Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review

Karandeep Singh  
*University of Michigan Medical School*

Kaitlin Drouin  
*Brigham and Women’s Hospital*

Lisa P. Newmark  
*Partners HealthCare System*

See next page for additional authors

Follow this and additional works at: [http://escholarship.umassmed.edu/oapubs](http://escholarship.umassmed.edu/oapubs)

Part of the [Health Services Administration Commons](http://escholarship.umassmed.edu/oapubs/healthservicesadministration) and the [Telemedicine Commons](http://escholarship.umassmed.edu/oapubs/telemedicine)

Repository Citation
Singh, Karandeep; Drouin, Kaitlin; Newmark, Lisa P.; Filkins, Malina; Silvers, Elizabeth; Bain, Paul A.; Zulman, Donna M.; Lee, Jae-Ho; Rozenblum, Ronen; Pabo, Erika; Landman, Adam; Klinger, Elissa V.; and Bates, David W., "Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review" (2016). Open Access Articles. 3019.
[http://escholarship.umassmed.edu/oapubs/3019](http://escholarship.umassmed.edu/oapubs/3019)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review

Authors
Karandeep Singh, Kaitlin Drouin, Lisa P. Newmark, Malina Filkins, Elizabeth Silvers, Paul A. Bain, Donna M. Zulman, Jae-Ho Lee, Ronen Rozenblum, Erika Pabo, Adam Landman, Elissa V. Klinger, and David W. Bates

Keywords
chronic disease, mHealth, mobile apps, review, self-management

Rights and Permissions
© Karandeep Singh, Kaitlin Drouin, Lisa P Newmark, Malina Filkins, Elizabeth Silvers, Paul A Bain, Donna M Zulman, Jae-Ho Lee, Ronen Rozenblum, Erika Pabo, Adam Landman, Elissa V Klinger, David W Bates. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. Citation: JMIR Mhealth Uhealth. 2016 Dec 19;4(4):e136. Link to article on publisher's site
Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review

Karandeep Singh¹, MD, MMS; Kaitlin Drouin², MS, MA; Lisa P Newmark³, BA; Malina Filkins⁴, BA; Elizabeth Silvers⁵, BA; Paul A Bain⁶, MS, PhD; Donna M Zulman⁶,⁷, MD, MS; Jae-Ho Lee⁸, MD, PhD; Ronen Rozenblum⁵,⁹, MPH, PhD; Erika Pabo⁵,⁹, MD, MBA; Adam Landman⁵,¹⁰, MD, MS, MIS, MHS; Elissa V Klinger⁹, SM; David W Bates⁵,⁹,¹¹, MD, MSc

¹Departments of Learning Health Sciences and Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States
²Department of Pediatric Newborn Medicine, Brigham and Women’s Hospital, Boston, MA, United States
³Information Systems, Partners HealthCare System, Wellesley, MA, United States
⁴University of Massachusetts Medical School, Worcester, MA, United States
⁵Harvard Medical School, Boston, MA, United States
⁶Division of General Medical Disciplines, Stanford University School of Medicine, Stanford, CA, United States
⁷Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, CA, United States
⁸Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
⁹Division of General Internal Medicine, Brigham and Women’s Hospital, Boston, MA, United States
¹⁰Department of Emergency Medicine, Brigham and Women’s Hospital, Boston, MA, United States
¹¹Department of Health Policy and Management, Harvard TH Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Karandeep Singh, MD, MMSc
Departments of Learning Health Sciences and Internal Medicine
University of Michigan Medical School
1161H NIB
300 N Ingalls St
Ann Arbor, MI, 48109-5403
United States
Phone: 1 734 936 1649
Fax: 1 734 647 3914
Email: kdpsingh@umich.edu

Abstract

Background: Self-management is essential to caring for high-need, high-cost (HNHC) populations. Advances in mobile phone technology coupled with increased availability and adoption of health-focused mobile apps have made self-management more achievable, but the extent and quality of the literature supporting their use is not well defined.

Objective: The purpose of this review was to assess the breadth, quality, bias, and types of outcomes measured in the literature supporting the use of apps targeting HNHC populations.

Methods: Data sources included articles in PubMed and MEDLINE (National Center for Biotechnology Information), EMBASE (Elsevier), the Cochrane Central Register of Controlled Trials (EBSCO), Web of Science (Thomson Reuters), and the NTIS (National Technical Information Service) Bibliographic Database (EBSCO) published since 2008. We selected studies involving use of patient-facing iOS or Android mobile health apps. Extraction was performed by 1 reviewer; 40 randomly selected articles were evaluated by 2 reviewers to assess agreement.

