Context-sensitive ecological momentary assessment: Application of user-centered design for improving user satisfaction and engagement during self-report

Authors:
Preethi Srinivas, PhD, 1,2
Kunal Bodke, MS, 1,2
Susan Ofner, MA, 4
NiCole R. Keith, PhD, 1,2
Wanzhu Tu, PhD, 1,4
Daniel O. Clark, PhD, 1,2,3

Affiliations
1 Indiana University Center for Aging Research, Indianapolis, Indiana
2 Regenstrief Institute, Inc., Indianapolis, Indiana
3 Department of Medicine, Division of General Internal Medicine and Geriatrics, Indiana University School of Medicine, Indianapolis, Indiana
4 Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, Indiana

Corresponding Author:
Daniel O. Clark, PhD.
1101 West 10th Street
Indianapolis, IN 46202, United States
daniclar@iupui.edu

Acknowledgements:
We wish to acknowledge the contributions from persons who provided key expertise and effort to this project. Cori McCorkle and Anietra Butler recruited study participants and completed assessments with those participants. Tony French and Andrew Martin provided computer programming expertise to the project. The work was funded by the National Institute of Health, Lung, and Blood Institute (R01 HL128494).
Abstract

**Background:** Ecological Momentary Assessment (EMA) can be a useful tool for collecting real time behavioral data in studies of health and health behavior. However, EMA administered through mobile technology can be burdensome and it tends to suffer from suboptimal user engagement, particularly in low health literacy populations.

**Objective:** To report a case study involving the design and evaluation of a mobile EMA tool that supports context-sensitive EMA reporting of location and social situations accompanying eating and sedentary behavior.

**Methods:** An iterative, user-centered design process with obese, middle-aged women seeking care in a safety-net health system was used to identify the preferred format of self-report measures, and the look, feel and interaction of the mobile EMA tool. A single arm feasibility field trial with 21 participants receiving 12 prompts each day for momentary self-reports over a 4-week period (336 total prompts per participant) was used to determine user satisfaction with interface quality and user engagement, operationalized as response rate. A second trial among 38 different participants randomized to receive or not a feature designed to improve engagement was conducted.

**Results:** The feasibility trial results showed high interface satisfaction and engagement, with an average response rate of 50% over 4 weeks. Qualitative feedback pointed to the need for auditory alerts. We settled on three alerts at 10-minute intervals to accompany each EMA reporting prompt. The second trial testing this feature showed a statistically significant increase in response rate among participants randomized to receive repeat auditory alerts versus those not (60% vs 40%).

**Conclusions:** This paper reviews the design research and a set of design constraints that may be considered in the creation of mobile EMA interfaces personalized to users’ preferences. Novel aspects of the study include the involvement of low health literacy adults in design research, the capture of data on time, place, and social context of eating and sedentary behavior, and reporting prompts tailored to an individual’s location and schedule.

**Trial Registration:** NCT03083964

Keywords: user-centered design; mhealth; health disparities; ecological momentary assessment
Introduction
Precision medicine is an approach to care that involves classifying individuals into subpopulations that differ in their susceptibility to a particular disease, or in their response to a specific treatment [1]. Subpopulations can be defined by genetics but also by behavioral and environmental exposures that lead to differential responses to biomedical, behavioral and environmental interventions. The latter targets might be referred to as precision health interventions, and may be advanced by better measures of behavioral and environmental exposures [2]. Environmental exposures can be relatively constant or vary over relatively short intervals of time, which together make up what we will refer to in this paper as situations.

Large-scale precision health research efforts are in final planning [3] or just under way [4] and include the measurement of real-world and real-time physiological data from sensors such as accelerometers, heart rate and glucose monitors, among others. While detailed physiological measures will help with early detection of changes in physiological states and thus improve prevention or early treatment, precision health will require also measuring the situations and behaviors that directly or indirectly affect physiology [5]. Social, behavioral, and environmental factors contribute as much or more to health and longevity as other major domains including medical care and genetics [6]. Thus, better measures of behavior and situations are keys to not only understanding behavior, but also to precision health interventions and ultimately better health and longevity.

One of the techniques widely used to obtain a situational or contextual understanding of daily life includes experience-sampling methodology [7]. This method includes self-report measurement, but in a form where a person responds to subjective questions multiple times a day. This technique has often been attributed to overcoming methodological problems owing to memory and recall [8-9]. Further, this method has high ecological validity and supports within-subject investigations [10-11]. Prior health research has attempted to execute experience sampling on technology devices, such as personal digital assistants and pagers. With advancing technology, experience sampling has also been executed on mobile devices such as smartphones. Often referred to as ecological momentary assessment (EMA) in health research, EMA is typically completed as a person is experiencing something in their natural environment.

