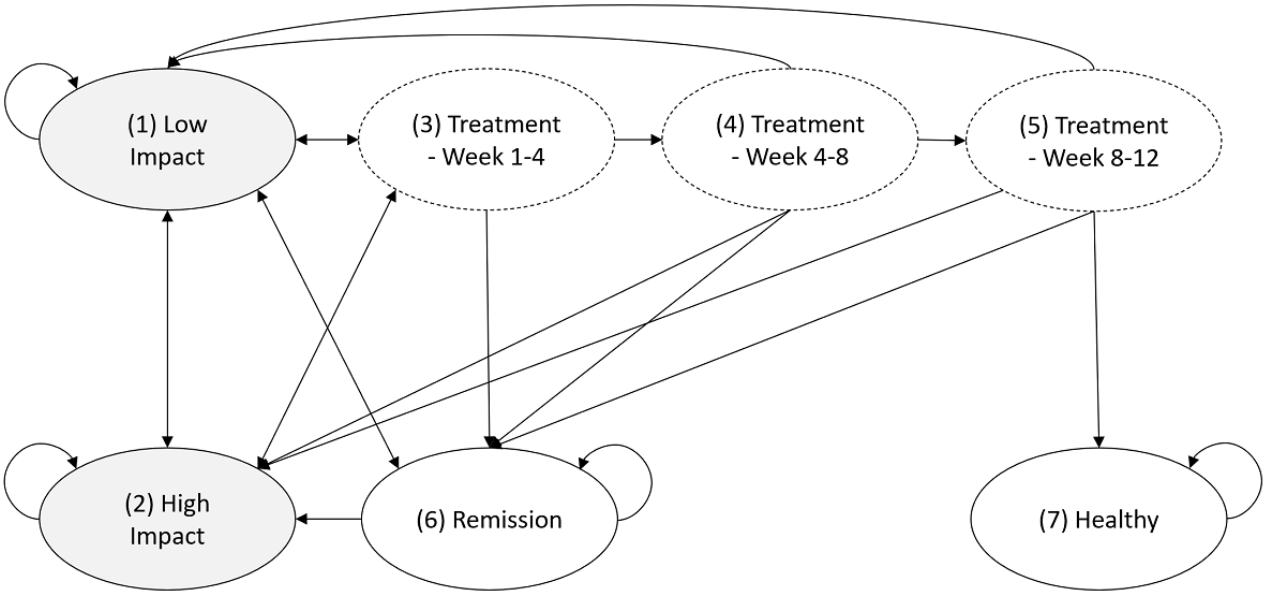
### Patient Population

We focused on a specific subcohort of nonspecific LBP patients who are insured via the German statutory health insurance and suffer from subacute or chronic LBP with a greater probability of recurring mild or heavy pain episodes. Particularly, we considered a hypothetical cohort of average 41-year-old participants with a mean BMI value of 24.4 in the intervention group, which was based on the cohort data from our reference RCT study by Toelle et al [18]. According to current clinical guidelines, this patient cohort should be treated with physiotherapeutic-based exercise programs and educational material for ergonomic and health-promoting behavior change, combined with temporary medication therapy for momentary pain reduction [9]. Conversely, we excluded acute LBP patients from this evaluation because, although remaining active is vital in this pain phase, extensive physiotherapy is not recommended for this subgroup [9]. Moreover, we did not intend to evaluate any treatment approach toward patients with existing red flags (eg, osteoporosis or any other severe comorbidities) or with yellow flags (eg, critical psychosocial factors). For instance, depressive disorders are a potential driving cause of the underlying pain condition and an essential predictor for future costs, but this would require additional psychotherapeutic care and involve the consideration of another treatment pathway, which was not the intention of this economic evaluation [6].

### Model Structure and Transition Probabilities

In our Markov chain model, displayed in Figure 1, we defined a total of seven discrete health states: (1) low impact, (2) high impact, (3) treatment weeks 1-4, (4) treatment weeks 4-8, (5) treatment weeks 8-12, (6) temporary remission, and (7) healthy.

**Figure 1.** Discrete health state-transition Markov chain with 7 health states.




We did not include "dead" as another absorbing state, since LBP is not affiliated with a higher death rate. All-cause mortality does not affect the overall outcome considering our middle-aged population in the defined time horizon of 3 years in the base case. We chose a cycle length of 4 weeks for our simulation for two reasons. First, a 4-week cycle length enables the consideration of three treatment phases, one for each month, which in turn represents the recommended 3-month treatment according to clinical guidelines. Second, recent studies of digital therapeutic care apps as well as the Toelle et al [18] study reported primary and secondary outcome measures based on monthly surveys. Hence, we found data to simulate monthly model cycles, but not enough data to increase the complexity such as by considering a weekly cycle length. We also applied the life-table method to account for corrected state membership counts [29]. Low-impact and high-impact health states (1) and (2) included untreated people with a respectively low or high impact on the QoL resulting from LBP. In the low-impact state, we classified people with low functional impairments and low limiting consequences. By contrast, people classified into the high-impact states were those suffering from high functional disability with moderately to severely limiting LBP. An intermediate impact state was excluded because we have not found sufficient data in the literature to further populate the model toward the relation to the other states. We decided to group the initial starting population based on the severity of symptoms. This approach enabled matching toward health states (1) and (2), which depict the corresponding level of the low and high impact of LBP. We found the Von Korff Graded Chronic Pain Scale to be the best choice for this matching, as this tool enables the classification of the severity of symptoms toward chronological pain grades [3]. For our model, we combined pain grades I and II to correspond to state (1) and the highest two pain grades III and IV to correspond to state (2) by adding up the correlated percentages for each category extracted from a German multiregional survey [3]. Thus, the hypothetical primary care cohort of 10,000 people was allocated to states (1), (2), and (6) with a starting distribution of 5320, 1120, and 3560 participants for each state, respectively [3,5].

Furthermore, we introduced treatment states in the form of three tunnel health states to represent a recommended treatment length of 12 weeks [30]. This approach was found to be highly useful when modeling time-to-treatment variations in a cost-utility analysis for internet-based cognitive behavioral therapy reported by Bauman et al [31]. We thus adopted the methodology of dedicated treatment states to optimize our model regarding the simulation of divergent monthly attrition rates. After each cycle, patients drop out of a treatment program for various reasons (eg, spontaneous remission phase with no present symptoms, stagnation of any health improvements, worsening condition with a fear of movement or pain, or simply a lack of motivation and time) [16,32]. In the RCT [18], for the base case, the attrition rates were 12.5% after both month one and month two in the digital therapeutic care app group, and were 6.5% after month one and 4.3% after month two in the control group. Conversely, in the digital therapeutic care strategy, patients continued treatment with a probability of 87.5% of proceeding to the next cycle, whereas this probability was 93.5% after month one and 95.7% after month two for the TAU strategy.

However, the discontinuation of either treatment strategy does not deduce any insights on the improvement or worsening effect on the patients' QoL. Therefore, we assumed an equal distribution of the remission rate of 50% among those who quit the program at each time point. According to a long-term cohort study in primary care performed in the United Kingdom, patients with LBP remain in a similar pain trajectory over time without any fluctuating patterns (ie, patients with mild pain will again with high probability experience mild pain in future episodes and vice versa) [33,34]. Following this argumentation, we inferred transition probabilities by proportionately splitting the remaining count of dropouts, meaning that 82.2% of nonremission dropouts will again transfer to the low-impact state group and 17.8% will transfer to the high-impact group [33].

Moreover, nonspecific LBP is characterized as a recurrent disease. Phases of pain and functional disability may frequently occur alternately to a temporal phase of relief [34]. The

"remission" state thereby serves as an intermediary simulation approach to include the temporary fluctuating and episodic nature of LBP with different intensities [31,34]. According to a German LBP survey with an adult sample size of 5650, 61.4% of participants experienced pain episodes repeatedly, which we utilized as the transition probability going from "remission" to either a low- or high-impact state [5]. In the case of a recurring episode, we further assumed that the GP or orthopedist will reevaluate the clinical findings through an imaging diagnostic procedure after the patient reenters the treatment pathway [35]. To capture the economic tradeoffs of recurrences and readmissions of patients to primary care, we chose a model time horizon of 3 years.