Results: Our final analysis included 175 studies. The populations most commonly targeted by apps included patients with obesity, physical handicaps, diabetes, older age, and dementia. Only 30.3% (53/175) of the apps studied in the reviewed literature were identifiable and available to the public through app stores. Many of the studies were cross-sectional analyses (42.9%, 75/175), small (median number of participants=31, interquartile range 11.0-207.2, maximum 11,690), or performed by an app’s developers (61.1%, 107/175). Of the 175 studies, only 36 (20.6%, 36/175) studies evaluated a clinical outcome.
Conclusions: Most apps described in the literature could not be located on the iOS or Android app stores, and existing research does not robustly evaluate the potential of mobile apps. Whereas apps may be useful in patients with chronic conditions, data do not support this yet. Although we had 2-3 reviewers to screen and assess abstract eligibility, only 1 reviewer abstracted the data. This is one limitation of our study. With respect to the 40 articles (22.9%, 40/175) that were assigned to 2 reviewers (of which 3 articles were excluded), inter-rater agreement was significant on the majority of items (17 of 30) but fair-to-moderate on others.

(JMIR Mhealth Uhealth 2016;4(4):e136) doi:10.2196/mhealth.6445

KEYWORDS
review; mobile apps; mHealth; chronic disease; self-management

Introduction
Caring for high-need, high-cost (HNHC) populations represents a complex problem because these individuals often suffer from multiple chronic conditions, functional limitations, behavioral health problems, socioeconomic challenges, and inadequate coordination of care [1,2]. Nearly half of all US adults suffer from a chronic illness and this group accounts for a large share of health care costs [3]. Advances in mobile phone technology coupled with increased availability and adoption of mobile health apps have changed the landscape of self-management [4]. Data increasingly support the role of patient-facing health information technology tools in improving patient-centered care outcomes, health services efficiency, and health outcomes [5-7]. Community health centers and clinics that care for vulnerable populations overwhelmingly perceive mobile health technologies as an ideal tool to engage their patient populations in chronic disease management [8].

Although more than 165,000 mobile health apps are available on the iTunes (iOS) and Google Play (Android) app stores in the Unites States [9] and billions of dollars are being invested in digital health [10], it is not clear how many of these apps focus on patients with chronic conditions and how well the scientific evidence supports their effectiveness. Prior reviews of the literature evaluating the use of patient-facing health apps have been limited by a narrow scope. Reviews have focused on a single medical condition [11-13], on a single aspect of a broad group of apps (such as identifying target populations, behavioral functionalities, privacy policies, and expert involvement) [14-17], or have included only clinical trial–based evidence [9,18], which represents a minority of the ongoing research. A recent systematic review of apps targeting diabetes mellitus, cardiovascular disease, and lung disease found only 3 studies in which a chronic disease management app was used as an intervention and a clinical outcome was measured [19]. Another review focused on how apps can be leveraged by nonprofessional caregivers to care for patients [20].

Although clinical trial evidence supporting the use of apps is generally lacking, this finding may be explained by several factors. First, health apps are fairly new as a medium for engaging patients in comparison with other digital media; therefore, research supporting their use may be ongoing but not yet published. If that is the case, evidence may be found in “gray literature” such as conference proceedings that has not yet made its way to peer-reviewed journals. Second, app developers may be participating in and using research findings to market their apps, which may favor obtaining lower-quality evidence because it is less costly and potentially biased toward a favorable result. Third, it is possible that high-quality evidence exists but that prior reviews failed to uncover it because they focused too narrowly on a small set of disease areas. Given this set of limitations, a “scoping review” may better describe the extent and quality of the literature as well as evidence gaps in comparison with a traditional systematic review [21].

To address the need for a comprehensive assessment of health app evidence, we performed a scoping review in order to (1) assess the breadth of app coverage across HNHC populations, (2) characterize the quality of the published literature (including full-length journal articles and work presented at scientific conferences), (3) evaluate the possibility of biases due to conflicts of interest, and (4) evaluate the types of outcomes measured.

