Designing and Evaluating mHealth Interventions for Vulnerable Populations: A Systematic Review

Elizabeth Stowell\textsuperscript{1}, Mercedes C. Lyson\textsuperscript{3}, Herman Saksono\textsuperscript{1}, René C. Wurth\textsuperscript{1}, Holly Jimison\textsuperscript{2}, Misha Pavel\textsuperscript{2}, Andrea G. Parker\textsuperscript{2}

\textbf{ABSTRACT}

Diverse disciplines, including Human-Computer Interaction have explored how mobile health (mHealth) applications can transform healthcare and health promotion. Increasingly, research has explored how mHealth tools can promote healthy behaviors within vulnerable populations—groups that disproportionately experience barriers to wellness. We conducted a systematic review of 83 papers from diverse disciplines to characterize the design and impact of mHealth tools in low-socioeconomic (low-SES) and racial/ethnic minority individuals. Our findings highlight that the diversity within low-SES and racial/ethnic minority groups was not reflected in the populations studied. Most studies focused on improving the health of individuals, often neglecting factors at the community and society levels that influence health disparities. Moreover, few improvements in health outcomes were demonstrated. We further discuss factors that acted as barriers and facilitators of mHealth intervention adoption. Our findings highlight trends that can drive critically needed digital health innovations for vulnerable populations.

\textbf{Author Keywords}

Systematic Review; Vulnerable Populations; Race; Ethnicity; Socioeconomic Status; mHealth; Minority.

\textbf{ACM Classification Keywords}

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

\textbf{INTRODUCTION}

Though still a relatively young field, there is a strong and growing literature on implementing and evaluating the impact of mobile health (mHealth) approaches. The resulting systems show promise, given that they can help people make healthy decisions when and where they need the support, and given that the prevalence of mobile devices across socioeconomic groups is increasing [99,100]. Within Human-Computer Interaction (HCI), a large body of work has yielded novel mobile applications to promote behaviors such as physical activity and healthy eating [20,52,68,81]. Although prior work has predominantly focused on these broad wellness domains, some researchers have begun to examine how mHealth systems can specifically meet the needs of vulnerable populations such as low-socioeconomic status (low-SES) and racial and ethnic minority groups [86,122,123]. Even given the proliferation of mHealth research in HCI generally, and the growing focus on vulnerable populations specifically, the field lacks a comprehensive synthesis and analysis of the body of mHealth research done inside and outside of HCI, and how well such tools are serving the needs of vulnerable populations. Such a review of prior work is needed to enable HCI researchers—and mHealth researchers more broadly—to take stock, identifying design implications and open research questions within the field.

To this end, we conducted a systematic review of scholarly papers describing work done to design and evaluate mHealth technologies in vulnerable populations. Through a literature search and a detailed analysis of 83 papers published across a variety of disciplinary venues, we synthesized the level of impact that mHealth interventions have had, as well as lessons learned, design implications, and open questions in this research space. Systematic reviews of mHealth research have examined the effectiveness of mobile technologies for improving health care and service outcomes [38,44], text messaging interventions [13], mobile tools for health behavior change and disease management [24,41], and mobile health information access [43]. Researchers have also systematically reviewed interventions that specifically focus on health equity, that is, improving the health of vulnerable populations who disproportionately experience barriers to health and poor health outcomes [6,123]. However, there are limited analyses of the design and impact of mHealth interventions in vulnerable groups. Montague and Perchonok [89] assessed health technology interventions amongst racial and ethnic minority groups, but included few studies that explicitly examined mobile technologies. Our review focuses specifically on mobile tools, within an expanded population purview that includes low-SES populations in addition to racial and ethnic minority groups given that these are...
vulnerable groups which also face disproportionate barriers to healthy living. (Note that our review also included individuals with disabilities as a target population; as only one paper in our corpus examined this population, we restrict our reporting to papers focused on low-SES and racial and ethnic minority groups.) Our focus on mobile technology use within these populations is particularly important given the prevalence of such technology, the ability for these platforms to reach broad audiences, and the ability for mobile systems to enable “just-in-time” interventions.

While there is considerable optimism that mHealth tools can transform health promotion, most meta-analyses find small to moderate impacts of such technology on health outcomes. One review found SMS (short messaging service) reminders for medication adherence in the general population have at best small effects [44]. The authors concluded that there remains insufficient high quality evidence of the beneficial effects of mobile technology on clinical outcomes to warrant implementing such tools for other areas of health behavior change [44]. Buchholz et al. [13] found that physical activity text messaging interventions show promising early results towards improving physical activity; yet these kinds of studies are characterized by "small sample sizes, heterogeneous (but positive) effect sizes, and a lack of specificity as to the characteristics of the text messages used in these studies." In addition, the effectiveness of these types of interventions depends on the details of the implementation and interactions with users [104]. In summary, prior reviews have frequently assessed the impact of SMS-based interventions, reporting mixed results. By analyzing a more diverse set of mHealth projects (including but going beyond SMS-based systems), we contribute to the field by summarizing the impact that these interventions have had on health outcomes in low-SES and racial and ethnic minority groups.