Finally, we assumed that only patients who underwent treatment and were fully engaged for 3 months can transfer to the "healthy" state [5]. We defined the "healthy" state as not reexperiencing LBP in the scope of this model's time horizon. Including a health state for both types of remissions, a spontaneous remission in state (6) and a long-lasting pain-free healthy condition in state (7), allowed us to consider the episodic nature of LBP and address those events with distinct transition probabilities, which are drawn from the literature [31]. In the underlying RCT [18], patients received access to the therapy app for 3 months, whereas the control condition only included six physiotherapy sessions. Hence, the 3-month app access duration with proven continuous high user retention affects the maintenance of performing exercises in the long term [18]. Moreover, decision-support interventions in the digital therapeutic care app have a high chance to induce positive behavior change by repeatedly informing and motivating the user through push notifications and in-app on-demand education material [36,37]. We thus derived that in the digital therapeutic care app strategy, the transition probability from the last treatment cycle to "healthy" is 5% higher. All model parameters regarding the transition probabilities and QoL utility scores are summarized in Table 1.

**Table 1.** Model input parameters: transition probabilities and quality of life (QoL) utility scores.

| Parameter | Base case[a] | | DSA[b] | | Reference |
|---|---|---|---|---|---|
| | DTC[c] | TAU[d] | Low[e] | High[f] | |
| **Key transition probabilities** | | | | | |
| Low to Low | 0.16 | 0.16 | —[g] | — | [33,34] |
| Low to High | 0.01 | 0.01 | — | — | [33,34] |
| Low to T[h]_W[i]1-4 | 0.75 | 0.75 | 0.60[j] | 0.90 | Assumption (75%) |
| Low to Remission | 0.03 | 0.03 | — | — | [33,34] |
| High to Low | 0.02 | 0.02 | — | — | [33,34] |
| High to High | 0.08 | 0.08 | — | — | [33,34] |
| High to T_W1-4 | 0.80 | 0.80 | 0.70[j] | 0.90 | Assumption (80%) |
| T_W1-4 to Low | 0.0514 | 0.0267 | — | — | [3,5,12,38] |
| T_W1-4 to High | 0.0111 | 0.0058 | — | — | [12,33] |
| T_W1-4 to T_W4-8 | 0.875 | 0.935 | — | — | [12,18] |
| T_W1-4 to Remission | 0.0625 | 0.0325 | 0.40[j] | 0.60 | Assumption (50%) |
| T_W4-8 to Low | 0.0514 | 0.0177 | — | — | [12,33] |
| T_W4-8 to High | 0.0111 | 0.0038 | — | — | [12,33] |
| T_W4-8 to T_W8-12 | 0.875 | 0.957 | — | — | [12,18] |
| T_W4-8 to Remission | 0.0625 | 0.0215 | 0.40[j] | 0.60 | Assumption (50%) |
| T_W8-12 to Low | 0.235 | 0.235 | — | — | [33,34] |
| T_W8-12 to High | 0.051 | 0.051 | — | — | [33,34] |
| T_W8-12 to Remission | 0.614 | 0.614 | 0.583[k] | 0.644 | [33,34] |
| T_W8-12 to Healthy | 0.10 | 0.05 | 0.095[k] | 0.105 | Assumption |
| Remission to Remission | 0.386 | 0.386 | 0.30[j] | 0.46 | [5] |
| Remission to Low | 0.505 | 0.505 | — | — | [3,5] |
| Remission to High | 0.109 | 0.109 | — | — | [3,5] |
| **QoL utility scores** | | | | | |
| Low impact | 0.655 | 0.655 | — | — | [18] |
| Higher pain | 0.610 | 0.610 | 0.5795[k] | 0.6405 | [39] |
| T_W1-4 | 0.655 | 0.655 | — | — | [18] |
| T_W4-8 | 0.699 | 0.717 | — | — | [18] |
| T_W8-12 | 0.748 | 0.729 | — | — | [18] |
| Remission | 0.806 | 0.806 | 0.7657[k] | 0.8463 | [39] |
| Healthy | 0.806 | 0.806 | 0.7657 | 0.8463 | [39] |

[a]Base case values are listed as the final calculated inputs used in the model. The raw numbers are provided in Multimedia Appendix 2.

[b]DSA: deterministic sensitivity analysis.

[c]DTC: digital therapeutic care.

[d]TAU: treatment as usual.

[e]Low: low-impact low back pain.

[f]High: high-impact low back pain.

[g]not applicable.

[h]T: treatment.

[i]W: week.

[j]Values are shown in raw numbers (ie, before the actual relative transition probability was calculated). All absolute transition probabilities are listed in Multimedia Appendix 2.

[k]Based on a –5% to +5% interval range.

### Measurement of Effectiveness: QoL

We retrieved the effectiveness data concerning the three treatment states from the reference RCT of Toelle et al [18]. In the intervention group, the authors found a significant improvement in the health-related QoL scores in both groups but no significant difference between groups, neither regarding pain levels nor QoL measurement [18]. However, after 12 weeks, patients in the app group had a higher mean QoL score compared with that of patients in the control group. For the QoL measurement, we adapted the QoL outcome results based on the Veterans RAND 12-Item Health Survey (VR-12) [18]. We gratefully retrieved the VR-12 data from the authors and calculated the single-utility index Veterans RAND 6-Item Health Survey (VR-6D) scores according to the approach proposed by Selim et al [40,41]. The detailed calculation steps for the VR-6D are listed in Multimedia Appendix 2. Notably, we applied QoL data for the remaining health states (2), (6), and (7) based on the Short Form 6-Item Health Survey (SF-6D) from another LBP study [39] because QoL data for these health states were not available from the reference study. We consider this methodological choice applicable because both strategies in our simulation do not differ regarding the QoL measure for these three health states. Consequently, the important difference in QoL outcome occurs only in three treatment states, (3), (4), and (5), for which the QoL data rely on the single study estimates. Even though interchanging different QoL metrics is not recommended in most cases, prior research has shown that the VR-6D and SF-6D are comparable indices with similar utility scores [40].

### Health Care Resource Utilization and Costs

For the evaluation of cost outcomes, we considered LBP-related utilization of health care resources and procedures as well as productivity losses that result from absenteeism from work. The direct cost components in our model include outpatient consultations, app subscription cost, F2F physiotherapy sessions, pharmacotherapy, and diagnostic procedures. Considering the patient visits in primary care, we multiplied the number of GP (“Hausarzt”) and specialist or orthopedist (“Facharzt”) consultations with provider-specific charges according to the German medical fee schedule for physicians (“Einheitlicher Bewertungsmaßstab”: EBM 13211 and 18211) [38,42]. Moreover, participants in the control group of Toelle et al’s [18] RCT received weekly, guideline-conforming individual physiotherapy for six sessions, each at least 20 minutes long. We calculated a charge of €21.11 (Position X0501 in the German “Heilmittelkatalog”) for each session and added the obligatory patient copayments of a one-time €10 prescription charge and an additional 10% own share for each physiotherapy session. Moreover, we extracted data on the utilization of pharmacotherapy and diagnostic procedures from the German cost-of-illness study performed alongside an RCT [6] instead of referring to the reference RCT study. We chose this approach because the reference study only reports on the outcome of the medication quantification scale, which is a value for a patient’s medication profile, and does not include data on the monthly frequency and dose of medication intake. Nonetheless, we adopted the finding that there was no significant between-group difference regarding medication intake [18]. We inflated the reported monthly mean cost of medication intake, including nonsteroidal antiinflammatory drugs as the most frequently used drugs, and monthly mean medical image diagnosing cost per patient to 2021 prices assuming an annual expenditure growth rate of 3% [6]. Since we did not consider patient treatment in inpatient care, surgical or rehabilitation costs were excluded. Finally, we utilized the price for the app’s 3-month subscription in Germany from the DiGa repository, in which another, yet similar in functionality, listed therapy app for people with LBP is currently listed and being reimbursed by statutory health insurance at a price level of €239.96 [43].