Methods

Data Sources and Searches
Studies that evaluated health-related apps for mobile devices were identified by searching PubMed and MEDLINE (National Center for Biotechnology Information), EMBASE (Elsevier), the Cochrane Central Register of Controlled Trials (EBSCO), Web of Science (including the Conference Proceedings Citation Indexes; Thomson Reuters), and the NTIS (National Technical Information Service) Bibliographic Database (EBSCO). The search was conducted between June 20, 2014, and July 14, 2014. The complete search strategy including search terms is available in Multimedia Appendix 1. Our search was designed to identify studies examining applications or software programs running on mobile devices such as mobile phones or tablets that are designed to address the health-related needs of specific HNHC populations. Populations included in the search were older adults (age ≥65 years); individuals with chronic conditions including coronary artery disease, congestive heart failure, hypertension, stroke, chronic obstructive pulmonary disease, cancer, diabetes mellitus, obesity, arthritis, chronic kidney disease, cirrhosis, organ transplantation, or chronic pain; the psychologically or mentally vulnerable who have been diagnosed with depression, bipolar disease, posttraumatic stress disorder, attention-deficit hyperactivity disorder, autism, substance-related disorders, dementia, cognitive impairment, developmental delays, or mental impairment; individuals with medication management issues (multiple medications); individuals with physical handicaps or disabilities, including the blind and deaf; and the socially vulnerable including those with low literacy or numeracy, limited English proficiency, minority status (Native
American, Hispanic, African American), low income or homelessness, or infection with human immunodeficiency virus. Appropriate controlled vocabulary terms were included when available (Medical Subject Headings and Emtree). The retrieval set was limited to articles published in 2008 or later; this start date was selected to coincide with when the iOS and the Android app stores were established. No language restrictions were applied, although non-English abstracts were excluded during title and abstract review (Figure 1). Articles not pertaining to native iOS and Android apps were excluded during the full manuscript review.

Figure 1. Article selection process.

Study Selection
All titles and abstracts were individually examined by 2 reviewers (KS and KD, or KS and LPN). Abstracts were included if they described original research written in the English language involving use of an iOS- or Android-based health-related patient-facing mobile app by study subjects. Patient-facing apps are those intended for use primarily by patients or their caregivers. We selected articles that described either iOS or Android apps because the 2 operating systems serve different demographics, with lower-income individuals, blacks, and younger adults preferring Android devices [22]. Articles describing apps focused on supportive technologies (eg, hearing or vision aids), communication technologies (eg, apps used to help autistic children communicate in school settings), or apps requiring a medical device (eg, an app to interact with artificial pancreas) were excluded. Study design was not a basis for exclusion. Full-length articles were obtained for all abstracts identified for inclusion by either reviewer. Certain included abstracts could not be linked to full-length articles.
manuscripts because they were associated with conference proceedings, including oral presentations or poster sessions; these abstracts were included despite the absence of a full-length manuscript as we wanted to capture gray literature in our review.

The full-length manuscripts and conference abstracts were evaluated by 2 reviewers (MF and ES) to confirm that they met the inclusion criteria. Articles identified for inclusion by both reviewers were selected for abstraction. Articles where the 2 reviewers disagreed were evaluated by a third reviewer (KD) to break ties.

**Data Extraction and Quality Assessment**

An abstraction survey tool was created to capture information about both the mobile app as described in the publication and the study itself, including the characteristics of the studied apps, quality of evidence, presence of conflict of interest, and types of outcomes evaluated. During a pilot phase, 8 study investigators each abstracted 3 articles using the tool (24 articles in total); changes were made based on feedback until there was consensus regarding the face validity of the tool.

Abstraction of the selected articles was then performed by 1 reviewer (MF or ES). A total of 40 randomly selected articles were evaluated by both reviewers to assess the level of agreement (Table 1).

Details regarding the abstracted items are presented in Multimedia Appendix 2. App engagement was assessed using a previously described framework [23]. We evaluated the following areas for each article:

- **General**
  - We captured information about the app studied, including its target population, platform, availability on the app store, and functionalities to support patient engagement.

- **Quality of the Evidence**
  - We ascertained factors that influence quality and generalizability of the article, including study design, enrollment, follow-up, role of the app in the context of the intervention, and inclusion of relevant patient populations.

- **Declaration of Conflicts**
  - We determined whether any members of the research team were developers of the app in question (or in a formal role supporting app development such as the advisory board) or whether the app developer directly funded the research. While a conflict of interest does not invalidate the results of a given study, literature written or funded by a company responsible for the product being researched is known to be systematically biased [24].

- **Outcomes Evaluated**
  - We evaluated the outcomes considered by each study and assessed their direction (ie, positive, neutral, or negative). Clinical outcomes were those directly related to patient care (eg, decreased hemoglobin \(A_1c\)). Safety or adverse event outcomes were those relevant to unintended negative consequences of an app. Usability outcomes were those describing an app’s ease of use—in some usability studies, multiple rounds of testing are performed, in which case the direction of the outcome was classified based on the final round of testing. Usage describes the amount of time users engaged with the app—this was not reported in a standard fashion and therefore we based the direction on the authors’ expectations, considering “sufficient usage” if observed usage matched expectations. Process outcomes refer to measures pertaining to actions taken in response to the app (eg, undergoing testing for hemoglobin \(A_1c\))—because the result of the action is not considered (eg, decreased hemoglobin \(A_1c\)), this is not a clinical outcome. A validation outcome was considered present when an app focused on measurement (eg, an app for assessing hepatic encephalopathy) was compared with a non–app-based clinical measure. We evaluated whether the app-based measure performed differently from a non–app-based clinical measure (eg, asterixis). If the article also used a gold standard test (eg, neuropsychiatric testing), we ascertained whether the app-based measure was better or worse than the non–app-based clinical measure. The user satisfaction outcome referred simply to whether users were satisfied with an app.

