HHS Public Access

Author manuscript

Proc ACM Int Conf Ubiquitous Comput. Author manuscript; available in PMC 2018 September 18.

Published in final edited form as:

Proc ACM Int Conf Ubiquitous Comput. 2016 September ; 2016: 1124–1128. doi: 10.1145/2971648.2971717.

μEMA: Microinteraction-based Ecological Momentary Assessment (EMA) Using a Smartwatch

Stephen Intille, **Caitlin Haynes**, **Dharam Maniar**, **Aditya Ponnada**, and **Justin Manjourides**
Northeastern University, Boston, MA

Abstract

Ecological Momentary Assessment (EMA) is a method of *in situ* data collection for assessment of behaviors, states, and contexts. Questions are prompted during everyday life using an individual's mobile device, thereby reducing recall bias and increasing validity over other self-report methods such as retrospective recall. We describe a microinteraction-based EMA method ("micro" EMA, or μEMA) using smartwatches, where all EMA questions can be answered with a quick glance and a tap – nearly as quickly as checking the time on a watch. A between-subjects, 4-week pilot study was conducted where μEMA on a smartwatch (n=19) was compared with EMA on a phone (n=14). Despite an ≈8 times increase in the number of interruptions, μEMA had a significantly higher compliance rate, completion rate, and first prompt response rate, and μEMA was perceived as less distracting. The temporal density of data collection possible with μEMA could prove useful in ubiquitous computing studies.

Keywords

Microinteractions; smartwatch; ecological momentary assessment; experience sampling; compliance; H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION

An important challenge in health, ubiquitous computing (ubicomp), and other human-centered disciplines is better measurement of behavior, state, and context in natural settings. Developing novel health technologies, for example, requires valid and reliable measures of behavior for surveillance, epidemiological, and intervention studies. In research related to human behavior, concerns over the validity of retrospective self-report due to recall biases have led to an increase in the use of technology to measure behavior, such as so-called "objective" measures that use electronic devices to measure a behavior directly. For instance, in physical activity research, activity and heart rate monitors can measure body movement without self-report. Nevertheless, despite ongoing research in objective measures, self-report is still required to measure many behavioral constructs known to impact health, such as stress, emotions, diet, pain, and fatigue [45]. In addition, self-reports can be used to fill-in missing contextual data that objective measures do not capture (e.g., [10]).

s.intille@neu.edu, cahaynes1202@gmail.com, {maniar.d, ponnada.a}@husky.neu.edu, j.manjourides@neu.edu.

Most self-report data collection is temporally-sparse because of the high burden of obtaining information. In ubicomp research, where researchers are interested in *in situ* deployments and evaluation of real-time interventions, *temporally-dense* self-report may be particularly valuable (e.g. [12, 28]). For example, suppose a researcher develops a system to detect an activity such as "eating" (e.g., [1]) or "smoking" (e.g., [47]) passively from wearable sensors, and then validates the technology in lab. One way to subsequently validate this system during real-world use would be to use self-report prompts on a wearable device delivered at a high temporal density asking questions about the person's activity (e.g., "Are you eating? Yes/no", "Are you smoking? Yes/no"). One concern about this approach might be unacceptable burden from frequent interruptions.

In this paper, we present a new data collection methodology for studying human behavior: microinteraction-based ecological momentary assessment, or μEMA. Standard ecological momentary assessment (EMA) [35 – 37, 44], also known as experience sampling [3, 23], is a widely used method to study subject-level effects of time-varying phenomena in user interface design, health behavior studies, and other fields where gauging behaviors of people *in situ* is important. In EMA, a person's mobile phone beeps and presents a set of multiple choice questions about momentary behavior. Standard EMA, however, is burdensome on a phone and cumbersome to use on smartwatches with small screens. We have implemented a modified version of standard EMA, i.e. μEMA, where all prompted surveys are reduced to fast, glanceable "micro-interactions" [2] that can be answered within just a few seconds. This requires changing the nature of how questions are prompted and asked, and it is best achieved using a smartwatch. The market share of wearable devices is predicted to reach 200 million units by 2019 [20], with smartwatches alone reaching ≈90 million units. This growth may support the use of μEMA delivered on smartwatches for future *in situ* data collection.