We estimated the indirect cost regarding the vocational outcome using the human capital approach. We multiplied the average hourly labor cost with productivity losses due to the LBP-related number of days of absences from work, assuming 21 working days per month and an average monthly gross wage of €3092 in 2020 in Germany [44]. Becker et al [6] reported that for one-third of patients, the mean frequency of short-term productivity losses was 8 days, whereas highly frequent utilizers with 5 or more days off work accounted for 98% of total absenteeism. Another German cost-of-illness study performed by Wenig et al [5] reported a mean value of 13.5 days of sick leave over 3 months. We thus assumed absenteeism to be on average the sum of 8 days for all states after each cycle in the model, which is also in alignment with a systematic review that reported a median duration of work absence of 5 to 28 days in the workplace samples [45]. However, we did not include productivity losses resulting from employees remaining at work with restricted operating activity, so-called presenteeism [5]. Furthermore, patients in the remission state do not cause any additional expenditure since we assumed that no follow-up pharmacotherapy nor primary care consultation is utilized in that health condition phase. All expenditure-related data that we used to populate the model as well as the mean monthly cost per patient and cycle are reported in Table 2.

**Table 2.** Model input: health care resource utilization and cost parameters (in Euro: €1=US $1.12).

| Parameter | Unit | Base case | Deterministic sensitivity analysis | | Reference |
|---|---|---|---|---|---|
| | | | Low | High | |
| **Direct cost** | | | | | |
| DTC[a] app | One-time access (for 3 months) | 239.96 | 99.96[b] | 299.96[b] | [43] |
| GP[c] | Consultation | 20.47 | —[d] | — | [5] |
| Orthopedic specialist | Consultation | 21.36 | — | — | [38,42] |
| Physiotherapist | Session | 21.11 | — | — | [38,42] |
| Physiotherapist | Cycle (6 session) | 149.33 | 102.88[e] | 288.65[e] | [38,42] |
| Pharmacotherapy | Per cycle | 16.81 | — | — | [6] |
| Diagnostic procedure | Per cycle | 29.24 | — | — | [6] |
| Indirect cost: productivity loss (absenteeism) | Daily wage | 147.24 | 132.52 | 161.96 | [44] |
| Discount rate | Annual | 0.03 | 0.00 | 0.05 | N/A |
| **Cost per cycle and per state** | | | | | |
| Low-impact LBP[f] | Cycle | 441.72 | 397.55 | 530.06 | See Multimedia Appendix 2 |
| High-impact LBP | Cycle | 588.96 | 471.17 | 706.75 | See Multimedia Appendix 2 |
| Treatment weeks 1-4 | Cycle | — | — | — | See Multimedia Appendix 2 |
| DTC | — | 475.08 | 335.08 | 535.08 | See Multimedia Appendix 2 |
| TAU[g] | — | 377.85 | 331.40 | 517.17 | See Multimedia Appendix 2 |
| Treatment weeks 4-8 | Cycle | 16.81 | — | — | See Multimedia Appendix 2 |
| Treatment weeks 8-12 | Cycle | 16.81 | — | — | See Multimedia Appendix 2 |

[a]DTC: digital therapeutic care.

[b]Manually set upper and lower bound values for price level of DTC app cost reimbursement.

[c]GP: general practitioner.

[d]not applicable.

[e]Assuming lower and upper bound values based on a divergent number of physiotherapy sessions: 4 and 12.

[f]LBP: low back pain.

[g]TAU: treatment as usual.

## Sensitivity and Scenario Analyses

We tested all previously mentioned assumptions in a comprehensive deterministic sensitivity analysis (DSA) to validate the robustness of our results. In the DSA, we manually set lower and upper bound values to increase the plausibility of our assumptions. Specifically, we chose divergent remission rates for all dropouts in the range of 40%-60% since it is not clear how many patients discontinue treatment because of sudden pain relief. Regarding the therapy app cost, we set the lower and upper bound values according to alternative price levels for the reimbursement rate of €99.96 and €299.96, respectively. We also varied the number of physiotherapy sessions in the TAU strategy from 4 to 12 to explore the influence of the F2F treatment cost on the overall outcome. For the remainder values, we used confidence intervals as reported in the respective studies or assumed 5% intervals. To address the use of different QoL indices, we tested the SD-6D index values with increased lower and upper bound values of 5%.

We refrained from performing a probabilistic sensitivity analysis (PSA) due to missing data on standard deviations or confidence intervals. Far-reaching assumptions would be necessary that would reduce the value and meaningfulness of a PSA.

Furthermore, we performed several scenario analyses. In scenario A, we simulated different time frames and extended our base case scenario to time horizons of 2 (A.1), 4 (A.2), and 5 (A.3) years. In scenario B, we investigated the impact of three alternative attrition rates (B.1-B.3 as described below) of digital therapeutic care app usage on the overall cost-effectiveness of the intervention compared to TAU. In the base case of our analysis, we used slightly higher attrition rate values in the digital therapeutic care app strategy as adopted from the findings of our reference RCT [18]. However, these numbers were found in a controlled clinical trial environment and do not represent real-world engagement and program dropout rates, which we previously explored in a review of different retrospective studies of digital therapeutic care apps [12].

We previously found divergent attrition rates of digital therapeutic care apps of up to 80% when retrospectively analyzing user databases of real-world app usage frequency [12,46]. Therefore, scenario B extends our base-case analysis by changing the transition probabilities for the digital therapeutic care app strategy in two ways. First, we assumed best-case attrition rates for the digital therapeutic care app strategy to be as low as that in the TAU strategy (B.1). Hence, we decreased attrition rates in the digital therapeutic care strategy to 6.5% after month one and to 4.3% after month two. As alternative worse-case scenarios, we assumed higher attrition rates for the digital therapeutic care strategy with equivalent values after months one and two (B.2: 14% each, B.3: 30% each) to explore the economic consequences when patients are reimbursed for the app but essentially stop using it shortly after.

## Results

In the base-case analysis of our simulation, the digital therapeutic care app strategy cost €121.59 more per patient but also generated additional 0.0221 QALYs compared to the TAU strategy. The incremental cost-effectiveness ratio (ICER) was €5486.05 per QALY. The total expenditures of both the digital therapeutic care and TAU strategies did not significantly differ and amounted to €2039 and €1998 per patient per year, respectively. The indirect cost aggregated to €1442 in the digital therapeutic care strategy and €1550 for TAU. The average QALY values per person and year aggregated to 0.697 in the digital therapeutic care app strategy and to 0.689 for TAU. According to data from a German cost-of-illness study performed by Becker et al [6] in 2010, we addressed 61% of the LBP-related direct cost in our model. By adding the indirect cost components, we addressed 81% of overall health care expenditure resulting from LBP [6]. In the digital therapeutic care strategy, a total of 4143 patients ended up in the “healthy” state, which is 1571 patients more compared with that in the TAU strategy, despite the higher attrition rate. In the TAU strategy, primary care consultations were substantially higher, with a mean number of consultations of 6151 and 5750 per year, respectively. The number of recurring prescriptions for physiotherapy was also 8% higher than the number of app prescriptions. In addition, fewer patients were located in the untreated states “low-impact” and “high-impact” LBP for the digital therapeutic care strategy, leading to a reduced indirect cost. However, the circumstance of the app cost being higher than six sessions of physiotherapy was superior and consequently the reason the digital therapeutic care app strategy was more costly than TAU. The results of the scenario analysis are summarized in Table 3.

The results of our base case and alternative scenarios in the cost-effectiveness plane are visualized in Figure 2. We also included manual threshold values of €10,000 and €20,000 per QALY to provide a better overview regarding the cost-efficiency and thus increase the comparability of the cost per QALY outcome.