Echoing well-established user-centered design principles, prior work has found that it is imperative that mHealth technologies be tailored specifically for the end user [41,56]. Projects that attempt to take a previously designed technology and implement it among underserved or vulnerable populations can have less than optimal success, particularly with respect to sustainability of use over time. As Montague and Perchonok [89] found, "personally relevant and contextually situated health technology is more likely than broader technology to create behavior changes." They also implore designers and researchers to reflect on and state their best practices, so a community of user-centered designers can learn from one another and work towards ever-evolving benchmarks for culturally informed health information technology. Accordingly, a second contribution of our systematic review is a synthesis of the usability and user experience findings across mHealth projects focused on health equity. A third contribution is our analysis of the extent to which these projects took a culturally and contextually tailored approach to creating the mHealth tools, and by what means.

The findings from our comprehensive systematic review will help guide mHealth research by characterizing best practices, and the limitations of existing research. We present opportunities for future mHealth research focused on health equity: designing more engaging mHealth software and evaluating such tools, constructing more representative study populations, and grounding the research more solidly in theory.

**METHODS**

Our systematic review was guided by four research questions. RQ1: How has the socioecological context of the patient/consumer informed the design of mobile health technologies for vulnerable populations? RQ2: What barriers hinder, and what facilitators stimulate the use of mHealth tools in vulnerable populations? RQ3: How are mHealth technologies used in U.S. vulnerable populations? RQ4: What is the evaluated impact of mHealth technologies in the U.S. vulnerable populations?

Our review comprised five steps: database search, abstract screening, full-text screening, data extraction, and analysis. Papers were included if they met the following criteria: (1) Published in a peer-reviewed publication, (2) Published in English, (3) Study was conducted or data collected in the United States, (4) Technology used is mobile (cell phone, smartphone, tablet, wearable sensor, etc.), (5) Target population, or the population recruited, included a significant proportion of subjects from one (at least) of the following vulnerable groups: low-SES, racial/ethnic minority, or individuals with disabilities; (6) Paper included the primary human subjects data, (7) Patients/lay people were the primary users of the technology developed and/or tested; if no technology was developed (e.g., the paper describes a formative study), then the paper must indicate that the goal was to design a patient-facing technology, (8) The technology was related to health.

Papers were excluded if: (1) The technology discussed was designed for/mean to be used primarily by clinicians or health-care professionals, (2) The technology described was Internet-enabled, but not explicitly designed/meant to be used in a mobile context or on a mobile hardware platform, (3) The paper described a technology that primarily helps health professionals or researchers collect data to assess patients, (4) The technology described did not include features intended to directly influence patient/consumer health-related behaviors, knowledge, attitudes, access to clinical information, or environments, (5) The article was an abstract, or otherwise not a full paper, or (6) The empirical component of the paper focused solely on process evaluations of recruitment, enrollment, and retention.

**Database Search**

Our Database Search consisted of three literature searches conducted in: July-September 2014, January-February 2015, and January-March 2016. We used 11 databases: ACM Digital Library (including specific searches within the SIGCHI database), IEEEExplore, Google Scholar, Science
Direct, PubMed, CHINAHL, MEDLINE, PsychINFO, Sage Journals, Compendex, and PAIS. An iteratively developed boolean search string was built including up to 74 AND/OR terms across three categories (technology, health, population characteristics), and up to 12 NOT terms. The terms in each search varied depending on the constraints and capabilities of each database. When permitted by the database, we used the following search filters: published in English, limited to the United States. We reviewed a total of 64,249 titles, resulting in 340 relevant titles. Additionally, the bibliographies of papers included after the full-text screening were searched for relevant titles.

**Figure 1. Process Figure. n = # of Papers.**

**Abstract and Full Text Screening**
When the titles appeared relevant, abstracts were read. Titles were considered relevant based on the inclusion of keywords relating to mobile technology and health. If the abstract indicated that the article was broadly related to mobile health technology or an intervention among a vulnerable population, the paper was included in the full text screening. A total of 208 abstracts were considered relevant and included for full text screening. Next, the bodies of 208 papers were assessed using our inclusion and exclusion criteria (stated above). A total of 91 papers met all criteria and were added to the corpus.

**Quality Assessment**
We assessed the methodological quality of each article in our corpus using established study quality assessment tools [1,128]. We also adapted the qualitative study assessment tool developed by the Critical Appraisal Skills Programme [1] to develop a new form for assessing user experience research. Two independent reviewers assessed the quality of each article separately, then discussed their assessment and any disagreements until they were resolved. A third researcher reviewed any disagreement that could not be resolved. 50 papers received a “good” quality rating, 34 received a “fair” rating, and 7 received a “poor” rating. Papers receiving a “poor” rating were removed from the corpus resulting in 84 studies in the corpus.

**Data Extraction**
Next, two researchers independently extracted information from each article based on a predetermined set of data fields, such as the characteristics of the study design and population samples, and the empirical findings (health outcomes and user experience). For all papers, the results of the study, the authors’ unexpected findings, conclusions, and recommendations for future work were extracted for analysis. For each paper, the two researchers met to reconcile any differences in data extraction.

**Thematic Analysis**
The results, discussion and conclusion data from each paper were entered into ATLAS.ti version 1.6.0. We inductively created a codebook that included 41 codes to address our research questions. Two independent coders used the codebook to conduct a deductive qualitative analysis to arrive at themes across papers. Any differences between the coders were discussed and resolved. When assessing the strength of the code, we examined the number of studies in which the code appeared.