We posit that this strategy of using frequent, but less burden-inducing, microinteractions delivered via smartwatches may permit collection of temporally-dense longitudinal data for health, ubicomp, and other fields where understanding contextual changes in behavior is important. We have assessed μEMA's feasibility by measuring compliance, completion, response latency and perceived burden of 33 participants in a 4-week study, and compared it with standard EMA on a phone. Our goal was to explore whether μEMA, with a high interruption rate that might initially seem unsustainable, would be no less tolerable than common EMA, currently extensively used.

BACKGROUND

In behavioral and health sciences, *in situ* data collection methods have played an important role in understanding individual behaviors. The roots of these methods stem from diary studies and retrospective self-reports. Retrospective self-reports, however, capture only "snapshots" of behaviors, and they provide little information about the dynamic nature of behavior and psycho-physiological processes that occur throughout the day. Understanding the *dynamics of behavior* throughout weeks, days, and even minutes may be critical to developing better ubicomp technologies, advancing the science of behavior change, and developing new "just-in-time" interventions that take advantage of moment-to-moment tailoring [31, 40, 43]. Major new health surveillance initiatives might benefit from low-

burden measurement methods that gather longitudinal behavioral data on massive cohorts of people (e.g., [5]). Various transdisciplinary health researchers have identified the lack of intensive longitudinal behavior measurement systems as a key barrier to scientific discovery and intervention design [17, 29, 32, 33, 42].

Advantages of Ecological Momentary Assessment

EMA is widely used to gather information on the purpose of behavior, social context, subject state, or causality that might help inform science and the development of successful ubiquitous systems that respond to behavior [35]. EMA has four beneficial characteristics [46]. First, it reduces recall biases and errors, capturing people's behaviors instead of retrospective memories/beliefs about their behaviors. Second, EMA occurs in the natural environment, thus increasing ecological validity. Third, multiple assessments occur over time, so that short-term shifts and temporal dynamics in specific contexts can be examined. Fourth, EMA, and especially context-sensitive EMA [21], can be used to learn about behavior as it happens and then use that data to intervene in 'real-time,' i.e., in just-in-time, adaptive interventions (JITAIs) [41].

EMA Limitations

The three most important limitations cited for EMA are cost, reactivity, and burden. Cost has plummeted due to widespread proliferation of mobile phones [22]. Reactivity has found to be small [6], although it is a topic that warrants more research. The Achilles heel of EMA is the interruption burden. This burden is especially disruptive if researchers desire to use this technique for longitudinal measurement in order to gather data at a high temporal density. Researchers employing the technique know that compliance will drop quickly – often within a week – if questions are too long or too frequent (e.g., [7, 14]). Researchers attempt to offset this interruption burden by making surveys short. However, standard protocols used in health or psychology often have 8 prompts a day, with question sets up to 36 items that take 1-2 min to answer (e.g., [19]), leading to EMA compliance and sustainability challenges [18]. Without high compliance, the goals of any EMA-based study are called into question because of selective non-response bias. Unfortunately, the more temporal density we wish to achieve, the higher the possibility of imposed burden on the participants because of the frequent task interruption from their devices prompting them (and therefore interrupting them). These interruptions are most burdensome when the prompting device is not instantly accessible. This could prevent data gathering that might support development of computational behavioral models supporting real-time interventions using individual-level intensive longitudinal data, sometimes called "small data" [8]. Despite smartphones being in the same room with users 90% of the time, they are within hands reach just 50% of the time [9]. This will contribute to burden with smartphone-based EMA.

Typical EMA delivered via smartphone, therefore, has two problems that may limit the temporal density of data collected: (1) the amount and length of interruption, and (2) the difficulty of accessing the device (thereby increasing the burden of each interruption further). To address these concerns, we propose using a variation on EMA where interruptions can be much more frequent because prompted questions can always be

answered extremely fast, thereby leading to an acceptable perceived burden when data are collected multiple times per hour.