**Table 3.** Results of the scenario analyses.

| Scenario | Incremental outcome[a] | | ICER[b] result |
|---|---|---|---|
| | Cost outcome[c] | Effect outcome | |
| A.1: Time horizon 2 years | 246.86 | 0.0098 | €25,189/QALY[d] |
| A.2: Time horizon 4 years | –99.23 | 0.0371 | DTC[e] dominant[f] |
| A.3: Time horizon 5 years | –381.80 | 0.0534 | DTC dominant |
| B.1: Equal attrition rates in both groups (6.5% and 4.3%)[g] | –288.58 | 0.0319 | DTC dominant |
| B.2: Higher attrition rates in DTC strategy (14% and 14%) | 213.47 | 0.0201 | €10,620/QALY |
| B.3: Higher attrition rates in the DTC strategy (30% and 30%) | –1263.62 | –0.0029 | TAU[h] dominant |

[a]Incremental outcome referring to the strategy: digital therapeutic care app intervention.

[b]ICER: incremental cost-effectiveness ratio.

[c]Presented in Euro (€1=US $1.12).

[d]QALY: quality-adjusted life year.

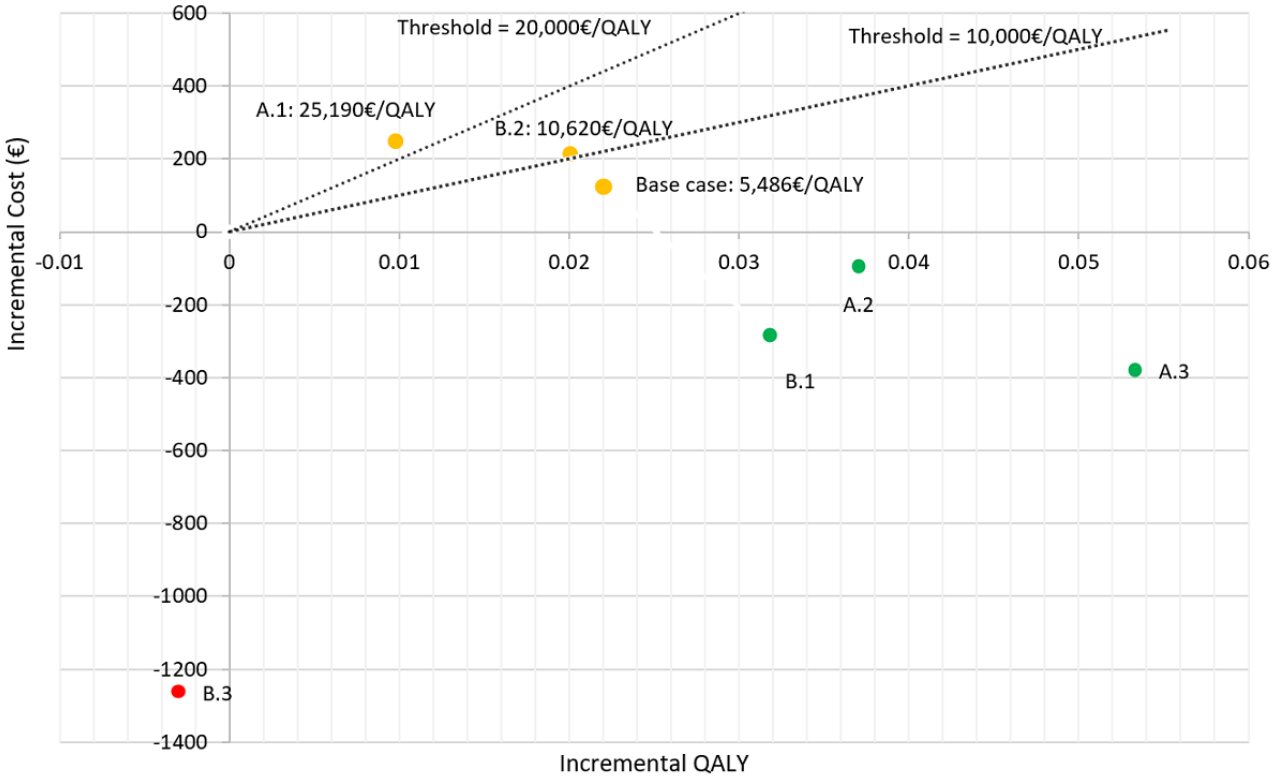
[e]DTC: digital therapeutic care.

[f]A dominant strategy: less costly and more generated QALYs.

[g]Monthly attrition rates: 6.5% when transferring from state (3) to (4) and 4.3% in the subsequent cycle when transferring from state (4) to (5).

[h]TAU: treatment as usual.

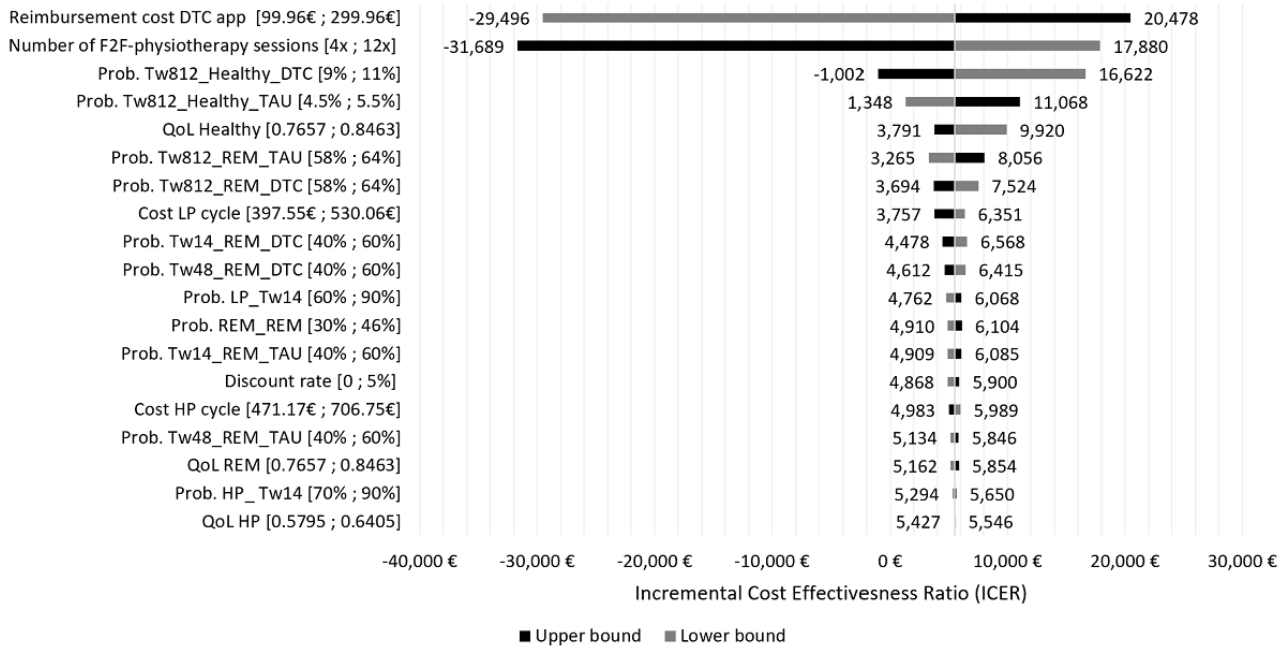
**Figure 2.** Cost-effectiveness plane, including base case and scenario analyses. The color code indicates when the digital therapeutic care strategy is dominant (green); both scenarios are comparable based on different quality-adjusted life year (QALY) thresholds (orange) or the treatment-as-usual strategy is dominant (red). €1=US $1.12.




The different model time horizons were revealed to have a substantial impact on the economic outcome. For the 2-year time horizon (A.1), the cost for one additional gained QALY increased significantly, whereas for the long-term observations (A2 and A3) the app intervention became less costly than TAU and thus the dominant strategy. Moreover, the divergent app attrition rates also showed a strong impact on the outcome. Considering equal attrition rates between app usage and TAU (B.1), digital therapeutic care became the dominant strategy. However, increasing the attrition rate of app users up to 30% (B.3) after each month resulted in TAU becoming the overall dominant strategy.