**Meta-analysis**
We also conducted a Meta-analysis, extracting quantitative findings to assess how successful projects were at improving the health of the populations studied. We began by computing descriptive statistics to assess trends across the 14 randomized control trials (RCTs) in our corpus, as these studies are a robust measure of quantitative effects [4,8,18,31,39,53,58,60,75,88,92,96,116,118]. The meta-analysis included only studies with an RCT design and health outcomes that were common across at least 2 studies, to enable the calculation of an aggregated effect size. These 5 studies [3,58,60,92,116] included 2 which contributed 2 health outcomes to the analysis [3,116]. G estimates, a commonly accepted unbiased estimate of the standardized mean difference (Cohen’s d), and variances were calculated [78], and used in a random effect omnibus test to show the aggregated effect and heterogeneity between studies with the same outcomes. The G estimates were then plotted in a forest plot with 95% confidence intervals to visualize the comparison. This analysis was performed with R statistical software, using the MAd, metafor, and rmeta packages.

**RESULTS**
Although our initial definition of vulnerable populations included individuals with disabilities, only 1 of the 84 papers included in the corpus focused on this population. Given the sparse data on disability-focused research, we report only on the 83 papers that focus on low-SES and racial/ethnic minority populations.

For each paper, where possible, information was collected and analyzed separately for formative and evaluation studies. Papers were considered to report on evaluation studies if they discussed work done to identify design requirements or guidelines for the technological intervention. In some cases, the paper
Study Design and Measures

Formative Studies (n=42) | Evaluation Studies (n=59)
--- | ---
**Population Density**
Urban | 52.38% (22) | 61.02% (36)
Suburban | 0% (0) | 0% (0)
Rural | 9.52% (4) | 1.69% (1)
Mixed | 7.14% (3) | 6.78% (4)
Other | 0% (0) | 1.69% (1)
Not Reported | 30.95% (13) | 28.81% (17)

**Study Duration (weeks)**

<table>
<thead>
<tr>
<th>Intervention Frequency</th>
<th>Evaluation Studies (n=59)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Session</td>
<td>71.43% (30)</td>
</tr>
<tr>
<td>&gt; 1, &lt;= 4</td>
<td>11.90% (5)</td>
</tr>
<tr>
<td>&gt; 4, &lt;=13</td>
<td>2.94% (1)</td>
</tr>
<tr>
<td>&gt;13, &lt;= 26</td>
<td>0% (0)</td>
</tr>
<tr>
<td>&gt;26, &lt;= 52</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Not Reported</td>
<td>14.28% (6)</td>
</tr>
</tbody>
</table>

Table 1. % of studies (and n) in settings with various population densities and with various study durations.

reported on both a formative and an evaluation study. 41 papers exclusively reported results from evaluation studies [3–5,7,8,11,15,16,18,19,27,29,31,34,35,39,40,42,46,50,52,53,58,60,62,65,80,88,90,92–96,103,111,115–119]. An additional 18 studies reported findings from both formative work and the subsequent evaluation studies [2,23,26,28,33,48,61,64,67,73,75–77,79,84,97,107,120]. 24 papers reported exclusively on formative studies [9,10,12,14,17,21,22,25,30,32,45,54,63,66,69–71,87,101,112–114,124,125]. This resulted in a total of 59 evaluation studies and 42 formative studies included in the corpus.

Our analysis of the research represented in our corpus offers an in-depth look into the state of mHealth research within low-SES and racial and ethnic minority groups. We first characterize trends in methodology and types of technology explored, reporting on the study design, study population characteristics, technology characteristics, and theoretical bases of the projects in our corpus. Through our analysis of study findings and lessons learned, we also characterize barriers to and facilitators of mHealth intervention adoption in vulnerable groups, as well as the impact of these interventions.

**Study Design and Measures**

**Study Design & Sample Size**

Of the 83 studies in our corpus, 28.92% (24) were qualitative, 20.48% (17) controlled intervention studies, 24.10% (20) UX/Usability, 15.66% (13) pre/post without a control group, and 10.84% (9) were observational cohort and cross sectional studies.

88.09% (29) of the 37 formative studies reported sample size and all of the evaluation studies reported sample size. A total of 4343 participants were included in formative studies (Median=27, IQR=68, Range=5-1157 participants per study) and 14,209 were included in evaluation studies and completed the intervention (Median=37, IQR=75, Range=5-7,574 participants per study).

**Study Location**

The context in which an individual interacts with mobile technology has implications for how applications should be designed and helps to contextualize how such tools are adopted (or not). However, most of the 59 papers with evaluation studies did not report on the type of setting in which the study was conducted (76.27%, 45). Settings that were reported included the home (6.78%, 4), neighborhood (6.78%, 4), health clinics (10.17%, 6), or other settings such as community centers or college campuses (6.78%, 4). For example, Sheats et al. [110] evaluated how active and passive photo capturing can help low-income Latinos identify elements of their neighborhood’s built environment that influence their opportunities for physical activity [110]. Additionally, Czaja et al. [31] designed a mobile intervention for the home setting targeted at caregivers, who spend significant amounts of time at home with the person for whom they care [31]. No evaluation studies were conducted in the context of school, workplace, or religious institutions, even though these environments are common sites for health interventions, and have been explored as venues for technological innovation in HCI research across application domains [37,82,109,126].

As Table 1 shows, over 60% of evaluation studies and over 50% of formative studies were conducted in urban settings. Rural settings were rarely studied and no papers reported on suburban settings. Almost 30% of formative and evaluation studies did not report on population density.