EMA is considered the gold standard of experiential sampling in health research [8]. However, self-report through EMA can be a burden given the need to administer instruments multiple times in a day. Further, the collected data can suffer from poor adherence and misreporting, especially if the instrument is cumbersome to use or doesn’t suit individually variable reporting needs and preferences (e.g., sleep and work schedules and location triggers) [12]. This motivates a need for sampling tools that not only support situation dependent, real-time self-report multiple times in a day but also are (1) low burden, (2) supportive of recurrent use, and (3) tailored to users’ needs.

The current study addresses these needs with a case study involving self-report measurement of location and social situations accompanying eating or sedentary behavior. This work was carried out in the context of a randomized trial among middle-aged obese women cared for in a safety-net health system [13]. We report the design and implementation of self-report measures of eating and movement behavior specific to users’ location and social contexts. Each measure was developed through an iterative, user-centered design process involving obese, middle-aged women and was deployed in a field trial to establish usability. The specific contributions of this
paper are: (i) a series of design constraints identified as important to consider and satisfy when designing mobile EMA interfaces that are personalized to users’ preferences, (ii) six refined measures for self-report of eating and sedentary behavior specific to location and social context using a mobile device, (iii) a characterization of the ways in which individuals prefer to self-report eating and movement, along with perceived benefits and challenges of this self-reporting, and (iv) field trials of feasibility with attention to response rates.

Methods
This study used an iterative user-centered research and design approach comprised of four phases (Table 1) to support the design and development of a mobile application (app). All sessions were audio- and video-recorded. Recordings were reviewed by stakeholders and designers prior to making design modifications to the EMA system. Qualitative thematic analysis was performed to identify and iteratively refine themes. The findings of Phases 1 and 2 were paired with design literature to guide the development of prototypes for evaluation in Phase 3 and 4 field trials. Overall, iterative participatory design and review sessions helped progress the identified measures from low-fidelity paper sketches to high-fidelity prototypes. This study is funded by the National Institute of Health, Lung, and Blood Institute (R01 HL128494) and is approved by the Indiana University Institutional Review Board.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Research method</th>
<th>Duration</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1: Exploratory ideation</td>
<td>Focus group with stakeholders</td>
<td>60 min</td>
<td>6 stakeholders</td>
</tr>
<tr>
<td></td>
<td>One-one design session with users</td>
<td>45-60 min</td>
<td>5 users</td>
</tr>
<tr>
<td>Phase 2: In-lab evaluation</td>
<td>Scenario-based think-aloud usability evaluation</td>
<td>45-60 min</td>
<td>6 users</td>
</tr>
<tr>
<td>Phase 3: Field trial 1</td>
<td>User evaluation of EMA system in the field (feasibility test)</td>
<td>4 weeks</td>
<td>21 users</td>
</tr>
<tr>
<td>Phase 4: Field trial 2</td>
<td>Response rate comparison of two versions of EMA system</td>
<td>4 weeks</td>
<td>38 users</td>
</tr>
</tbody>
</table>

Setting and users
The target users of EMA system live within a single city-county area of the Midwest. This study recruited patients aged 35-64yrs and currently receiving care in one of Eskenazi Health’s federally qualified health centers (FQHCs). Eskenazi Health is one of the five largest safety-net health systems in the nation. FQHCs gave us access to obese middle-aged women who had had a provider-referral to the Healthy Me program. A primary care provider may refer a patient with a BMI of 30 or higher to meet with a Healthy Me coach. In this clinic-based program, health coaches counsel adult obese patients and create an action plan for increased physical activity, encourage patients to make healthy food choices with an emphasis on controlling portions of food they consume. Healthy Me coaches are certified in behavior change counseling and fitness instruction, and are present two or more days per week in each of the FQHCs [14].

Phase 1: Exploratory design ideation
This phase involved ideation and design of self-report measures based on existing literature, and user- and stakeholder- requirements. It included: (i) a focus group session with the research team’s primary stakeholders (social scientist, exercise physiologist, visual communication experts, and personal trainers), and (ii) a one-on-one design session with five Healthy Me patients (P1–P5).

The focus group session lasted for an hour and identified the core self-report measures of behavior and situation specific to the project – food/drink consumption, physical movement, users’ location, and social co-presence. The focus group session also resulted in the development of four questions, each representing a single measure (version 1, V1). Questions in V1 were straightforward and required selection from a list of pre-populated response options. For instance, questions on eating/drinking and social co-presence included options yes and no; while questions on location and physical activity included options created by stakeholders from expert knowledge and prior literature.