The results of our DSA are shown in a tornado diagram centered around the base case result of €5486/QALY in Figure 3. The reimbursement rate of the app and a diverging number of prescribed F2F physiotherapy sessions had the most considerable impact on the results. Increasing the app cost to around €99 per month also increased the cost per generated QALY up to €20,478, while decreasing the app cost to €99 for all 3 months, making digital therapeutic care dominant and significantly less costly than TAU. Similarly, increasing the number of F2F physiotherapy sessions from 6 to 12 within 3 months also made digital therapeutic care dominant over TAU. Except for these two outliers, our simulation results are robust, and parameter uncertainty did not significantly influence our findings.

**Figure 3.** Tornado diagram from the deterministic sensitivity analysis. DTC: digital therapeutic care; F2F: face to face; TAU: treatment as usual; QoL: quality of life; Tw: treatment week; HP: high impact; LP: low impact; REM: remission; Prob: probability of changing states.




## Discussion

### Principal Findings

The results of our analysis show that in the base-case analysis, reimbursing a digital therapeutic care app at a rate of €296 for a prescription duration of 3 months results in a cost of €5486 per additional gained QALY compared to TAU. Our decision-analytic model addressed 81% of total direct and indirect costs resulting from LBP and determined an average cost of around €2000 per patient per year in each group. Considering that we excluded inpatient and rehabilitation care, our cost estimate lies in the range of available data from different cost-of-illness studies in Germany with estimates ranging from €1322 and €3580 to €13,745 [5-7]. Our cost outcome value also confirmed our model assumption regarding the amount of absenteeism in the respective health cycles, and subsequently the amount of indirect cost predicted through our model, which averaged at around 70% of the total cost. Although our base-case approach following 8 days of sick leave per cycle is based on scientific findings, we had to make assumptions on when these days of sick leave occur in our model. Hence, we inferred that the majority of days off work occur when patients experience untreated LBP or at the very beginning of the treatment cycle. After treatment begins, medication use enables people to return to work regularly.

The pricing of digital therapeutic care as currently listed in the German DiGa directory is higher than that of six F2F physiotherapy sessions. Concerning commercial apps on the market, therapy app manufacturers offer access to their program for out-of-pocket payers at different price levels, which can vary significantly from around €10 to €99 per month. We have covered this broad range of app pricing within our DSA, in which we defined lower and upper bound values in the range of €99 to €296 for a 3-month treatment. Moreover, therapy apps also have much higher attrition rates, which made it overall unclear if digital therapeutic care is a cost-effective alternative to TAU. Nevertheless, in a digital therapeutic care program, the patients are supported by daily exercise and education material for 12 weeks, in contrast to 45 minutes once per week for 6 weeks of physiotherapy. This also includes personalized decision-support notifications that guide the user, reinforce daily achievements, and thus lead to long-term healthier behavior. It is notable that our explored incremental cost and incremental effect outcomes deviated only by €41 and 0.08 gained QALY per person per year, respectively. Looking at the cost components, the two strategies digital therapeutic care and TAU mainly differ regarding the initial treatment cycle cost, which is the cost of reimbursing a 3-month app prescription versus six F2F physiotherapy sessions. There was no significant between-group difference regarding the VR-6D index values in the QoL data we retrieved from the authors of the RCT [18]. The QoL improvements in the treatment and control group average from 0.671 at baseline to 0.748 after 3 months postintervention in the digital therapy group and from 0.639 to 0.729 in the control group, respectively. Hence, these minor differences ultimately lead to consequential comparable incremental outcomes in the simulation, making a difference only in the long term.

### Determinants for Reimbursement Price Predictions

The scenario and sensitivity analyses revealed the significant drivers and fundamental tradeoffs in our model: the cost of the app and the number of F2F physiotherapy sessions, the impact of digital therapeutic care on behavior change, and overall app attrition rates. We varied both the reimbursement cost of the therapy app and the number of physiotherapy sessions to account for the fact that digital therapeutic care could be equivalent to having up to 12 guided sessions. Both adjustments were found to have the most considerable influence on the ICER of digital therapeutic care. Similarly, the third and fourth bars in the tornado plot of Figure 3 ascertain that personalized health

assistance interventions are a decisive factor toward the cost-efficiency outcome of our analysis. We define a decisive factor as an element of the analysis that changes the ICER outcome of our simulation significantly. Since the tornado plot shows that this is the case for these two bars, referring to the transition probability going from state (5) to (7), the effect of decision-support interventions on prolonged behavior change has a substantial impact on the ICER of digital therapeutic care compared to TAU. Conversely, the more people improve their coping behavior with LBP (ie, reach health state 7) in our simulation, the more recurring episodes of LBP can be prevented.

From the perspective of our model, the actual recommended price for reimbursing the app is therefore directly dependent and derivable from two factors: (1) the retention and attrition rates while using the app, and (2) the effect of the implemented decision-support interventions to provoke behavior change and support long-term coping with LBP. If future trials can prove that apps achieve lower attrition rates in real-world usage as they currently do in a controlled clinical trial environment, our analysis confirms that digital therapeutic care becomes dominant over in-person physiotherapy. Moreover, if apps can support behavior change and the patient's self-management of LBP, our analysis confirms that the cost per generated QALY decreases for the digital therapeutic care strategy. Both factors significantly impact the formation and reevaluation of the reimbursement price and thus determine the amount of profit contribution for the app developers.

It is inevitable that the level of benefit and the patient's perception of the actual value of digital therapeutic care ultimately determine future reimbursements rates. Hence, more clinical effectiveness trials, including patient-reported outcome measures, are needed to generate more insights on these critical factors to further increase our model's usefulness and make actual price predictions for the statutory health insurance.

### Need for Further Cost-Effective Analyses Considering Different Patient Cohorts and Scenarios

In our model, we focused on a specific use case that implies the prescription of the therapy app to patients with subacute or chronic LBP as an alternative modality to in-person physiotherapy. However, there are other application areas that amplify the advantages of using an app and could potentially increase the cost-effectiveness of digital therapeutic care significantly. A cluster randomized trial has found that stratified care or immediate access to the app without prolonged waiting times on F2F physiotherapy is highly effective in preventing the worsening of LPB to a chronic condition. Thereby, early reduction of overall persistent pain levels could have a tremendous positive impact on the economic burden of LBP [16].

Moreover, we based our analysis on an RCT that has elaborated the efficiency of a multidisciplinary therapy app, including exercises, education material, and push notifications for people with LBP [18]. Considering the multidisciplinary capabilities of digital therapeutic care and the fact that multimodal offline rehabilitation programs are much more expensive in Germany, therapy apps could be a cost-effective alternative to these offline resource-extensive programs. It remains unclear if future effectiveness trials could show that therapy apps are as efficient as a multimodal rehabilitation program or if digital therapeutic care also provides synergistic effects as an add-on supporting modality. Subsequently, future studies should further explore how other therapeutic apps with different in-app features, dashboards, or even backends should be treated, and if cost-effectiveness analyses have to be conducted individually for each app based on the respective clinical effectiveness data.

Finally, our study relied on single study–based estimates with a small-sized and narrow cohort. In our simulation, we assumed a middle-aged cohort with above-average education and a medium BMI, thus possibly constraining the transferability of our results to a broader population. Coping with LBP by self-managing the digital therapeutic care program may be more challenging for other populations. For example, people with a higher BMI might experience more insecurities and fear-avoidance beliefs and drop out earlier, or less educated people might find it challenging to understand and adopt the necessity of behavior change [47]. Hence, further economic evaluations, including different patient characteristics, scenarios, and control groups, are required to make a profound conclusion on the cost-efficiency for including digital therapeutic care apps into the statutory health insurance.