**Data Synthesis and Analysis**

Data from the reviewers were imported into R version 3.2.2 (R Foundation for Statistical Computing). Descriptive statistics were calculated and accompanied by a narrative summary.

**Results**

**Article Selection and Abstraction**

We identified 7301 titles and abstracts, of which 800 were identified for inclusion by either reviewer (Figure 1). Two reviewers evaluated the full-length manuscripts and 146 articles were selected by both reviewers for abstraction. Of the 90 articles identified for inclusion by only 1 reviewer, a third reviewer selected 37 for abstraction, resulting in a total of 183 articles being selected. During the abstraction process, 8 articles were identified as not meeting the inclusion criteria. After examination by a second reviewer, consensus was achieved on all 8 articles to exclude from the analysis. Thus, in total, 175 articles were abstracted (Multimedia Appendix 3).

Of the 40 articles (22.9%, 40/175) randomly selected for evaluation by both reviewers, 3 were excluded after further examination. As a result, 37 articles were abstracted by both reviewers, and the level of agreement was generally good, with some exceptions such as patient engagement and whether caregivers were included as subjects (Table 1).
Table 1. Level of agreement on items on the abstraction form.

<table>
<thead>
<tr>
<th>Question</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
</tr>
<tr>
<td>Who is the primary population(s) that would benefit from the app studied?</td>
<td>.79b</td>
</tr>
<tr>
<td>Which platform(s) is used by the app(s) mentioned in the study?</td>
<td>1.0b</td>
</tr>
<tr>
<td>Is the app(s) studied currently available on the iTunes or Google Play app store?</td>
<td>.85b</td>
</tr>
<tr>
<td>Based on the app's description in the article, how does it engage patients?</td>
<td>.26b</td>
</tr>
<tr>
<td>Did the app link to a medical device in the study (eg, glucometer)?</td>
<td>1.0</td>
</tr>
<tr>
<td>Did the app link to a consumer wearable device in the study?</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Quality of evidence</strong></td>
<td></td>
</tr>
<tr>
<td>What is the study design?</td>
<td>.62b</td>
</tr>
<tr>
<td>How many total subjects are enrolled in this study (including controls for controlled trials)?</td>
<td>.72</td>
</tr>
<tr>
<td>Is the app studied a standalone intervention (or are there multiple interventions concurrent with app use)?</td>
<td>.59</td>
</tr>
<tr>
<td>What was the average length of follow-up reported (in months)?</td>
<td>.54</td>
</tr>
<tr>
<td>Was this associated with a conference proceeding (abstract, poster, presentation, etc)?</td>
<td>.93</td>
</tr>
<tr>
<td>Does the study have a clinicaltrials.gov registration number?</td>
<td>1.0</td>
</tr>
<tr>
<td>Was at least one of the above vulnerable populations included as subjects in the study?</td>
<td>.49</td>
</tr>
<tr>
<td>Does the study include children as subjects (people under 18 years old)?</td>
<td>.85</td>
</tr>
<tr>
<td>Does the study include people aged 65 or older as subjects?</td>
<td>.79</td>
</tr>
<tr>
<td>Were caregivers for at least one of the above vulnerable populations included as subjects in the study?</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Conflict of interest</strong></td>
<td></td>
</tr>
<tr>
<td>Did the research team or their employer contribute to the design or development of the app?</td>
<td>.52</td>
</tr>
<tr>
<td>What is the source of external funding for this study?</td>
<td>.35b</td>
</tr>
<tr>
<td><strong>Outcomes evaluated</strong></td>
<td></td>
</tr>
<tr>
<td>Was a clinical outcome considered in this study?</td>
<td>.43</td>
</tr>
<tr>
<td>If yes, in what direction was the clinical outcome with use of the app?</td>
<td>.36</td>
</tr>
<tr>
<td>Was a safety or adverse event outcome (caused by the use of the app) considered in the study?</td>
<td>0c</td>
</tr>
<tr>
<td>Was a usability outcome considered in the study?</td>
<td>.53</td>
</tr>
<tr>
<td>Was a usage outcome considered in the study?</td>
<td>.71</td>
</tr>
<tr>
<td>If yes, in what direction was the usage outcome with use of the app?</td>
<td>.61</td>
</tr>
<tr>
<td>Was a process measure considered in this study?</td>
<td>.41</td>
</tr>
<tr>
<td>If yes, in what direction was the process measure with use of the app?</td>
<td>.39</td>
</tr>
<tr>
<td>Was a validation outcome considered in this study?</td>
<td>.80</td>
</tr>
<tr>
<td>If yes, in what direction was the validation outcome with use of the app?</td>
<td>.69</td>
</tr>
<tr>
<td>Was user satisfaction considered in this study?</td>
<td>.84</td>
</tr>
<tr>
<td>If yes, in what direction was the satisfaction outcome with use of the app?</td>
<td>.71</td>
</tr>
</tbody>
</table>