**Study Duration & Intervention Frequency**

Most of the 42 formative studies lasted for a single session (71.43%, 30). 14.28% (6) of papers did not report the formative study’s duration. There was great variation in the duration of the evaluation studies, which ranged from a single session to a year (Table 1). The average evaluation study length was 11.91 weeks (SD = 12.02).

In the 59 papers reporting evaluation studies, participants were asked to use the technology at different frequencies. (Table 2) In most studies (61.02%, 36), participants were asked to use the technology at least multiple times per week, with 45.43% of studies (27) asking participants to use the technology at least daily. 15.25% (9) of studies did not report the frequency of the intervention, crucial information for interpreting user engagement and impact.
Common technology evaluation measures included feasibility or acceptability on health outcomes, services utilization, biometric measures of health (e.g., efficacy, behaviors, eating). Measures of health outcomes included self-assessments of factors like blood pressure, HbA1c, diabetes perceptions and attitudes that arose through technology use, and frequency of access to technology. 13.56% of studies did not report on participant interaction with the technology.

**Study Population Characteristics**

The demographic characteristics of the study populations were determined to investigate the appropriateness of the interventions for the populations targeted and groups that were understudied (Table 3).

**Race & Socioeconomic Status**

96.39% (80) of papers in our corpus targeted or included a significant proportion of participants from racial or ethnic minority groups. Both formative and evaluation studies mostly included individuals who were reported to be African American or Black and/or who were Latino or Hispanic. Participants who were reported to be Native American or Pacific Islander were sparsely represented. 63.86% (53) of studies targeted or included a significant proportion of participants of lower-socioeconomic status (SES). There were no comparative studies examining populations of different SES.

**Language**

English and Spanish were the dominant languages represented in our corpus in both formative and evaluation studies. In the few studies that included participants who spoke a language other than English or Spanish, these participants made up less than 2% of the study population.

**Gender**

Gender distributions were only calculated from papers that provided the sample size as well as the percent or number of participants of a gender that was included in the study. This includes 38 formative studies and all 59 evaluation studies. While females were over-represented across the formative and evaluation studies in our corpus, this was particularly the case in the formative studies (Table 4).

**Age**

Average age was calculated from papers that provided the sample size and participants; mean age. Variance statistics were inconsistently reported, thus the pooled variance was not calculated. In the 21 formative studies that were included, the weighted mean age of participants was 22.9 years old. In the 36 evaluation studies that were included, the weighted mean age of participants was 30.4 years old. As such, the mean age for participants in formative studies was almost 10 years younger than in the evaluation studies.

**Education**

Education level data was calculated from papers that provided the sample size and participants; mean age. Variance statistics were inconsistently reported, thus the pooled variance was not calculated. In the 21 formative studies that were included, the weighted mean age of participants was 22.9 years old. In the 36 evaluation studies that were included, the weighted mean age of participants was 30.4 years old. As such, the mean age for participants in formative studies was almost 10 years younger than in the evaluation studies.

**Measures of Health Outcomes**

Just over half of the evaluation studies (55.93%, 33) assessed the health-related behaviors of participants (e.g., eating habits [52], receipt of vaccine [117], and sexual behaviors [77]). Many studies (30.51%, 18) measured knowledge (e.g., pre-diabetes perceptions [19] and prostate cancer knowledge [120]) and health-related self or collective efficacy (25.42%, 15). An additional 23.73% (14) used biometric measures of health (e.g., Body Mass Index [92], blood pressure [7], HbA1c [3]). 32.20% (19) used other methods (e.g., psychosocial measures [96] and health services utilization [3]). 30.51% (18) of studies did not report on health outcomes, focusing on assessments of intervention feasibility or acceptability.

**Technology Evaluation Measures**

Common technology evaluation measures included technology satisfaction (62.71%, 37), frequency of use (42.37%, 25), user experience—such as participant behaviors and attitudes that arose through technology use (42.37%, 25), and ease of use (33.90%, 20). 15.25% (9) used other measures (e.g., cultural appropriateness of intervention wording, message understandability [73], parental rules around technology use, and frequency of access to technology [8]). 13.56% of studies (8) did not report on participant interaction with the technology.
intervention within a population. Only 57.14% (24) of formative studies and 54.23% (32) of evaluation studies reported the education level of their participants. Most evaluation studies had a majority of participants with some college education or below (Table 3).

**Technology Characteristics**

Amongst the evaluation studies, the device most commonly assessed was a cell phone without Internet capability (61.02%, 38). Papers also reported using smart phones (cell phones with Internet capability; 13.56%, 8), tablets (8.47%, 5), PDAs (6.78%, 4), and wearable fitness trackers (3.39%, 2). 15.25% (9) of studies used other devices such as digital scales or videophones.

Although mobile platforms can connect people to social networks that can support health, most of the evaluated technology was intended to be used by an individual (88.13%, 52). Very few systems were designed to be used by social groups, including families (5.08%, 3), neighborhoods (5.08%, 3) or other groups (10.16%, 6).

The technologies used in the 59 evaluation studies supported a number of functions. In 59.32% (35) of studies, the mHealth tools enabled one and/or two-way text messaging. In 8.47% (5) of papers, the software supported photo or video messaging. 5.08% (3) of studies used phone calls and 1.69% (1) used interactive voice response. In 15.25% (9) of papers, the technology enabled daily caloric records (e.g., calorie counters), and in 3.39% (2) of papers, database access was provided (e.g., recipe indexes). Additional features, such as photo capture and viewing, videos, voice messages, social networking, and games, were leveraged to support interventions’ health promotion goals.