### Transferability of Results to Other Countries

The lack of transferability also applies for implications toward the cost-efficiency of alternative health care systems in other countries. Our model's input data are specifically tailored toward the addition of digital therapeutic apps into the German statutory health insurance reimbursement catalog by drawing on insights from various German cost-of-illness studies. However, efficiency studies from other therapy apps for people with LBP have proven similar positive health effects on QoL and pain intensity, and might thus imply the same impact on the economic outcome [12]. The fact that our results imply an ICER of €5486 per QALY is highly promising for other countries, such as the UK National Institute for Health and Care Excellence in England with a cost-effectiveness threshold in the range of £20,000 to £30,000 (US $26,764 to US $40,146) per QALY [48].

### Comparison With Prior Work

Prior model-based economic evaluations concerning the long-term cost-efficiency of various treatment interventions and management of LBP were found to be sparse and with a poor standard of modeling [49,50]. A recent systematic review performed by Hall et al [49] identified a total of five studies that encompass a health economic decision model (ie, based on a decision tree or state transition for any treatment modality for LBP) [49]. The authors concluded a predominantly poor quality of modeling techniques, especially regarding the applied health states concerning a suboptimal representation of LBP health conditions and treatment pathways or inadequate time horizons and model cycle lengths [49]. Remarkably, concerning the three Markov model–based studies in their review, all constructed models included a total of three or four health states each to represent the respective treatment approach, which might entail oversimplification biases. Among the studies in their review, the authors favored a state-transition model in which the initial

health states served as a temporal classification of LBP (eg, “acute,” “subacute,” or “chronic” LBP, and “healthy”) [51]. However, the heterogeneity of pain severity and functional disability cannot sufficiently be reflected in a chronological arrangement of health states, while recurrent LBP episodes could result in a false state classification. Therefore, a modeling approach considering the severity of symptoms (eg, focusing on the level of pain) is recommended as the current best course of action [52].

In advance of our economic evaluation, we performed a snowball sampling search method to complement the view from the review of Hall et al [49] with any more recently published studies, allowing us to draw further insights on best-practice modeling techniques in this field. We searched the reference lists and used Google Scholar’s “cited by” function to find additional model-based studies concerning any treatment interventions for LBP. In total, we found another four publications, including three distinct Markov models [52-55]. First, Hall et al [52] performed a cost-utility analysis exploring the STarT Back stratified care model compared to usual care. Based on a six-health-state transition model, the authors concluded that stratified care is cost-effective for managing LBP over a 10-year horizon [52]. Second, Hermann et al [53] constructed a four-health-state transition model and investigated the cost-efficiency of 17 nonpharmacologic therapies compared to usual care over a 1-year time horizon. The authors updated their model in a subsequent publication by adding five additional trials with further alternative treatment modalities into their analysis [53,54]. Lastly, to complete the list of related work regarding available Markov model–based economic analyses of managing LBP, Olafsson et al [55] constructed a hybrid decision tree state-transition model to establish a lifetime treatment pathway model based on Swedish national registry data to extrapolate a mean lifetime total cost of €47,452 per LBP patient. We analyzed all additional studies more profoundly according to the individual model approaches, model conceptualization, and underlying techniques, which are summarized in Multimedia Appendix 3. Economic evaluations alongside clinical randomized controlled or observational trials are more prevalent and have been summarized for various LBP treatment modalities in numerous systematic reviews published in the previous 3 years alone [56-59].

### Limitations

Our model and analyses have several limitations and constraints. First, we did not include all cost components and treatment dimensions related to LBP. By addressing only 81% of LBP occurred costs and excluding inpatient and rehabilitation care, we may have caused over- or underestimation of costs and neglected the coherences and impact of digital therapeutic care on resource utilization of alternative treatment modalities. Although we specifically included the interventions recommended by German treatment guidelines, we excluded other minor yet prevalent applied interventions in outpatient care, such as injections therapy, because we would lack relevant data to integrate this alternative pathway into our model [35]. The high numbers of different options and treatment considerations make it challenging to develop a model that considers a broader patient cohort than we did in our simulation [50].

Furthermore, we did not perform a PSA because of the lack of information. For most of our input parameters, essential data such as standard deviations or confidence intervals were not available so that using a recommended beta or gamma distribution in the PSA was not feasible. Another limitation is the use of two different metrics, the SF-6D and VR-6D, as part of the QoL measurement of effectiveness. Although we do not expect our results to differ significantly because these two metrics provide comparable indices and the health states with the SF-6D utility values are equal for both strategies, this methodological choice is a limitation of our study. Finally, we emphasize that the QoL data of the treatment health states are taken from one RCT and are not derived from synthesis-based estimates such as a meta-analysis of QoL effectiveness studies, since more data are not yet currently available in the scientific literature. Although we tested all relevant uncertain parameters within our sensitivity analysis, more research is required on the cause and consequences of fluctuating LBP intensity as well as the reasons behind early and spontaneous treatment discontinuations.

### Conclusion

We developed a best-practice model for evaluating the cost-effectiveness of digital therapeutic care compared to TAU for people with LBP, and provided the first long-term economic evidence for reimbursing an app by the statutory health insurance in Germany. The current reimbursement cost set at €296.99 for a 3-month app prescription can be considered cost-effective compared to TAU with an ICER of €5486 per generated QALY. Future value-based price targets should focus on additional outcome parameters besides the effect on the QoL or reduction in pain intensity. Including the app’s attrition rate and the effect on the patient’s coping ability and behavior change induced by the app’s personalized assistance interventions will essentially influence the setting of a holistic value-based reimbursement price.

## Acknowledgments

This work has received funding from the European Union Horizon 2020 Research and Innovation Program under grant agreement number 826117 Smart4Health-Building a citizen-centered European electronic health record exchange for personalized health.

## Authors' Contributions

DL conceptualized the study, and performed the literature screening, model development, analytical analysis, interpretation of results, and writing of the manuscript. AW and EB contributed to refining all sections and critically editing the manuscript. All authors approved the submitted manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Items to include when reporting economic evaluations of health interventions (Consolidated Health Economic Evaluation Reporting Standards).
[PDF File (Adobe PDF File), 88 KB - mhealth_v10i2e35042_app1.pdf ]

Multimedia Appendix 2
Markov model input parameters.
[DOCX File , 25 KB - mhealth_v10i2e35042_app2.docx ]

Multimedia Appendix 3
Overview of related work.
[DOCX File , 19 KB - mhealth_v10i2e35042_app3.docx ]

## References


1. Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. Lancet 2012 Feb 04;379(9814):482-491. [doi: 10.1016/S0140-6736(11)60610-7] [Medline: 21982256]
2. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jun 24;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428] [Medline: 24665116]
3. Schmidt CO, Raspe H, Pfingsten M, Hasenbring M, Basler HD, Eich W, et al. Back pain in the German adult population: prevalence, severity, and sociodemographic correlates in a multiregional survey. Spine 2007 Aug 15;32(18):2005-2011. [doi: 10.1097/BRS.0b013e318133fad8] [Medline: 17700449]
4. Krankheitskosten: Deutschland, Jahre, Krankheitsdiagnosen (ICD-10, M54) im Jahr 2015. Statistisches Bundesamt. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Krankheitskosten/_inhalt.html [accessed 2021-10-01]
5. Wenig C, Schmidt C, Kohlmann T, Schweikert B. Costs of back pain in Germany. Eur J Pain 2009 Mar;13(3):280-286. [doi: 10.1016/j.ejpain.2008.04.005] [Medline: 18524652]
6. Becker A, Held H, Redaelli M, Strauch K, Chenot JF, Leonhardt C, et al. Low back pain in primary care: costs of care and prediction of future health care utilization. Spine 2010 Aug 15;35(18):1714-1720. [doi: 10.1097/brs.0b013e3181cd656f] [Medline: 21374895]
7. Brömme J, Mohokum M, Disch A, Marnitz U. Multidisciplinary outpatient care program vs. usual care: cost-benefit analysis in patients with chronic low back pain. Schmerz 2015 Apr 17;29(2):195-202. [doi: 10.1007/s00482-014-1508-1] [Medline: 25595921]
8. Engel C, von Korff M, Katon W. Back pain in primary care: predictors of high health-care costs. Pain 1996;65(2-3):197-204. [doi: 10.1016/0304-3959(95)00164-6] [Medline: 8826507]
9. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot J, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018 Nov;27(11):2791-2803. [doi: 10.1007/s00586-018-5673-2] [Medline: 29971708]
10. Chenot J, Greitemann B, Kladny B, Petzke F, Pfingsten M, Schorr S. Non-specific low back pain. Dtsch Arztebl Int 2017 Dec 25;114(51-52):883-890. [doi: 10.3238/arztebl.2017.0883] [Medline: 29321099]
11. Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). National Care Guideline (NVL) Non-specific low back pain, 2nd edition. Nationale Versorgunges Leitlinen. 2017. URL: https://tinyurl.com/2p8v2za9 [accessed 2022-01-28]
12. Lewkowicz D, Slosarek T, Wernicke S, Winne A, Wohlbrandt AM, Bottinger E. Digital therapeutic care and decision support interventions for people with low back pain: systematic review. JMIR Rehabil Assist Technol 2021 Nov 19;8(4):e26612 [FREE Full text] [doi: 10.2196/26612] [Medline: 34807837]
13. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]

14. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
15. Hou J, Yang R, Yang Y, Tang Y, Deng H, Chen Z, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e10201 [FREE Full text] [doi: 10.2196/10201] [Medline: 30785406]
16. Priebe JA, Haas KK, Moreno Sanchez LF, Schoefmann K, Utpadel-Fischler DA, Stockert P, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020;13:1823-1838. [doi: 10.2147/JPR.S260761] [Medline: 32765057]
17. Karstens S, Lang S, Saunders B. Patients' views on the implementation potential of a stratified treatment approach for low back pain in Germany: a qualitative study. Health Serv Insights 2020 Dec 07;13:1178632920977894 [FREE Full text] [doi: 10.1177/1178632920977894] [Medline: 33343197]
18. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34. [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
19. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1. [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
20. Von Korff M, Balderson B, Saunders K, Miglioretti D, Lin E, Berry S, et al. A trial of an activating intervention for chronic back pain in primary care and physical therapy settings. Pain 2005 Feb;113(3):323-330. [doi: 10.1016/j.pain.2004.11.007] [Medline: 15661440]
21. Behner P, Klink A, Visser S, Böcken J, Etgeton S. Effekte einer gesteigerten Therapietreue: Bessere Gesundheit und höhere Arbeitsproduktivität durch nachhaltige Änderung des Patientenverhaltens. Booz & Company, Bertelsmann Stiftung. 2012. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Presse/imported/downloads/xcms_bst_dms_36214_36215_2.pdf [accessed 2021-10-02]
22. Gesundheitsausgaben in Deutschland als Anteil am BIP und in Millionen € (absolut und je Einwohner - Gliederungsmerkmale: Jahre). Gesundheitsberichterstattung des Bundes. URL: https://www.gbe-bund.de/gbe/pkg_isgbe5.prc_menu_olap?p_uid=gastd&p_aid=20703791&p_sprache=D&p_help=0&p_indnr=522&p_indsp=&p_ityp=H&p_fid [accessed 2021-10-11]
23. National Health Expenditures 2020 Highlights. US Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/files/document/highlights.pdf [accessed 2021-12-10]
24. The fast-track process for digital health applications according to section 139e SGB V: A guide for manufacturers, service providers and users. Federal Institute for Drugs and Medical Devices (BfArM). URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html [accessed 2020-11-05]
25. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3(1):94. [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
26. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998 Apr;13(4):397-409. [doi: 10.2165/00019053-199813040-00003] [Medline: 10178664]
27. Filipović-Pierucci A, Zarca K, Durand-Zaleski I. Markov models for health economic evaluations: the R package heemod. arXiv. 2017. URL: http://arxiv.org/abs/1702.03252 [accessed 2021-02-28]
28. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health 2013 Mar;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
29. Naimark DMJ, Bott M, Krahn M. The half-cycle correction explained: two alternative pedagogical approaches. Med Decis Making 2008 Jun 30;28(5):706-712. [doi: 10.1177/0272989x08315241]
30. Sonnenberg FA, Beck JR. Markov models in medical decision making. Med Decis Making 2016 Jul 02;13(4):322-338. [doi: 10.1177/0272989x9301300409]
31. Baumann M, Stargardt T, Frey S. Cost-utility of internet-based cognitive behavioral therapy in unipolar depression: a Markov model simulation. Appl Health Econ Health Policy 2020 Aug 15;18(4):567-578 [FREE Full text] [doi: 10.1007/s40258-019-00551-x] [Medline: 32060822]
32. Vlaeyen J, Kole-Snijders A, Boeren R, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 1995 Sep;62(3):363-372. [doi: 10.1016/0304-3959(94)00279-N] [Medline: 8657437]
33. Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. BMJ Open 2013 Dec 11;3(12):e003838 [FREE Full text] [doi: 10.1136/bmjopen-2013-003838] [Medline: 24334157]
34. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? BMC Musculoskelet Disord 2016 May 21;17:220 [FREE Full text] [doi: 10.1186/s12891-016-1071-2] [Medline: 27209166]

35. Werber A, Schiltenwolf M. Treatment of lower back pain-the gap between guideline-based treatment and medical care reality. Healthcare (Basel) 2016 Jul 15;4(3):44 [FREE Full text] [doi: 10.3390/healthcare4030044] [Medline: 27417632]
36. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
37. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020 Sep 09;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
38. Einheitlicher Bewertungsmaßstab. Kassenärztliche Bundesvereinigung. 2005. URL: https://www.kbv.de/html/ebm.php [accessed 2021-06-02]
39. Cassidy JD, Côté P, Carroll LJ, Kristman V. Incidence and course of low back pain episodes in the general population. Spine 2005 Dec 15;30(24):2817-2823. [doi: 10.1097/01.brs.0000190448.69091.53] [Medline: 16371911]
40. Selim A, Rogers W, Qian S, Brazier J, Kazis LE. A preference-based measure of health: the VR-6D derived from the veterans RAND 12-Item Health Survey. Qual Life Res 2011 Oct;20(8):1337-1347. [doi: 10.1007/s11136-011-9866-y] [Medline: 21336657]
41. Buchholz M. The German version of the Veterans Rand Six Dimension (VR-6D) in the field of quality-of-life research: Translation, evaluation of methodological properties and application-oriented comparison. Dissertation. Universitat Greifswald. URL: https://nbn-resolving.org/urn:nbn:de:gbv:9-opus-38915 [accessed 2021-05-02]
42. Bundeseinheitliches Positionsnummernverzeichnis für Heilmittelleistungen. GKV-Spitzenverband. 2020. URL: https://www.gkv-datenaustausch.de/media/dokumente/leistungserbringer_1/sonstige_leistungserbringer/positionsnummernverzeichnisse_archiv/Heilmittel_20200424.pdf [accessed 2021-06-02]
43. DiGa-Verzeichnis: Information für Fachkreise. Federal Institute for Drugs and Medical Devices (BfArM). URL: https://diga.bfarm.de/de/verzeichnis/387/fachkreise [accessed 2021-08-12]
44. Höhe des durchschnittlichen Bruttogehalts im Monat je Arbeitnehmer in Deutschland von 1991 bis 2020. Statista. URL: https://de.statista.com/themen/293/durchschnittseinkommen/#dossier-chapter1 [accessed 2021-05-12]
45. Wynne-Jones G, Cowen J, Jordan JL, Uthman O, Main CJ, Glozier N, et al. Absence from work and return to work in people with back pain: a systematic review and meta-analysis. Occup Environ Med 2014 Jun 01;71(6):448-456 [FREE Full text] [doi: 10.1136/oemed-2013-101571] [Medline: 24186944]
46. Priebe JA, Utpadel-Fischler D, Toelle T. Less pain, better sleep? The effect of a multidisciplinary back pain app on sleep quality in individuals suffering from back pain - a secondary analysis of app user data. J Pain Res 2020;13:1121-1128. [doi: 10.2147/JPR.S232792] [Medline: 32547175]
47. Wettstein M, Eich W, Bieber C, Tesarz J. Pain intensity, disability, and quality of life in patients with chronic low back pain: does age matter? Pain Med 2019 Mar 01;20(3):464-475 [FREE Full text] [doi: 10.1093/pm/pny062] [Medline: 29701812]
48. Guide to the processes of technology appraisal. National Institute for Health and Clinical Excellence (NICE). 2018 Apr. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf [accessed 2021-10-04]
49. Hall JA, Konstantinou K, Lewis M, Oppong R, Ogollah R, Jowett S. Systematic review of decision analytic modelling in economic evaluations of low back pain and sciatica. Appl Health Econ Health Policy 2019 Aug 3;17(4):467-491. [doi: 10.1007/s40258-019-00471-w] [Medline: 30941658]
50. van Dongen JM, Ketheswaran J, Tordrup D, Ostelo R, Bertollini R, van Tulder M. Health economic evidence gaps and methodological constraints in low back pain and neck pain: results of the Research Agenda for Health Economic Evaluation (RAHEE) project. Best Pract Res Clin Rheumatol 2016 Dec;30(6):981-993. [doi: 10.1016/j.berh.2017.09.001] [Medline: 29103555]
51. Kim N, Yang B, Lee T, Kwon S. An economic analysis of usual care and acupuncture collaborative treatment on chronic low back pain: a Markov model decision analysis. BMC Complement Altern Med 2010 Nov 25;10(1):74 [FREE Full text] [doi: 10.1186/1472-6882-10-74] [Medline: 21106108]
52. Hall JA, Jowett S, Lewis M, Oppong R, Konstantinou K. The STarT Back stratified care model for nonspecific low back pain: a model-based evaluation of long-term cost-effectiveness. Pain 2021 Mar 01;162(3):702-710. [doi: 10.1097/j.pain.0000000000002057] [Medline: 32868748]
53. Herman P, McBain R, Broten N, Coulter I. Update of Markov Model on the cost-effectiveness of nonpharmacologic interventions for chronic low back pain compared to usual care. Spine 2020 Oct 01;45(19):1383-1385 [FREE Full text] [doi: 10.1097/BRS.0000000000003539] [Medline: 32516169]
54. Herman P, Lavelle T, Sorbero M, Hurwitz E, Coulter I. Are nonpharmacologic interventions for chronic low back pain more cost effective than usual care? Proof of concept results from a Markov model. Spine 2019 Oct 15;44(20):1456-1464 [FREE Full text] [doi: 10.1097/BRS.0000000000003097] [Medline: 31095119]
55. Olafsson G, Jonsson E, Fritzell P, Hägg O, Borgström F. A health economic lifetime treatment pathway model for low back pain in Sweden. J Med Econ 2017 Dec 11;20(12):1281-1289 [FREE Full text] [doi: 10.1080/13696998.2017.1372252] [Medline: 28840772]