aSee Multimedia Appendix 2 for additional information regarding the questions.
bItems where reviewers could select multiple options. Only perfect agreement was considered agreement in the kappa calculation.
cThere was only 1 article evaluated by 2 reviewers in which 1 reviewer marked safety or adverse event outcome as being present.

Identification of Overlapping Research

Of the 175 selected articles, we found 15 sets of articles that assessed the same app, in some instances using different study designs, numbers of participants, or end points. For the purposes of our analysis, we considered each article as separate. We found 2 articles that evaluated the ActiveLifestyle app [25,26], AsthmaCare [27,28], EncephalApp [29,30], iMigraine [31,32], iStepLog [33,34], the Mayo Clinic app [35,36], Multiple Sclerosis Performance Test [37,38], My Meal Mate [39,40].
Ready–Steady [41,42], a cognitive stimulation app for alcoholics [43,44], USMART [45,46], ClinTouch [47,48], and a food addiction intervention [49,50]. We found 3 articles that evaluated a mobile application in the Women with Epilepsy: Pregnancy Outcomes and Deliveries (WEPOD) study [51-53] and 3 that evaluated the SaGAS 20/10 app [54,55].

Characteristics of the Mobile Apps

Of the 27 vulnerable populations targeted by the literature search, the groups most commonly targeted by apps included individuals with obesity, physical handicaps, diabetes, older age, and dementia or mild cognitive impairment (Table 2).

Table 2. Primary population that would benefit from the app studied.

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of articles (N=175), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity</td>
<td>24 (13.7)</td>
</tr>
<tr>
<td>Physical handicap or disability (including blindness or deafness)</td>
<td>19 (10.9)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>15 (8.6)</td>
</tr>
<tr>
<td>Older adults</td>
<td>15 (8.6)</td>
</tr>
<tr>
<td>Dementia or mild cognitive impairment</td>
<td>14 (8.0)</td>
</tr>
<tr>
<td>Cancer</td>
<td>11 (6.3)</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>10 (5.7)</td>
</tr>
<tr>
<td>Alcohol or drug abuse</td>
<td>7 (4.0)</td>
</tr>
<tr>
<td>Chronic pain</td>
<td>7 (4.0)</td>
</tr>
<tr>
<td>Depression</td>
<td>7 (4.0)</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>6 (3.4)</td>
</tr>
<tr>
<td>Schizophrenia or psychosis</td>
<td>5 (2.9)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>4 (2.3)</td>
</tr>
<tr>
<td>Stroke</td>
<td>4 (2.3)</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td>Posttraumatic stress disorder</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td>Developmentally delayed or mentally impaired</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>HIV(^b) or AIDS</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Attention-deficit hyperactivity disorder</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Low income or poor</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Low literacy or low numeracy</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Posttransplant</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Smoking</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>None of the above</td>
<td>38 (21.7)</td>
</tr>
</tbody>
</table>

\(^a\)These are not mutually exclusive categories. Articles may evaluate multiple apps and individual apps may target multiple populations.

\(^b\)HIV: human immunodeficiency virus.

Of the 175 selected articles, 60.6% (106/175) involved iOS apps, 32.0% (56/175) involved Android apps, and 7.4% (13/175) involved both. Reviewers evaluated the availability of these apps on both the iTunes (iOS) and Google Play (Android) app stores. Reviewers were unable to search for the app being studied in 40.0% (70/175) of the articles because the name of the app was not mentioned. Apps from an additional 29.7% (52/175) articles were searched and unable to be found on either app store. Among the articles where an app was found, 66% (35/53) were found on the iOS app store, 6% (3/53) on the Android app store, and 28% (15/53) on both. The ways in which apps engaged patients were assessed based on the functionality described in the articles (Table 3).
Table 3. How health apps engage patients.