The technology fulfilled a myriad of health promotion functions. In most studies (71.19%, 42), the technology provided information to the user (e.g., delivery of expert knowledge, peer information sharing). Most studies (71.19%, 42) also evaluated technology that provided behavioral support (e.g., skill development, benchmarking, incentives). In 16.95% (10) of studies, the technology increased access to clinical care (e.g., access to personal health records). In 6.78% (4) of studies, the technology was intended to change the physical and/or social environment to promote health or prevent disease (e.g., connecting friends in a mobile fitness tracking application).

**Population-Targeted Design**

Only 54.24% (32) of the 59 evaluation studies discussed how the technology was targeted at the specific population under study. The most common forms of population targeting included targeting users’ language (20.34%, 12), literacy level (10.17%, 6), and cultural health beliefs and attitudes (8.47%, 5). Interventions were also customized depending on the user population age group (5.08%, 3), eating practices (3.39%, 2), and family values and dynamics (1.69%, 1).

**Theoretical Basis**

We grouped the theories guiding the studies in our corpus into one of four categories: Intrapersonal, Interpersonal, Community, or Ecological. These clusters are based upon the classification of health behavior theories presented by Glanz et al. [49]. The *intrapersonal* level encompasses an individual’s behaviors, and factors such as their beliefs, knowledge, and skills [102]. The *interpersonal* level describes the influence of other people on an individual’s health behavior [57]. The *community* level examines the influence of larger organizations such as a company, coalition, or government on an individual’s health behavior [47]. *Ecological* theories examine the interplay of health behavior influences at multiple levels—including each of the three aforementioned levels as well as societal factors such as public policies [106].

Just over half the 83 papers in our corpus did not report a theoretical basis for their work (55.42%, 46). Most papers that did report a theoretical grounding used intrapersonal (20.48%, 17) or interpersonal (16.87%, 14) level theories. The most commonly used intrapersonal theory was Social Cognitive Theory (7.23%, 6). Only 6.02% (5) of studies used community level theories of behavior change, with Diffusion of Innovation and McGuire’s Input/Output Persuasion Model being the most commonly used. Only 2.41% (2) of studies reported using an ecological theory, and they relied upon the Social Ecological Model and the Bioecological Model.

**Technology Use: Barriers and Facilitators**

We qualitatively analyzed the findings and discussions of the findings reported in our corpus to determine common barriers to, and facilitators of, effective intervention implementation and adoption. In many cases, a barrier or facilitator was not confidently determined within a study, but rather discussed by the authors as a potential contributor to the intervention’s success or ineffectiveness. Therefore, in addition to the barriers and facilitators codes used in our analysis, each instance of the code was assigned one of three secondary categorizations. Codes categorized as *data grounded* (DG) represented findings that were derived from data collected and analyzed in the study. Codes categorized as *data grounded hypothesis* (DGH) were defined as instances of the code in which: 1) the researchers make a hypothesis or assertion that a particular factor would have been a barrier or facilitator, 2) that hypothesis is based on data collected from participants, and 3) the researchers did not test this hypothesis in their study. Codes categorized as *hypothesized* (H) were defined as instances in which: 1) the researchers hypothesize that a particular factor would have been a barrier or facilitator to the intervention, but 2) the researcher did not test this hypothesis in their study. The sections below discuss common facilitators and barriers identified across the 83 papers in our corpus.

**Technology Familiarity, Accessibility & Cost**

The use of technology that participants were already familiar with was the most commonly reported facilitator of intervention success, and was reported by 23 studies.
(27.71%, 18DG, 1DGH, 4H). Rodgers et al. reported on a project in which participants used Photobucket, a freely available smartphone app, to track meals. However, many of these participants suggested using a more familiar app such as Instagram [103]. Devine et al., [33] used text messages to deliver a pregnancy prevention program to teens, pairing teens without cell phones with those who had access to cell phones. Teens who were already comfortable with text messaging were open to receiving the intervention content via text message [33]; however, teens who did not own cell phones often missed intervention content, as their buddy had already deleted the message.

Using technology that was unfamiliar to the participant was a barrier reported by 7 studies (8.43%, 5DG, 2H). Use of unfamiliar technology resulted in complications such as lengthy training sessions [42], in participants being unsure of what functions were possible with the technology [114], and anxiety around operating the technology [5]. Similarly, lack of access to the technology outside of the research setting was a barrier identified in 9 studies (10.84%, 8DG, 1H). An additional 9 (10.84%, 4DG, 1DGH, 4H) studies cited the high cost of technology and technology use as a barrier to recruiting and retaining participants in intervention studies. These costs included the cell phone plan prices and the cost of replacing lost devices. For example, Evans et al. [39] found that many declined to participate in their study out of a concern for the cost of cell phone minutes required for the intervention. Cornelius and St. Lawrence [30] conducted formative work to develop a text messaging intervention for HIV prevention. Their participants felt it would be beneficial if the study covered all costs of the text messages.

**Timing of Content Delivery**
Frequent delivery of intervention content was acceptable and even desirable in 13 studies (15.66%, 10DG, 3DGH). Steinberg et al. sent daily text messages prompting participants to report on their weight loss goals to promote self-monitoring of their progress [116]. Participants felt that the daily text messaging was important and an appropriate frequency to promote weight loss. Fischer et al. sent SMS reminders for blood sugar readings every other day, and found that half of participants wanted daily messages [40].