Following the initial focus group session, we conducted one-on-one design sessions with five Healthy Me patient volunteers (Participant 1 through Participant 5; or P1 through P5). Participants/volunteers to this session were identified through announcements by coaches at the end of HealthyMe classes. The 45-60 minute design session had two parts. During the first part, participants were presented with the problem domain and V1 questions created during the focus group session. This was followed by a brainstorming phase, where participants helped identify components for version 2 (V2). Each participant was compensated with a $25 gift card at the end of a session.

Fundamental principles of human–computer interaction informed us in designing an interactive prototype to support a user in responding to V2 EMA questions. As such, differences between V1 and V2 included refined and context-based response choices with images, icons, and a time component to every question to aid users’ recall from within a referenced timeframe. Specifically, V2 questions included: (i) context-based suggestions for location in addition to search option (e.g., pre-populated options to choose from based on device’s current location), (ii) simplified response choices for physical activity, (iii) detailed response choices for eating behavior to help users provide more detailed data (e.g., meal or snack as opposed to a yes), and (iv) detailed response choices for social co-presence.

Phase 2: In-lab evaluation studies
In this phase, six female Healthy Me patients aged 35-64yrs (50% African American, 50% non-Hispanic White) recruited through snowball sampling evaluated the second version of measures. We refer to these participants as P6–P11. Each session lasted 45–60 minutes where participants qualitatively evaluated and provided feedback on V2 questions delivered through an interactive, mobile prototype. Findings from this session resulted in the development of version 3 (V3) questions.

Some of the findings from in-lab evaluation session included mixed support for image icons placed next to each response option and the need for simplified question and response options. For instance, while some participants found the prototype intuitive, others were confused with associating an image icon with the response option. This led to an executive decision on rolling
back or excluding image icons since the immediate project scope was to design a mobile EMA tool supporting the core qualitative findings from Phases 1 and 2. Overall, differences between V2 and V3 questions included excluding image icons, refined questions with fewer words and a timeframe reference, and refined response choices that allowed user to maintain privacy. Specifically, V3 questions included: (i) response choices that helped users select the type of location as opposed to providing the actual physical address, (ii) simplified response choices for physical activity that focused on users reporting whether they had walked or not, (iii) simplified response choices for eating behavior that directs users to consider everything other than drinking water as an eating event, and (iv) means to either pre-populate social connections with names or relationships.

**Six Key Findings of Phases 1 and 2**

Phase 1 (ideation) and Phase 2 (in-lab evaluation) work produced the six qualitative themes below that guided our design decisions on questions and response choices between versions V1, V2 and V3 (Figure 1).

**First, privacy:** most participants disliked the idea of sharing their location with a technology solution:

“My husband is very paranoid when it comes to technology and often says to me that someone is watching us through it. I would rather not share my exact address with the app.” – P3

Participants attributed this feeling towards lack of trust as to where their data were stored and who had access to them (e.g., “I don’t know who is going to look at this data. It is like I am being policed” – P1). Such participants suggested reporting an approximate location by selecting names associated with the location’s address. For example:

“I don’t mind saying I am at a restaurant as opposed to reporting I am at the McD on the corner of 16th street, if you know what I mean.” – P11

**Second, active or sedentary:** Participants reported difficulties in identifying if/not they had been moving enough to self-report physical activity while responding to an EMA question. For example:

“If I were sitting down but moving my arms up and down, wouldn’t that still count as movement?” – P4

A few participants valued and thought an interface would be more intuitive and natural if the question included images depicting the action that was of interest. For example:

“It would be easier for me to recall what I was doing when I see a picture of someone performing the activity, like a walking figure.” – P2

On the other hand, some participants reported they would be confused if they saw images with actions that represented an unfamiliar action (“I see a picture of a figure standing on one leg with the text other movement. Is this yoga or dancing? Perhaps something else?” – P10).
Participants felt strongly about the need for a simplified and direct list of options from which to choose when reporting activity (e.g., “…if all the app wants to know is if I moved or not, why not just ask that as opposed to identifying in detail what activity I was performing?” – P2).

**Third, reporting food/drink consumption:** Participants were very diverse in the ways they thought about and wanted to report food or drink consumption. While some thought water would fall under food/drink self-report, others considered water being independent of any other food or drink they consumed. For example:

“I wouldn’t want to report drinking water. I am drinking water all day. The app would think I am eating or drinking something all day.” – P5

In such cases, participants suggested improving clarity on the question to help remind them not to think about water while self-reporting food/drink consumption (“…you could re-word the question to include did you eat or drink something other than water.” – P8).