MICRO-EMA: OVERVIEW

"Microinteractions" are brief interactions with an interface that take just a few seconds to start and complete [2]. A microinteraction is akin to glancing at a watch – so short that it does not significantly disrupt any ongoing activity. A microinteraction has two phases – device *access time*, and interface *usage time* [2]. Access time refers to the time it takes to access a device to start a task. Usage time refers to the time it takes to complete that task. An ideal microinteraction's access plus usage time is short enough to be perceived as a non-disruptive interruption (even when in the midst of another activity, such as a conversation). μEMA uses a smartwatch to minimize access time [2]. Usage time can be minimized by enforcing single questions (rather than sets of questions), each with a small set of responses (e.g., 3) that easily fit on a smartwatch display. By guaranteeing to the user that each interruption is always answerable in a microinteraction, μEMA removes user uncertainty about how long responses might take – *every* prompt can be answered in just a few seconds. This guarantee could reduce user hesitation to engage with the system, even at surprisingly high interruption rates, keeping perceived burden lower than might be expected with traditional EMA.

μEMA on Smartwatches

We implemented a μEMA smartwatch system using Moto 360 Android smartwatches (paired with an Android phone running Android 4.3+). The Android ecosystem was selected for the study because it allows precise question timing and logging, as well as use of phone sensor data (for future studies). When the μEMA software is installed on the accompanying smartphone, participants receive a prompt via a vibration on the watch running a custom watch app (Figure 1). When they rotate their wrist to look at the display (just like they would to check time), the survey is already displayed on the watch-screen. A question is answered with a single tap – no scrolling is required. As soon an answer is tapped, the question disappears and the interaction is over. Unlike most EMA protocols on phones, μEMA on a watch presents only one question at a time, without exception, guaranteeing to the user that all interruptions can be handled in a microinteraction.

For our pilot study, our concern was not the validity of the information acquired, but rather overall burden resulting from interruption. We therefore adapted EMA questions used in prior work on assessment of physical activity and context from standard EMA to μEMA [11]. The five mood questions used the standard Positive and Negative Affect Survey (PANAS) [51] and addressed how nervous, upset, stressed, excited, and alert participants were feeling when the phone prompted. In EMA, each multiple choice question was answered using a 5-point rating scale ("not at all," "a little," "moderately," "quite a bit," and "extremely"). One additional activity question asked the participant if s/he was doing one of three activities (selected randomly): sitting, lying down, or walking. The activity question was answered with "Yes," "No," or "Sort of." The 5-answer EMA questions, did not fit well on a watch screen without introducing scrolling (slow), small fonts (unreadable for middle-

aged individuals without reading glasses), and/or small buttons (difficult to tap). We also wanted to ensure answering each question was cognitively simple and that all interruptions be answerable with a microinteraction. Therefore, we converted the PANAS EMA survey to μEMA as follows. First, only one question appeared at a time instead of 6 questions back to back. Second, questions with five answer options were broken down into two questions with three answer options. For instance, a standard question on "How excited are you?" with five options was broken down as "Feel excited right now?" with options "Yes", "Just a little" and "No". If users tapped "Yes", a follow up question was scheduled – some time between 3-20 min in the future, at which time another question was prompted: "How excited are you?" with options "Moderately," "Quite a bit," and "Extremely." Delaying the follow-up questions guaranteed all interactions were single-question, 3-response micro-interactions. If some prompts had one question, and some prompts had two, participants might hear the prompt and assume the worst-case of a 2-question set, which takes longer to answer. A fundamental goal of μEMA is to ensure that when participants are prompted, they know that they only need to answer a single question – without uncertainty [4]. The activity question had the form of, "Are you sitting?" (or "lying" or "standing") with answer options of "Yes," "No," and "Sort of."

μEMA Questions Scheduling

During each 2-hour time-window between 8 AM to 8 PM, 6 to 11 questions were prompted on the watch. The number of questions depended upon how the previous questions were answered. Questions were prompted using vibration, without audio. The vibration used a 'stuttering' pattern that started subtly (hardly noticeable) and increased in intensity at the end, lasting 11.4 seconds. The vibrations occurred regardless of the watch's notification mode.