15 studies (18.07%, 5DG, 7DGH, 3H) found that personalizing the timing of the intervention was a facilitator. This included sending messages at critical times (e.g., reminders around personalized meal times for nutrition advice [103]) or when the user would likely have access to their mobile device (e.g., after school [29]). Participants across these studies seemed to prefer receiving intervention content at times that were relevant to the health challenge they were trying to manage.

**Engaging Content**
In 20 studies (24.10%), Participants enjoyed the use of multimedia within various platforms (14GH, 4DGH, 2H). This included use of audio, visuals, and videos to supplement or replace text. Chinn et al. found that photos on the mobile app helped participants recall what their children had eaten that day [23]. Lu et al. found that participants preferred the multimedia tablet-based depression education module over the paper pamphlets and hypothesized that the use of multimedia minimized the barrier of low literacy in the community health clinic population in the study [79].

Other methods of increasing enjoyment or engagement with the interventions included personalization, variation, and gamification of the intervention content. 16 studies (19.28%, 4DG, 5DGH, 7H) argued that personalization of the intervention content would increase satisfaction or effectiveness of the intervention. Varying the content delivered through the intervention was suggested as a means of increasing intervention engagement in 8 studies (9.64%, 3DG, 3DGH, 2H). Additionally, 6 studies, (7.23%, 2DG, 1DGH, 3H) discussed how gamification of the target health behavior may increase intervention engagement. While personalization, variation, and gamification were often hypothesized as beneficial to an intervention, they were scarcely evaluated in the corpus. Only 13 studies evaluated technologies that enabled personalized timing or content and only 2 studies evaluated health promotion games, both of which focused on nutrition education.

**Confidentiality and Privacy**
Technology can enable more confidential methods of information and support seeking. 16 studies (19.28% 5DG, 3DGH, 8H) found this affordance to be a facilitator of intervention success. Of the 16 papers that discussed confidentiality, 12 focused on sexual and reproductive health. In a study of text-message based intervention for HIV prevention, participants listed confidentiality as an advantage to the technology platform when seeking HIV information [29]. Jones and Lacroix [64] streamed an HIV risk-reduction soap opera to participants’ personal smartphones, which allowed users to view the content in the location or at the time of their choosing-- supporting a feeling of privacy.

Confidentiality was also a factor in the recruitment of subjects. In their efforts to recruit from a health clinic that serves predominantly low-income, non-English speaking immigrants, Evans et al. [39] noted anecdotally that some patients refused to enroll in the study because they were concerned about providing personal information before having completed their immigration paperwork.

While confidentiality and privacy was most often discussed in studies addressing sexual and reproductive health, the theme was also found in other health domains. In a family-based nutrition app, Schaeferauer et al. found that sharing diet information among family members can create tension [107]. As a result, it was important to develop an intervention that balanced the values of individual privacy and family transparency. These findings highlight the need to account for the sensitivity of health information when working in all health domains, and not only in traditionally “sensitive” or stigmatized health domains such as sexual and reproductive health or mental health. It is additionally important to
understand how the sensitivity of information is perceived in different populations.

Impact on Health Outcomes
To begin identifying how successful mHealth interventions have been at improving the health of vulnerable populations, we examined the impact on health outcomes reported by the randomized controlled trials (RCTs) in our corpus. These studies were chosen given that RCTs are widely considered the gold standard for experimentally measuring the impact of health interventions.

Our findings show a mixed level of success. 64.29% of the RCTs (9 out of 14) yielded at least mildly significant improvements (p<0.10) in the evaluated health measures, including weight loss and increased self-efficacy for health knowledge. Projects in the following health clusters saw the most success, achieving a highly significant impact (p<0.05) in the intervention group as compared to the control in most papers: weight loss (2 of 3 papers), nutrition and physical activity (3 of 4), and vaccine adherence (2 of 3). In interventions focused on diabetes, only studies involving video messages resulted in a significant decrease in hemoglobin A1c (2 of 3).

SMS was the most commonly evaluated technology in RCTs. Only 57.14% of SMS RCTs in our corpus (4 of 7) demonstrated at least mildly significant (p<0.10) improvements on health outcomes. This is an important finding, given the popularity of text messaging as an intervention approach in our corpus (60.71% of evaluation papers) and in prior reviews [24,56,127]. Interestingly, all 3 RCTs that incorporated voice or video into their mHealth system design (e.g., for counseling) had significant results.

Meta-analysis

We conducted a meta-analysis to statistically combine and evaluate the results across RCTs in our corpus. Figure 2 shows the individual g estimates (95% CI) for each study in the meta-analysis, as well as overall g estimate (95% CI) for each outcome.

After running the omnibus test, weight change outcomes had a summary statistic (overall g) of -0.73 (95% CI=-1.41, 0.04), which was significant (p=0.04), with insignificant heterogeneity between these studies (QE=1.67; p=0.20). For BMI studies, the overall g was -0.18 (95% CI=-0.58, 0.21), which was not significant (p=0.36) and did not have significant heterogeneity (QE=4.16; p=0.12). HbA1c studies had an overall g estimate of 0.01 (95%CI=-0.20, 0.23), that was not significant (p=0.91) and also not found to be significantly heterogeneous across the two studies (QE=0.65; p=0.41). In summary, apart from the small success demonstrated in the weight change studies (owing to the highly significant results of one of these two papers), overall we do not find evidence that the interventions in our meta-analysis have successfully impacted health outcomes in the vulnerable populations studied.