Several participants suggested improving their recall ability by including more response options. For example:

“It would be more easier if the question listed out options like snack or meal for food and maybe drink with and without calorie to help me remember. Yes or no is fine, but it would help if I saw better response options.” – P6

However, a few other participants felt they would have a challenge identifying and associating a category to whatever they had consumed. Several participants doubted the accuracy of this technique. For example:

“I think half a sandwich is a snack, that is just me. But I am not sure if that is a snack according to your app.” – P9

**Fourth, who is around me:** Participants discussed the need for maintaining privacy of their social connections while using the app. Some felt strongly about not sharing the name of the person they were spending time with when responding to the EMA questions since they thought they were violating someone else’s privacy without their consent (e.g., “I don’t want to tell a technology solution Sue is with me without her consent.” – P4). In such cases, participants suggested means for the app to allow selecting a relationship name as opposed to a person’s name (e.g., “Why don’t I say my sister is with me instead?” – P4).

Several participants reported increased desire and flexibility in pre-populating the names or relationships of most people with whom they interacted. This pre-populated list was used to create response options for the social co-presence question. Participants appreciated reporting “other” in instances the response options did not hold a specific name of a person or relationship (e.g., “…it is easy to choose other when I don’t see the name of the person who is with me when I am responding to the EMA question.” – P7).
Fifth, quick interaction: All participants strongly preferred selecting a value through a single-tap (e.g., “I like how I tap and can move onto next question. This is really quick” – P8). That being said, some participants discussed instances when they thought single-tap to move onto next question could be error-prone. For example:

“This is really quick. But, what would happen if I tapped yes to the eating question by mistake? Can I go back and re-do the selection?” – P11
Figure 1. Three versions of EMA questions resulting from iterative design process
Sixth, time-based reporting: Several participants expressed concern towards disruptions at times they prefer not being disturbed (e.g., while asleep). For example:

“Will I be receiving these messages during the night? I don’t think I can respond. Besides, this would be so disturbing. It would be nice if I received messages when I am awake and don’t want to be disturbed.” – P2

All participants felt strongly about the need to restrict the timeframe based on which they recalled before self-reporting. For example:

“I think it will make more sense if the question focused on a specific time. If the app asked me what I did in the past 15 or 20 minutes, I could mentally compute and report accordingly.” – P6

The other half of participants rejected time-oriented interface and strongly preferred a more relaxed qualitative instrument (e.g., “It is hard to know what I was doing exactly 15 minutes ago” – P8) or something they can tune to their own subjective experience. These participants explained that the more relaxed timeframe such as since last time they responded would enable them to be more accurate in remembering from the last time they responded. For example: “I can use the last time I responded as a point from when I can remember what I did as opposed to missing out all that information I wouldn’t have possibly reported if I limited to last couple of minutes.” – P9

Five Design Constraints and Decisions
The six qualitative findings from the exploration (described above) and in-lab evaluation phases were coupled with self-report and mobile usability literatures [15-24] to create the five core design constraints (described below) that guided our design decisions for phases 3 and 4 (Table 2):

1. Design to support rapid recall – a significant portion of phase 1 and 2 findings included discussing the need to reduce effort during recall. First, participants discussed the need for aiding users in rapidly recalling their behavior based on simplified response choices. Second, while the stakeholders discussed the need for capturing responses that represent behavior in the time between two EMA prompts, the Healthy Me patients expressed confusion in determining a time window they had to use as reference while performing a recall. Discussions included confusion when users skip/miss EMA prompts and the need for a time window as reference to help recall information. For instance, P1 stated – “What would happen if I didn’t respond all day and see a prompt at the end of the day?” and P3 stated, “Do I respond to the question considering what I was doing at the moment I received it or am I responding based-off of something in the recent past?”