<table>
<thead>
<tr>
<th>Type of engagement</th>
<th>Number of articles (N=175) n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records information</td>
<td>132 (75.4)</td>
</tr>
<tr>
<td>Provides guidance</td>
<td>64 (36.6)</td>
</tr>
<tr>
<td>Displays a patient’s health information</td>
<td>55 (31.4)</td>
</tr>
<tr>
<td>Reminds or alerts patients</td>
<td>45 (25.7)</td>
</tr>
<tr>
<td>Provides educational information</td>
<td>36 (20.6)</td>
</tr>
<tr>
<td>Enables data sharing with clinician</td>
<td>36 (20.6)</td>
</tr>
<tr>
<td>Enables data sharing with caregiver</td>
<td>15 (8.6)</td>
</tr>
<tr>
<td>Engages through social media</td>
<td>14 (8.0)</td>
</tr>
<tr>
<td>Not enough information to determine</td>
<td>10 (5.7)</td>
</tr>
<tr>
<td>None of the above</td>
<td>6 (3.4)</td>
</tr>
</tbody>
</table>

*a*These are not mutually exclusive categories. Apps may engage patients in multiple ways.

The most common functionalities were recording information, providing guidance, and displaying health information, and the least common were engaging with social media and enabling communication with family members. A total of 6 (3.4%, 6/175) articles described apps with the ability to connect with a medical device and 5 (2.9%, 5/175) described apps able to connect with a consumer wearable device.

**Quality of Evidence**

The method of dissemination involved full-text publications for 136 articles and conference proceedings (eg, oral or poster presentation) for 39 articles. Cross-sectional studies accounted for 42.9% (75/175) of studies (Table 4). Methodologies with lower bias were represented as follows: randomized controlled trials (10.3%, 18/175), nonrandomized controlled trials (2.9%, 5/175), and prospective cohort studies (21.7%, 38/175). The median number of participants in the studies was 31 (interquartile range, IQR, 11.0-207.2, maximum 11,690). The median length of follow-up for non–cross-sectional studies—weighted for the number of participants when articles involved multiple substudies—was 1.4 months (IQR 0.6-3, maximum 42.6).

Table 4. Study designs used in abstracted articles.

<table>
<thead>
<tr>
<th>Study design</th>
<th>Number of articles (N=175) n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional study</td>
<td>75 (42.9)</td>
</tr>
<tr>
<td>Prospective cohort study</td>
<td>38 (21.7)</td>
</tr>
<tr>
<td>Qualitative research</td>
<td>34 (19.4)</td>
</tr>
<tr>
<td>Before-after study</td>
<td>22 (12.6)</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>18 (10.3)</td>
</tr>
<tr>
<td>Nonrandomized controlled trial</td>
<td>5 (2.9)</td>
</tr>
<tr>
<td>Case report or case series</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td>Randomized trial with no control</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Interrupted time series</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Not enough information to determine</td>
<td>1 (0.6)</td>
</tr>
</tbody>
</table>

*a*These are not mutually exclusive categories. Articles may use multiple study designs or may describe multiple substudies.

Among 26 studies with a control arm, the app was the sole intervention in 21 (81%, 21/26) articles and just one part of a multipart intervention in 5 (19%, 5/26) articles. In the remaining 149 articles, all study subjects were exposed to the app. Only 7.4% (13/175) studies were registered on ClinicalTrials.gov. Of the 137 articles for which a specific population was identified that may benefit from the app (eg, patients with heart disease), 78.1% (107/137) included members of that particular population in the study; 13.1% (18/137) of studies focused on screening or prevention and therefore participants were healthy individuals; the remaining 8.8% (12/137) of studies did not include participants from the relevant population.

Of the 175 studies, 38 (21.7%, 38/175) studies included children, 53 (30.3%, 53/175) included adults aged 65 years or older, and 17 (9.7%, 17/175) included caregivers of HNHC patients.
Declaration of Conflicts

The authors of the 175 identified studies or their employer directly contributed to the design or development of the app in 107 (61.1%, 107/175) articles; among these, however, 5 (2.9%, 5/175) did not state this explicitly in the body of the paper. The authors of the identified studies were not involved in app development in 28 (16.0%, 28/175) articles, and in 40 (22.9%, 40/175) we were unable to confirm the presence or absence of involvement. Of the 175 studies, 61 (34.9%, 61/175) studies were funded by a government agency, 41 (23.4%, 41/175) by a nonprofit organization, 12 (6.9%, 12/175) by a for-profit company, and 6 (3.4%, 6/175) by a medical professional society; 5 (2.9%, 5/175) studies reported they had no external funding. No statement about the funding source was present in 82 (46.9%, 82/175) articles.