At the prompt, the question appeared on the watch display and remained for 1 min, or until answered. Each two-hour window was broken into 5, 24-min segments. A question was randomly assigned to one segment and then randomly scheduled at a time within that segment. The remaining questions for that two-hour window were then assigned to other segments in a similar way. The same type of question was never used back to back at any time throughout the day. Once all mood questions were scheduled, one activity question was scheduled during each two-hour segment, randomly at a minute between 0-117. When the second part of a mood question was required, it was scheduled randomly between 3-20 min after the prompt for the original question. If there was no answer to a question prompt for 60 s, the watch re-prompted once. If no answer was received within 1 min of the re-prompt, the question closed and was no longer visible and logged as not answered. If a question was dismissed by the participant by swiping it away or pushing the watch's physical button (i.e., using watch's 'cancel' function), that question was logged as not answered. We hypothesized that μEMA on the watch, where all assessments were single questions that could be answered in just a few seconds but where there were far more prompts, would have lower perceived burden than traditional EMA on the phone, while at the same time allowing the acquisition of similar or complementary information (although validation of the information gathered is left for future studies). μEMA may also permit acquisition of more temporally dense information about behavioral dynamics. In this pilot study, we compared

μEMA with EMA in terms of perceived burden to determine if the μEMA technique may be feasible for a multi-week momentary assessment of behavior.

Increasing Interruption to Reduce Burden?

Interruption burden is the barrier to effective EMA use, and so increasing interruption may seem at odds with reducing burden. Researchers using EMA often focus on reducing interruptions, but then they extend the length of each survey to gather more information at a single interruption. Here we propose an alternative. Interrupt much more frequently, but make the length of interruptions quick and predictable.

The perceived burden of interruption tasks will result from several dimensions of the interruption and associated task, including the frequency of interruption, task duration, content and task complexity, interruption timing (moderated by context and social norms), and perceived value of the task or the information received (moderated by information source) [15, 38]. Devices that use sensors to detect and estimate human interruptibility may offer the possibility of mitigating some of these factors [13, 39]. For situations where data can be obtained without proactive prompting, strategies such as simple lock-screen surveys have been proposed [49]. In most EMA studies, however, the device has no knowledge of the context – it interrupts blindly – and in many validation studies, random prompting is a desired property to simplify statistical analysis.

Researchers are therefore most likely to be able to manipulate frequency of interruption, response task duration, task cognitive complexity, and perceived value (i.e., reward). Week-long EMA studies often report high compliance (e.g., >75%), but researchers see compliance decline quickly due to interruption burden [21]. Researchers typically assume that increased prompting dominates interruption costs and leads to declines in compliance, which is why they reduce interruptions but increase task complexity, and then provide a reward to offset the burden, often in the form of small financial payments. The strategy explored here is to sharply reduce response time and the cognitive complexity of each question, and to make response time predictable, thereby minimizing interruption cost. Response time is reduced by making the prompting device highly accessible using smartwatches [2]. Cognitively complexity of prompts is reduced by breaking them into smaller “Yes/No” answer type questions. Although researchers are beginning to explore the potential of watch-type devices to quantify behavior using EMA e.g., perceived exertion [48], cardiovascular activity [16], depressive mood [24], here we focus on the use of μEMA to make temporally-dense self-report sustainable, so that each interruption contains only a single, simple question.

EVALUATION OF MICRO-EMA ON A SMARTWATCH

To test the feasibility of μEMA, we conducted a pilot study with 33 participants on a typical behavioral data gathering task. Self-reported data on positive and negative affect and activity were collected throughout the day using either the standard EMA PANAS phone survey or the modified μEMA watch survey. All study participants had an Android phone as their personal phone; half of the participants were loaned a Moto 360 Android smartwatch. Participants were asked to respond to all audio and/or vibration prompts delivered on the

phone or the watch. The same amount of information was requested in both conditions (but establishing the validity of the PANAS data acquired is beyond the scope of this paper and left for future work). Our focus in this study was to test the feasibility of running μEMA on the smartwatch, measuring compliance and perceived burden relative to a similar, more traditional EMA data collection task.

Hypothesis 1 (*Compliance*): Participants in the μEMA condition will answer a higher percentage of questions than participants answering the same questions using standard EMA taking all sources of data loss into account (i.e., all questions the participants should have received, regardless of whether devices were turned off, silenced, or disabled).

Hypothesis 2 (*Completion*): Participants in the μEMA condition will respond to a higher percentage of received prompts (i.e., prompts actually delivered via audio or vibration) than standard EMA on smartphone.

Hypothesis 3 (*Response Latency*): Participants in the μEMA condition will respond to the first of the prompted questions more often and quickly after being prompted than participants providing the same data using standard EMA on a smartphone (i.e., suggesting they were less hesitant and/or better able to engage with the application).