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Design decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Design decisions based on constraints
<table>
<thead>
<tr>
<th>Design to support rapid recall</th>
<th>Rapid recall can be supported through the provision of a reference timeframe where users can perform recall by focusing on the time window.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design for low effort from user</td>
<td>A system that is capable of sending self-report questions as a group of message notifications can allow users to respond on the go, with a single tap, on a mobile device. Burden can be further reduced if the group of messages is dependent on the users’ context. For instance, the EMA system can skip asking about food/drink consumption if the user is physically located in a restaurant when a self-report message group is sent. Or, the EMA system can provide suggestions for location based on the device’s location to help users from searching an actual address.</td>
</tr>
<tr>
<td>Design to capture situations that accompany a behavior</td>
<td>To ensure a response is captured close to a situation, the EMA system should not allow users to respond after a set number of minutes have passed because it is likely the context of the user had changed over time. To support this need, participants suggested a 30-minute window for capturing responses and context – that is, a notification with a question disappears from the user screen if the user does not respond within 30 minutes of receiving the notification.</td>
</tr>
<tr>
<td>Design to capture better quantity and quality of data</td>
<td>To maximize data capture, the EMA system can prompt response at times when the user is awake and does not want to be disturbed. One way to personalize this experience for a user is to have an onboarding process where the user can set preferences for when they will typically be available to receive EMA questions (Figure 2.) One way to capture data with improved quality, specifically for eating behavior is to include response options where users can choose the type of food or drink. Similarly, social co-presence can also include detailed response choices that users can pre-populate with their social connections prior to EMA usage and select those when prompted for social co-presence.</td>
</tr>
<tr>
<td>Design for user’s safety</td>
<td>A notification can be held back if the EMA system identifies the user is moving in a vehicle. This avoids putting the user in danger.</td>
</tr>
</tbody>
</table>

2. **Design for low effort from user** — self-report can be a burden given the need to respond multiple times a day and while in the midst of daily activities. The need to respond to a notification can force users to interrupt an ongoing activity, in turn, leading to increased burden/frustrations. This is true especially when the task of responding is effortful. Given this, it
was a critical need for the system to require minimal effort while responding to a question post-recall.

3. **Design to capture situations that accompany a behavior** – it became clear from the ideation sessions that users are present in dynamically changing contexts of time, space, activity, and social connection. Hence, the EMA system should not only capture self-report measures, but also the context or situations that entail particular user responses.

4. **Design to capture better quantity and quality of data** – both phase 1 and 2 participants/users (whatever term you decide to use) pointed to the need for capturing better quality in addition to quantity of data. Stakeholders discussed the need for capturing detailed information on social co-presence to help identify patterns in eating or movement behaviors specific to social connections. Similarly, several participants from the one-on-one design sessions expressed concern about lack of detail in reporting eating behavior. As such, all participants had differences in perception in determining when they have had something to eat or drink or not. For instance, P2 stated, “If I had a piece of candy do I still report yes to this question? I don’t think that is eating really.” while P4 stated “I usually know when I have eaten a meal, which can be a bigger portion size, as opposed to something like a small snack. It would be nice if I can report clearly what I had to eat because otherwise the system is going to think I am eating something all day.”

5. **Design for user’s safety** – participants from both phases raised the issue of impacting user safety in instances the user is required to but is unable to respond to message notifications. Hence, there is a need to design for safety and to avoid penalizing a user for a non-response at unsafe moments (e.g., while driving).

**EMA System: V3-simple**

From the six qualitative findings and five design constraints, an interactive and functioning EMA application was designed and developed to run on an Android smartphone. To achieve maximum timely awareness and context learning, this system also included an onboarding process (Figure 2). The onboarding process included users performing a one-time setup by selecting approximate times when they woke up, ate, slept, and wished to not be disturbed. We envisioned this information to guide the timing of EMA questions for each participant.
The EMA system was programmed to begin sending EMA questions the day following the onboarding setup to ensure complete capture of data in a day. Any EMA question was designed to mimic a notification as identified by Google Material Design\(^1\) and consisted of components such as header, content, and action buttons. Further, EMA questions were developed to always appear first in the device’s notification drawer (Figure 2) and included playing a sound clip (device’s default for receiving a notification) every time a question was received on the device.

\(^1\) https://material.io/guidelines/patterns/notifications.html
The system was programmed to send EMA questions in groups during every attempt. Overall, 12 attempts were pre-scheduled at the beginning of the day for every participant. Each attempt was timed to occur within the waking time and not during sleep or do not disturb times as selected by a participant during onboarding process. By default, each message group was comprised of the four EMA questions regarding location, social presence, eating, and movement behaviors. The system however, removes EMA questions from a group based on the participant’s context, such as device’s movement and location. At instances when the system is unable to locate the device, the participant’s response to the location question is used to present appropriate follow-up EMA questions specific to that context. The logic rules that guide the system and ultimately user experience are shown in Figure 3.