Types of Outcomes Evaluated

Among the 175 articles, 87 (49.7%, 87/175) articles evaluated user satisfaction, finding users to be generally satisfied in 74 (85%, 74/87), generally unsatisfied in 2 (2%, 2/87), and neutral in the remaining 11 (13%, 11/87) articles. A total of 74 (42.3%, 74/175) articles evaluated usability—often multiple cycles of testing were described, and the first cycle typically had worse performance than later cycles after modifications were made. A total of 61 (34.9%, 61/175) articles looked at usage, finding “sufficient” usage in 53 (87%, 53/61) and lower-than-expected use in 8 (13%, 8/61). Of the 175 articles, 56 (32.0%, 56/175) articles validated the measurement ability of apps in comparison with a clinical measure, finding the app to perform better than the clinical measure in 6 (11%, 6/56), worse than the clinical measure in 17 (30%, 17/56), and no different from the clinical measure in 34 (61%, 34/56) studies. A total of 40 (22.9%, 40/175) studies assessed a process measure (eg, increased administration of smoking cessation counseling), as opposed to a clinical outcome (eg, decreased rate of lung cancer). Of these, there were 35 (88%, 35/40) studies with improvement, 1 (2%, 1/40) with worsening, and 4 (10%, 4/40) with no change in the process outcome. A total of 36 (20.6%, 36/175) articles evaluated clinical outcomes, with 26 (72%, 26/36) demonstrating improvement in clinical outcomes and 10 (28%, 10/36) with no change. Only 9 (5.1%, 9/175) articles considered a safety or adverse event outcome caused by use of the app.

Discussion

Principal Findings

While there is optimism that mobile health apps may support the health of HNHC populations, existing research does not robustly evaluate this potential. Our review of the evidence supporting patient-facing mobile health apps identified a number of gaps in the current body of research. A few HNHC groups (older adults and people with obesity, physical handicaps, diabetes, and dementia) are more commonly studied, and we found less than 10 studies published for 20 of the 27 HNHC groups included in our review. The majority of apps studied were unavailable to consumers, the study designs were primarily cross-sectional, non–cross-sectional studies had a fairly short length of follow-up, and study sizes were small. In most cases, developers were often the ones evaluating the apps, sample sizes were small, funding sources were ambiguous, and clinical outcomes were evaluated in a minority of studies. Even among high-quality studies, drawing an inference about the usefulness of an app was frequently limited by intervention arms in which the app was a small piece of a much larger intervention.

Some of the methodological problems we identified such as small sample sizes and short length of follow-up could be addressed if apps incorporated the consent process and data collection into the apps’ functionality. Many of the studies used a traditional “in-person” consent process in order to enroll study subjects. While this may conform to the standards of traditional clinical research, using this approach may limit the number of subjects who can be enrolled and the length of follow-up. New methodological approaches that enable large-scale app outcomes research are needed [56]. Controlled trials where the consent process and data collection occur entirely in the context of a publicly available app may enable such work. The barrier to entry for integrating research into apps has been lowered by frameworks such as Apple ResearchKit, which was used to enroll 11,000 participants for a cardiovascular study in 24 hours [57].

Recommendations

On the basis of our findings, we make the following recommendations for researchers undertaking the study of mobile apps for the purposes of dissemination:

First, the researchers should consider evaluating apps in understudied HNHC groups to address the current imbalance in the body of research between HNHC groups. Second, reports should include the name of the app or intervention, so that literature about the app can be linked to it definitively; every effort should be made to include a bundle ID, permanent app store weblink, or other unique identifier to facilitate identification of the app. Third, researchers conducting interventional studies should consider the inclusion of both a control arm and an app-only intervention arm to make clearer the link between the app and the outcome. Fourth, studies should clearly state the nature of the relationship between the study contributors and the app developers; if the researchers are also the app developers, researchers should consider validating their work at an additional site supervised by a nondeveloper. Fifth, studies should clearly state the funding source or note if no external funding was used. Finally, researchers should report negative results.

In addition, funders will need to support additional evaluations of apps and should target evaluations that target clinically important outcomes and are large enough to deliver meaningful results. With newer enrollment approaches, it may be possible to enable much larger clinical trials, which may be feasible at low expense because much of the data usually collected may be extracted from existing electronic health records.

We used a robust multistage scoping review process involving 2 reviewers in most steps. We included gray literature in our analysis through a search of conference proceedings and did not limit our analysis to only high-quality evidence.
Limitations

Although 2 reviewers participated in the process of screening abstracts and full-length manuscripts for eligibility, data abstraction was carried out by only 1 reviewer for most studies. While the majority (17 of 30) of abstracted items had good agreement among the reviewers, interrater agreement was moderate (kappa .41-.6) for 7 items and fair (kappa .21-.40) for 6 questions, which limits the reliability of conclusions drawn from them. To put the kappa values into context, 2 reviewers agreed on ascertainment of a clinical outcome (kappa .43) on 31 out of 37 articles. In particular, the questions with the lowest interrater agreement involved determining whether safety or adverse event outcomes were considered, whether caregivers were included in the study, how the app mentioned in the study engaged patients, and the funding source. The low agreement levels for these questions may be attributed to the heterogeneity in the detail to which studies reported information about the apps and study conducted, which is partly due to the inclusion of conference abstracts in our analysis. The agreement was moderate when reviewers abstracted the types of outcomes measured in the studies. We attribute this partly to the breadth of populations that we considered because what constitutes a clinical outcome differs significantly between chronic conditions. Additionally, differentiating clinical outcomes from process outcomes carries some subjectivity and may introduce disagreements. We did revise our abstraction form based on a review of 24 articles during a pilot phase, but additional cycles of revision may have further improved interrater agreement.