56. Miyamoto GC, Lin CC, Cabral CMN, van Dongen JM, van Tulder MW. Cost-effectiveness of exercise therapy in the treatment of non-specific neck pain and low back pain: a systematic review with meta-analysis. Br J Sports Med 2019 Feb 20;53(3):172-181. [doi: 10.1136/bjsports-2017-098765] [Medline: 29678893]
57. Al Mazrou S, Elliott R, Knaggs R, AlAujan S. Cost-effectiveness of pain management services for chronic low back pain: a systematic review of published studies. BMC Health Serv Res 2020 Mar 12;20(1):194 [FREE Full text] [doi: 10.1186/s12913-020-5013-1] [Medline: 32164720]
58. Salathé CR, Melloh M, Crawford R, Scherrer S, Boos N, Elfering A. Treatment efficacy, clinical utility, and cost-effectiveness of multidisciplinary biopsychosocial rehabilitation treatments for persistent low back pain: a systematic review. Global Spine J 2018 Dec 19;8(8):872-886 [FREE Full text] [doi: 10.1177/2192568218765483] [Medline: 30560041]
59. Andronis L, Kinghorn P, Qiao S, Whitehurst DGT, Durrell S, McLeod H. Cost-effectiveness of non-invasive and non-pharmacological interventions for low back pain: a systematic literature review. Appl Health Econ Health Policy 2017 Apr 22;15(2):173-201. [doi: 10.1007/s40258-016-0268-8] [Medline: 27550240]


## Abbreviations

**CHEERS:** Consolidated Health Economic Evaluation Reporting Standards
**DiGa:** Digital Health Applications directory
**DSA:** deterministic sensitivity analysis
**F2F:** face to face
**GP:** general practitioner
**ICER:** incremental cost effectiveness ratio
**LBP:** low back pain
**PSA:** probabilistic sensitivity analysis
**QALY:** quality-adjusted life year
**QoL:** quality of life
**RCT:** randomized controlled trial
**SF-6D:** Short Form 6-Item Health Survey
**TAU:** treatment as usual
**VR-6D:** Veterans RAND 12-Item Health Survey
**VR-12D:** Veterans RAND 12-Item Health Survey


*Edited by G Eysenbach; submitted 18.11.21; peer-reviewed by L Sandal; comments to author 09.12.21; revised version received 17.12.21; accepted 29.12.21; published 07.02.22.*

*Please cite as:*
*Lewkowicz D, Wohlbrandt AM, Bottinger E*
*Digital Therapeutic Care Apps With Decision-Support Interventions for People With Low Back Pain in Germany: Cost-Effectiveness Analysis*
*JMIR Mhealth Uhealth 2022;10(2):e35042*
*URL: https://mhealth.jmir.org/2022/2/e35042*
*doi:10.2196/35042*
*PMID:35129454*

Corrigenda and Addenda

# Correction: A Novel mHealth Approach for a Patient-Centered Medication and Health Management System in Taiwan: Pilot Study

Wen-Ting Hsieh[1*], PhD; Yung-Cheng Su[2,3*], MD, MPH; Hsin-Lien Han[1], RN; Ming-Yuan Huang[1,4,5], MSc, MD

[1]Biomedical Research and Innovation Incubation Center, MacKay Memorial Hospital, Taipei, Taiwan

[2]Department of Emergency, Dalin Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Chiayi, Taiwan

[3]School of Medicine, Tzu Chi University, Hualien, Taiwan

[4]Department of Medicine, MacKay Medical College, New Taipei City, Taiwan

[5]Department of Emergency, MacKay Memorial Hospital, Taipei, Taiwan

[*]these authors contributed equally

**Corresponding Author:**
Ming-Yuan Huang, MSc, MD
Biomedical Research and Innovation Incubation Center
MacKay Memorial Hospital
9F, No 96, Chia Hsin Bldg (Second Bldg)
Sec 2, Zhongshan N Rd, Zhongshan District
Taipei, 10449
Taiwan
Phone: 886 225433535 ext 3065
Fax: 886 225239631
Email: jimmy.5684@mmh.org.tw

**Related Article:**

Correction of: https://mhealth.jmir.org/2018/7/e154/

***(JMIR Mhealth Uhealth 2022;10(2):e27292)*** doi:10.2196/27292

In "A Novel mHealth Approach for a Patient-Centered Medication and Health Management System in Taiwan: Pilot Study" (JMIR Mhealth Uhealth 2018;6(7):e154) one addition was made.

In the Methods section of the originally published paper, the subsection "Ethics Approval" has been added, containing the following statement:

> *This study was approved by the Institutional Review Board of Taipei Mackay Memorial Hospital, Taiwan (No.18MMHIS016e).*

The correction will appear in the online version of the paper on the JMIR Publications website on February 15, 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.

*Submitted 20.01.21; this is a non–peer-reviewed article;accepted 03.02.21; published 15.02.22.*

*Please cite as:*
*Hsieh WT, Su YC, Han HL, Huang MY*
*Correction: A Novel mHealth Approach for a Patient-Centered Medication and Health Management System in Taiwan: Pilot Study*
*JMIR Mhealth Uhealth 2022;10(2):e27292*
*URL: https://mhealth.jmir.org/2022/2/e27292*
*doi:10.2196/27292*
*PMID:35167485*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org