DISCUSSION
This systematic review examined the design and impact of mHealth tools in a set of groups that disproportionately experience barriers to wellness: racial/ethnic minority and low-SES individuals. Leveraging the increasing prevalence of mobile phones, wearable devices, and other computational platforms, there is a large and growing body of research examining how mobile platforms can support health promotion and management. An increasing subset of this work has explored opportunities for innovation and improved health in vulnerable groups, where barriers to wellness are highest and health outcomes are poorest. While our review identified many papers reporting such health equity research within the health, social, and behavioral sciences, there have been fewer contributions from the HCI community. Many HCI researchers have studied the role of smartphone applications, wearable activity trackers, and other systems in promoting health generally (e.g., [13,44,82,105]), but much less work has been done to determine how such innovations can be best designed and delivered within vulnerable populations.

With this systematic review, we have highlighted several trends to help catalyze research within HCI and other disciplines. Our synthesis of the methodological and technology design decisions made in prior work characterize the state of the art, several opportunities for improvement, and open research questions. By characterizing the barriers and facilitators to intervention adoption and success across previous work, we further highlight promising directions for the design of future mHealth applications. Finally, we have described how, despite all the innovation in this research space, there have been few improvements in health outcomes within low-SES and racial and ethnic minority groups. These findings underscore the need for continued research to determine how mHealth interventions can be designed to
improve health equity. We now turn to a discussion of important considerations for future research.

**Study Design**

Our findings highlight several implications for the design of formative and evaluation studies focused on mHealth interventions in vulnerable groups.

**Study Location**

First, over 76.27% of evaluation studies did not report the location in which the intervention was implemented. In one sense, these findings might be expected given that we focused on mobile interventions, which by definition are meant to support people as they travel through different settings. However, it is surprising to see that so few studies were conducted in specific community contexts, given that each setting can support or inhibit healthy decision-making, and hence may impact how mHealth interventions should be designed. As the papers in our corpus show, neighborhood locales can act as valuable anchors for mobile interventions—providing: a sense of legitimacy to interventions (e.g., a shared environment for participants to relate to [51]), logistical and operational support to encourage sustainability and social accountability (e.g., developing rapport with a local organization through volunteer hours to promote technology adoption [107]), complimentary real world supports that enhance the mobile component (e.g., reminder text messages for influenza vaccinations at community clinics [119]), and a focus point for change (e.g., a photo capture tool that helps community residents identify neighborhood attributes affecting physical activity [110]). Indeed, while many interventions have been grounded in settings such as health clinics, the home, neighborhoods, and religious institutions, little research has evaluated mHealth tools in such contexts, mirroring trends in HCI research more broadly [108]. We encourage future work that is grounded in community contexts, and increased reporting on the implications of varied settings on technology delivery and adoption.

**Study Length**

Our findings also suggest a need for longer formative studies. The formative studies in our corpus were most often conducted as a single session (71.43%, 30). Additionally, of the 59 evaluation studies in the corpus, 47.46% did not report having done formative work prior to designing and implementing their intervention. We suggest that future work conduct longer, more intensive formative studies as they may provide a more nuanced understanding of the population under study, resulting in more effective and culturally appropriate interventions.

Extended interactions with participants through formative work may also help researchers avoid making premature judgments and conclusions and to develop a more comprehensive understanding of the populations under study as well as their context [91]. Such understanding is particularly critical when conducting research with populations such as low-SES and racial and ethnic minority groups. Within these groups, individuals’ experience with and attitudes toward health and the healthcare establishment may be particularly impacted by multiple, complexly-intertwined levels of influence with a fraught historical legacy. Such influences must be thoroughly understood to appreciate the scope and nature of the challenges these populations face and to develop effective interventions which address these challenges. Without longer, more in-depth formative work, researchers risk designing evaluations based on a surface-level view of the challenges, attitudes, and resources within a community that affect health inequities.

**Measuring Health Outcomes, Usability & User Experience**

We identified significant barriers to comparing findings across studies. First, there was high variation in how interventions measured outcomes such as ease of use, technology satisfaction, and health behaviors (e.g. frequency of exercise). 30.51% (18) of evaluation studies did not measure health outcomes at all, further limiting our ability to assess the appropriateness of several interventions for improving health. Also, most studies only measured user experience only at the end of the study. While this may be appropriate for shorter studies, it can also limit knowledge of how user experience evolves over time. We encourage researchers to leverage standardized health and technology measures consistently across mHealth studies to enable comparisons, and more robust evaluations of usability and user experience over time.