We did not evaluate articles that were not in English, which limits our generalizability toward apps targeting non-English speakers. Finally, we conducted our literature search in 2014, which does not capture recent trends.

Recent Trends in the Literature

Recent studies of patient-facing apps have provided supporting evidence for the role of apps in several areas. In a 2016 randomized controlled crossover study of a mobile app focused on supporting drug intake and vital sign documentation, researchers found that patients who used the iPad app showed greater adherence for both medication intake and blood pressure measurement than a paper-based control group [58]. Another randomized controlled study published in 2016 found that overweight and obese adults who used a social support app lost on average 3 kg more than patients using a self-monitoring control app over the course of the study [59]. Evidence from other recent trials has demonstrated the ability of apps to reduce consumption of sugar-sweetened beverages in women and nutrient-poor foods in men, increase activity level and reduce fatigue following stroke, and improve respiratory parameters with a reduction in corticosteroid usage among individuals with uncontrolled asthma [60-62].

Conclusions

In the future, providers may routinely prescribe apps for their HNHC patients, and health care systems may invest in them. However, given the limited availability of high-quality evidence for most of the HNHC groups included in our review, we would not expect systematic reviews or meta-analyses focused on these groups individually to yield enough evidence to assess the effectiveness of disease-specific apps. Additionally, apps are being lost in translation from research to the app stores, resulting in a lack of commercial impact of existing research. Despite these limitations, the body of evidence overwhelmingly reports early results that favor the use of mobile health apps.

Acknowledgments

We would like to acknowledge the efforts of Shreyas Ramani, a master’s student in health informatics at the University of Michigan, for providing assistance with revisions of the manuscript. This research was supported by The Commonwealth Fund.

Conflicts of Interest

EP serves as a mentor at Rock Health and previously served as Chief Medical Officer at Twine Health, a mobile health company. AL currently serves as an advisor to the Hacking Medicine Institute, a nonprofit organization that evaluates digital health apps, where he also serves as senior editor on the RANKED Health project.

DWB is a coinventor on Patent No. 6029138 held by Brigham and Women’s Hospital on the use of decision support software for medical management, licensed to the Medicalis Corporation. He holds a minority equity position in the privately held company Medicalis, which develops Web-based decision support for radiology test ordering. He serves on the board for SEA Medical Systems, which makes intravenous pump technology. He is on the clinical advisory board for Zynx, Inc, which develops evidence-based algorithms. He consults for EarlySense, which makes patient safety monitoring systems. He receives equity and cash compensation from QPID, Inc, a company focused on intelligence systems for electronic health records. He receives cash compensation from CDI (Negev), Ltd, which is a not-for-profit incubator for health information technology start-ups. He receives equity from Enelgy, which makes software to support evidence-based clinical decisions. He receives equity from Ethosmart, which makes mobile apps to help patients with chronic diseases. He receives equity from Intensix, which makes software to support clinical decision making in intensive care. He receives equity from MDClone, which takes clinical data and produces deidentified versions of the data. DWB’s financial interests have been reviewed by Brigham and Women’s Hospital and Partners HealthCare in accordance with their institutional policies.

The views presented here are those of the authors and not necessarily those of The Commonwealth Fund or its directors, officers, or staff. The Commonwealth Fund was not involved in any of the following: design or conduct of the study; collection, management,
analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Multimedia Appendix 1
Search Terms.

[PDF File (Adobe PDF File), 37KB - mhealth_v4i4e136_app1.pdf]

Multimedia Appendix 2
Abstraction Form.

[PDF File (Adobe PDF File), 56KB - mhealth_v4i4e136_app2.pdf]

Multimedia Appendix 3
Articles included in review (n=175) and complete results of abstraction.

[XLSX File (Microsoft Excel File), 60KB - mhealth_v4i4e136_app3.xlsx]

References

7. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222 [FREE Full text] [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]


Abbreviations

HNHC: high-need, high-cost
IQR: interquartile range

©Karandeep Singh, Kaitlin Drouin, Lisa P Newmark, Malina Filkins, Elizabeth Silvers, Paul A Bain, Donna M Zulman, Jae-Ho Lee, Ronen Rozenblum, Erika Pabo, Adam Landman, Elissa V Klinger, David W Bates. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 19.12.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.