![Flowchart depicting logic used to identify the EMA questions in a packet](image)

**Phase 3: Field Trial 1**

This field trial included testing the feasibility of the EMA system (V3-simple), with a focus on participants’ understanding of EMA questions and response choices. Prior to the trial, research assistants completed an in-home baseline assessment that captured socio-demographic and body-mass index data, and the New Vital Sign score for health literacy [25] data (see Table 3). During the 4-week trial, twenty-one HealthyMe patients (P12–P33) received twelve EMA message groups per day, every day based on their onboarding times. The actual EMA questions for each group were determined based on the phone’s context and participant responses to the location question. Each EMA question behaved similar to a SMS notification in a smartphone, with a sound alert every time a question was received on the phone. This field trial included gathering qualitative feedback on question and response choices, overall satisfaction [26] and perceived

---

**Other rules in place**

- 12 EMA attempts will be made/day
  - Semi-random timing, 3 attempts for each period of the day
  - Use user’s onboarding responses to wake up, sleep, do not disturb times while scheduling
  - No two attempts at sending EMA is longer than 60 minutes
- EMA is not sent if the device speed is more than 4mph
- With anyone’ question will not be sent if EMA attempt x1 AND device’s location is identified as home AND user lives alone
- If device’s location has not changed over 50m, the same set of questions as previous EMA attempt will be sent, excluding the question question
interface quality [27]. Table 3 shows the characteristics of participants in Field Trial 1 as well as Field Trial 2.

Table 3. Descriptive statistics for participants in field trials 1 and 2

<table>
<thead>
<tr>
<th></th>
<th>Field Trial #1 n=21</th>
<th>Field Trial #2 n=38</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>Mean ± SD</td>
<td>52.2 ± 7.6</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African</td>
<td>16 (76.2)</td>
<td>33 (89.2)</td>
</tr>
<tr>
<td>American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>5 (23.8)</td>
<td>4 (10.8)</td>
</tr>
<tr>
<td><strong>Number Household</strong></td>
<td>Mean ± SD</td>
<td>2.2 ± 1.5</td>
</tr>
<tr>
<td><strong>Household Income</strong></td>
<td>Mean ± SD</td>
<td>20,480 ± 23,022</td>
</tr>
<tr>
<td><strong>Education Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>7 (33.3)</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>College/University</td>
<td>14 (66.7)</td>
<td>22 (59.5)</td>
</tr>
<tr>
<td><strong>Work Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>17 (81.0)</td>
<td>14 (37.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>4 (19.0)</td>
<td>23 (63.9)</td>
</tr>
<tr>
<td><strong>Hours Worked</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>1 (25.0)</td>
<td>13 (36.1)</td>
</tr>
<tr>
<td>11-20</td>
<td>1 (25.0)</td>
<td>2 (15.4)</td>
</tr>
<tr>
<td>21-30</td>
<td>1 (25.0)</td>
<td>11 (84.6)</td>
</tr>
<tr>
<td>31-40</td>
<td>1 (25.0)</td>
<td>7 (53.8)</td>
</tr>
<tr>
<td><strong>Shift Work</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Shift (6-8am)</td>
<td>2 (50.0)</td>
<td>7 (53.8)</td>
</tr>
<tr>
<td>2nd Shift (2-5pm)</td>
<td></td>
<td>3 (23.1)</td>
</tr>
<tr>
<td>Varies</td>
<td>2 (50.0)</td>
<td>3 (23.1)</td>
</tr>
<tr>
<td><strong>Weight (lbs)</strong></td>
<td>Mean ± SD</td>
<td>257.0 ± 57.6</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td>Mean ± SD</td>
<td>45.5 ± 8.4</td>
</tr>
<tr>
<td><strong>Low Health Literacy</strong></td>
<td>No</td>
<td>11 (52.4)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>10 (47.6)</td>
</tr>
</tbody>
</table>

This trial was also used to determine participants’ response rate to the EMA system. A common finding on reviewing quantitative data suggested that all participants who began responding to the first question in a message group continued to complete all the questions in that group. Hence, the presence of a response to the first question within an EMA message group was used to calculate a participants’ response rate. That is,

\[ \text{Response rate} = \frac{\text{Total number of first question responses}}{\text{Total number of EMA packets}} \]
Based on this formula, the overall average response rate for participants in Field Trial 1 was identified as 50.3%. The mean weekly response rates were consistently close to 50% (Figure 4). At end of 4 weeks, participants reported on a Likert scales (1=strongly agree to 7=strongly disagree) delivered by research assistants and captured overall satisfaction, perceived likability and pleasantness of the smartphone app. Data captured on paper-based survey forms were transcribed and stored to a secure database. As such, participants reported being satisfied (M_{Satisfaction} = 1.3) and expressed pleasantness and likability (M_{Interface quality} = 1.4)