**Population Sample**

Our review focused on the broad categories of low-SES and racial and ethnic minority groups, which are comprised of many subpopulations. Yet, the participants in our corpus did not reflect this diversity. First, very few studies included participants who were Native American or Pacific Islander—two groups who disproportionately experience health problems. Our findings mirror those described in a review of CHI proceedings from 1983-2016, which found no matches for the keywords First Nations, Alaskan Native, and Pacific Islander, among other identity keywords in their search of the ACM Database [108]. Second, males were underrepresented across studies in our corpus, despite the fact that even within socioeconomic, racial, and ethnic categories, there exist gender disparities in outcomes such as mortality, with males having shorter life expectancies [121]. Third, despite the diversity of languages spoken in the United States, only three studies reported including participants who spoke a language other than English and Spanish. Finally, only one evaluation study was conducted in a rural population. The disproportionate focus on urban contexts is reflective of health disparities research generally, despite the unique health barriers in rural settings [36,127]. Exploring how mHealth tools can meet the needs of rural populations is thus an open area for future research. Furthermore, the lack of reporting on population density (e.g., rural, urban) in 17 (28.81%) papers weakens the potential for those findings to be applied and translated in future projects. Increased reporting on population density is needed in future work.
Prior research has demonstrated that the unique experiences, attitudes, and linguistic nuances of populations must be considered to develop effective health interventions [108]. Thus, there is a need for research that teases apart the broad categories of “low-SES” and “racial and ethnic minority” to identify how mHealth tools can effectively address the needs of diverse subgroups.

Throughout our corpus, population characteristics were underreported, such as non-binary gender, race and ethnicity including multi-racial and ethnic groups, and languages spoken. Additionally, the intersectionality of these social categories was rarely discussed. Our results echo Schlesinger et al.’s findings that, in general, HCI studies rarely directly address how multiple identity characteristics (e.g., race and gender) interact to create unique challenges, opportunities, and assets [108]. Within health technology research, the interplay of identity categories are crucial to report and study, to help researchers interpret and apply findings around intervention adoption and impact.

Technology

The papers in our corpus highlighted several ways in which mHealth tools can better engage users, such as varying content and utilizing rich multimedia. One benefit of mobile platforms is their ability to personalize many aspects of an intervention. Indeed, personalization of the content and timing of the intervention delivery are usage facilitators highlighted within our review and in prior work [56,85]. Despite widespread acknowledgement of the importance of personalization, only 13 studies in our corpus leveraged the technology’s ability to personalize an intervention. Similarly, in our corpus, gamification was a commonly proposed solution to increasing participant engagement; yet, health-promoting games were seldom evaluated. Both trends highlight opportunities for increased empirical study to identify best practices for translating the high-level concepts of personalization and gamification into effective and engaging tools for vulnerable groups.

Theoretical Grounding and Individual Focus

Only 44.58% of the studies had a theoretical foundation, mostly based on intra and interpersonal theories of behavior change. Few studies relied on community (6.02%, 5) or ecological (2.41%, 2) theories. These findings may be partially explained by the fact that most of the technologies in our corpus targeted individuals, with very few designed for use by social groups (e.g. families or neighborhoods). Additionally, the interventions in our corpus most frequently provided information or support for behavior change, with only 4 evaluation studies focused on making changes to participant environmental context. While mobile technology is often viewed as personal, it has the ability to connect people to address health disparities on a larger scale and to leverage the resources and abilities of a community to make sustainable changes on a population level. By overemphasizing the role of an individual’s actions in health behavior change (and neglecting community and ecological levels of analysis), we ignore the impact of social, political, and physical environments on health outcomes, exacerbating health disparities. Choice of theory shapes the focus and structure of an intervention, resulting in real-world consequences [72]. Most current theoretical frameworks provide qualitative intervention guidance, although emerging approaches suggest system-theoretic, quantitative models [59]. There is a need to combine these theoretical advances with the design and evaluation of mHealth interventions that explore interpersonal, community and society level factors that impact health inequities.

LIMITATIONS AND FUTURE WORK

This systematic review represents a broad survey of the field. As a result, the papers in the corpus used a myriad of measures that impeded a more robust meta-analysis. Additionally, our corpus is restricted to studies that took place in the US in order to manageably scope the review [98]. Our analysis addresses recent calls for increased HCI research focused on vulnerable populations [74], by articulating concrete recommendations for how such work can be done in the domain of health. Future analyses of mHealth research from a global perspective can provide valuable directions for the field. For example, by replicating our systematic review protocol for mHealth research in other countries, researchers can compare and contrast findings across countries. Doing so will require analysis of the many important differences in the healthcare and health management experiences of individuals in different parts of the world (e.g., politically, culturally, and economically).

CONCLUSION

This systematic review reports on 83 papers focused on mHealth interventions in racial/ethnic minority and low-socioeconomic groups. Our findings reveal trends that indicate gaps in the current literature, including a need for more: in-depth formative studies, reporting of population characteristics, research on diverse subgroups, evaluation of user engagement strategies (e.g., personalization and gamification), and design and evaluation of community-level, technology-based interventions.

ACKNOWLEDGEMENTS

This work was supported by a grant funded by the Aetna Foundation. We thank Christine Gordon, Xuan Li, and Emma Simpson for their assistance with the review.

REFERENCES


37. Elsa Eiriksdottir, Yan Xu, Erika Poole, Andrew Miller,


73. Kuang-Yi Wen; Suzanne M Miller; Linda Kilby; Linda Fleisher; Tanisha D Belton; Gem Roy; Enrique Hernandez. 2014. Preventing postpartum smoking relapse among inner city women: Development of a theory-based and evidence-guided text messaging intervention. Journal of Medical Internet Research 16, 4: 1–12. https://doi.org/10.2196/resprot.3059


111. Jylana L Sheats, Sandra J Winter, Priscilla Padilla-Romero, Lisa Goldman-Rosas, Lauren A Grieco, and Abby C King. 2013. Comparison of passive versus active photo capture of built environment features by