![Field Trial 1: Mean Weekly Response Rates](image)

Figure 4. Mean weekly response rates for participants using V3-simple in Field Trial 1

A common qualitative finding from this trial included participants’ misunderstanding of EMA questions, especially when they perceived a group of questions to be related to one another. For example:

“When I saw the ‘with anyone’ question after the eating questions, I thought the system was asking me if I was eating with someone.” – P30

Feedback from the Field Trial 1 also included the need for balancing repetitive reminders vs. interruptions that may be caused by an ongoing activity. While some participants suggested re-sending an unanswered EMA question every few minutes as a reminder (e.g., “It would help if the phone dinged after a couple of minutes to remind me in case I missed hearing the first time” – P18), others discussed their inability to respond irrespective of the reminder since they cannot interrupt the ongoing activity (e.g., “…it won’t matter sending reminders really. My phone is not with me if I am busy doing something.” – P20).
As such, we identified an additional design constraint for the EMA system, namely, **designing for increased engagement**. One way to increase user engagement is through auditory reminders. We proposed playing an auditory alert (sound) on the mobile phone two times in addition to the instance when it is played on receipt of an EMA question. That is, the phone will play an audio sound indicating that the first EMA question in a packet is awaiting a response when a phone has received the question but the user has not responded. Specifically, the EMA system was programmed to sound an alert every 10 minutes if the mobile phone has an unanswered first EMA question in its notification drawer. As such, in Field Trial 2 we tested the hypothesis that sounding an alert every 10 minutes in a 30-minute window of sending the first EMA question would increase response rate as opposed to only sounding an alert once when the first EMA question is sent to the mobile device.

**Phase 4: Field trial 2**

This field trial included testing the hypothesis from Field Trial 1 for its effect on response rate. We used two versions, V3-simple and V3-ding for use by two groups of participants. Thirty-eight HealthyMe patients (P34–P72, Table 3) were randomly assigned to use either V3-simple or V3-ding for 4 weeks. During the 4-week field trial, both groups of participants received twelve EMA prompts daily based on their onboarding times. Similar to Field Trial 1, the actual EMA questions received by a participant were dependent on their phone’s context or their response to the location question. All participants were instructed at study setup to consider each question independent of the other while responding to overcome the confusion experienced by participants in Field Trial 1. While participants using V3-simple received a single auditory alert, participants testing V3-ding received two additional auditory alerts when the first EMA question in a packet is received by the mobile phone. The two sound alerts were spaced to occur at 10-minute intervals when a participant did not respond. Both versions included the same group of questions seen by participants in Field Trial 1. This trial also included gathering qualitative feedback (semi-structured interview or occasional phone conversations where research assistants assisted participants during the trial when needed) on repeated notifications in addition to perceptions on question and response choices either at the end of the trial or during the trial at times participants needed assistance.

We calculated response rate using the formula identified in Field Trial 1. Response rates of the participants using V3-simple were compared between trials 1 and 2 by means of a linear model with independent variable representing trial 1 vs 2. In this model, mean response rates of participants using V3-simple between trials 1 and 2 were not significantly different ($p = 0.1550$).

We found a statistically significantly higher response rate for V3-ding respondents. However, a greater percentage of participants worked in the V3-ding group as opposed to those in V3-simple group ($p = 0.0291$). Hence, the statistical test comparing the response rates between V3-simple and V3-ding groups in Field Trial 2 was estimated again with an adjustment for work status. This adjustment did not change the findings; there was a significantly higher response rate for participants in V3-ding group than in the V3-simple group (Table 4).
Table 4. Estimated Means for response rate adjusted for Group and Work Status

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
<th>Difference</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>V3-simple</td>
<td>40.35</td>
<td>4.66</td>
<td>0.0097</td>
<td>-19.96</td>
<td>(-34.76, -5.16)</td>
</tr>
<tr>
<td>V3-ding</td>
<td>60.32</td>
<td>5.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Qualitative reports on V3-ding indicated that some participants considered repeat sound alerts as non-disruptive while others perceived this feature as a drawback. For example:

“It is not a big deal.” – P45 and “It felt like the system was eating up my phone’s battery.” – P60

Overall, participants expressed disappointment at trial completion since they felt the system helped increase their awareness of eating or movement behavior. For example:

“I was disappointed it was over, it was like someone was there checking on me. I have paid more attention to what I ate like it asks about eating or drinking, then if I hadn’t I’d make a healthy choice.” – P38

Another common finding included participants being confused between whether/not they had to consider drinking water as a drinking activity. For example:

“Do I select yes when I have had water to drink?” – P38 or “Water is a drink too. I reported yes every time I had just had water when I received eat question. I hope the system doesn’t think I am always eating or drinking something” – P42
Discussion
Our aim in this research was to identify constraints and design parameters for a mobile EMA system capable of capturing self-report measures of eating and movement behavior coupled with participants’ location and social contexts. We view the capture of these and similar health-related behaviors as they occur in everyday contexts as critical to progress in precision health research and interventions. By working with a diverse set of individuals, we were able to design an effective reporting interface. Following an iterative exploration and ideation process, we identified constraints that users and stakeholders reported to be particularly important. We evaluated the interface and the self-report measures through two field trials. While the first trial helped identify feasibility, the second trial tested a design feature that improved user engagement. Our interface was well received with high satisfaction and interface quality ratings.

Our target users were female patients with obesity seeking care in a safety-net health system; most had very limited household resources or technology experience, and many had scores indicating low health literacy. Achieving a satisfactory interface for real-world, real-time, self-reports of weight-related behaviors and context required significant time with users and many design iterations and prototypes over an 6-month period. From this extensive effort, we suggest that future work attempting to conduct momentary assessments achieve the following minimum specifications to support increased user engagement: 1) an onboarding process to personalize the times when an assessment is delivered to a participant, 2) question pruning through passive sensing of device location to present only questions most relevant to a participant’s context, 3) increase participant’s attention with a notification that is prioritized over any other application notifications in the receiving device, 4) limit the time-window within which a participant can respond to a question to capture situations accompanying a behavior, and 5) repeat auditory alerts to remind participants to respond. Specific to reducing burden while capturing a user response, we suggest designing a system that: 1) supports tap-interaction to record a response, 2) uses simple-worded, direct questions with fewer words that are easier to read and quicker for the participant to understand, 3) simple response options that are easier to read, quicker for the participant to understand and select from, and 4) includes instruction as to whether/not a participant has to consider each question in the assessment independent of one another while selecting a response.

Overall, we highlight that personalizing or tailoring the EMA experience to an individual can have several positive outcomes. First, customizing the times for EMA prompts to suit individuals’ daily routine can increase adherence. For instance, participants are more likely to respond to an EMA question at times they do not want to be disturbed. Second, personalization can help only present EMA questions that are most relevant to an individuals’ situation or context. For instance, our system can automatically identify an individual’s location from their smartphone and present only the most relevant EMA questions. This in turn can reduce response burden and better match a respondent’s cognitive process for recall and self-assessment. Finally, our solution involving tap-to-report interaction highlights a design solution for reducing response burden.
One potential side effect of EMA indicated in existing literature (cite) is that EMA systems may serve as an intervention and not just a measurement tool. We heard some feedback supporting/suggesting that this may be the case. For instance, several participants expressed becoming more aware of their eating and movement behavior after using our EMA system. In other words, while participants were self-reporting eating or movement behavior, they were also being nudged to pay close attention to these self-behaviors. To what extent this affected behavior or weight over time would require work outside of the scope of this report. Relatedly, it is also outside the scope of this report to carryout validation of the EMA data. Future work could investigate the validity of EMA reports of moving, as this can be compared to accelerometry data. Social co-presence could theoretically be validated by corroboration of the co-present individual. In the case of eating, validation would be more difficult as there is no objective measure of eating beyond direct observation however we do have work in the field using 24 hour dietary recall and EMA eating questions in the same time window.

With full appreciation of the potential limitations, we are intrigued by the possibilities of this EMA platform in three broad areas of future work. First, if asking an individual about a behavior frequently and in varied context has implications for the practice of that behavior, broad use of this EMA platform could result in an extraordinarily low burden, low cost, highly scalable intervention tool. Thus, we believe the impact of the EMA reporting process on behavior, particularly weight-related behavior in our case, deserves careful investigation. Second, expanding this EMA platform to provide contextually appropriate feedback, whether social or machine, is an interest of ours. Keeping to the design principles learned here while developing easy, simple positively reinforcing feedback could enhance any effect EMA reporting may be having on behavior. Third, the use of this tool to measure behavior and context for other projects and programs is significant as precision health advancements may well depend on EMA. Examples in the literature point to exciting possibilities for work in smoking cessation, [28] drug abuse and recovery, [29-31] and mental health [32] to name a few. We are hopeful that ongoing work in these and other areas could lead to a new generation of behavioral data and interventions.

References


12. (Tull and Ebner-Priemer, 2009)


25. Weiss et al, 2005