Hypothesis 4 (*Perceived Burden*): Participants in the μEMA condition will self-report lower burden, despite receiving far more interruptions than participants providing similar amounts of information using mobile-EMA.

EXPERIMENT DESIGN

This study was approved by the Northeastern University IRB. Participants were randomly assigned to μEMA on a smartwatch and standard EMA on a phone conditions.

Standard EMA on Phone

In the EMA-phone condition, participants were prompted 6 times a day with 6-question question sets from the PANAS. Question sets were asked at a randomly-selected time during 2-hour time windows between 8AM to 8PM. If a question set was not fully answered within 5 min, it was re-prompted once. If no answer was received 5 min after the reprompt, the question was closed and was no longer visible. Since this is a measurement tool used in prior studies, we maintained a design consistent with previous work. Therefore, a persistent progress bar was displayed so that participants could monitor their compliance anytime on their phone [11] (Figure 1). The notification showed the number of prompted surveys the participant had completed or missed that day so far. This capability was retained to ensure maximum compliance among the EMA participants, i.e. we biased our experiment to reject our hypotheses.

Recruitment

Participants were recruited from a major private university and a local school through mailing lists, flyers and word of mouth. These institutions were chosen because it was possible to establish procedures that ensured participants would be accountable for returning

the loaned smartwatches. Inclusion criteria were: (1) employee/student at the university, or an employee of the school, (2) age 18-55, (3) minimum education level of high school diploma, (4) using an Android v4.3+ phone, (5) self-reporting fluency in written and spoken English, (6) self-reporting willingness to allow use of their phone's data allocation for the study, (7) self-reporting willingness to wear a smartwatch during all waking hours, and (8) self-reporting a willingness to charge the watch or phone nightly. Due to concerns about eyesight or prior smartwatch use biasing results, exclusion criteria were: (1) self-reporting use of reading glasses to read a phone, (2) self-reporting regular use of a smartwatch, and (3) self-reporting an intention to switch phones within 4 weeks.

Compensation

Unlike most EMA studies, there was no financial compensation in this study either for participation or achieving high compliance. Participants were only offered the opportunity to try a normally-configured smartwatch for 4 days after completing the study.

Participants

Participants for the pilot study were recruited between December 2014 and June 2015. Eighty-eight individuals expressed interest:15 were ineligible, 32 refused participation, and 3 participated in the pre-pilot. Thirty-eight individuals were randomly selected for EMA-phone (n=18) and μEMA-smartwatch (n=20) arms in the main study. Four participants in the EMA condition and one participant in the μEMA condition dropped out within 24 h. As a result, we have used data from the remaining 33 participants (μEMA n=19, EMA n=14).

Procedure

Day 1: A research assistant met the participant at a convenient location for 15-25 min, obtained informed consent, and installed the software on the participant's phone. Participants in the μEMA condition were loaned a Moto 360 Android smartwatch. ***Days 7, 14, 21***: Each participant received an email with a link to an online survey requesting feedback about the previous week. The email was distributed between 9 AM and noon, and participants were asked to complete the survey within 5 h. If they did not, they were sent a reminder email between 5-8 PM. If the survey was still not received by 10 AM the following morning, the participant was called between 10 AM-10 PM; if the participant was not reachable a brief voicemail was left requesting that the survey be submitted as soon as possible. All surveys were completed. The survey included 11 multiple-choice items and 3 open-ended questions on a participant's experiences answering prompts, perceived burden and disruptiveness (adapted from [26, 27]). Questions also asked about perceptions of ease of learning and using the system, and the length of time needed to respond to prompts. ***Day 28:*** Participants were emailed a final, 20-question online survey about their overall experience. It addressed the same perceptions as the weekly surveys but also asked participants to list the most negative and positive aspects of the system and to offer suggestions on how to make the system less disruptive. ***Day 32:*** After allowing a participant to try a watch as desired for four days, a research assistant retrieved the watch from a convenient location.

Remote Updating and Compliance Monitoring

The μEMA/EMA application sent summary data about participants' use of the app to a secure server each hour, which allowed for real time visualization of participant data. Full log data, including sensor data from the phone/watch and extra logging for monitoring performance, was transmitted nightly. This system facilitated remote study compliance monitoring and proved indispensable in pre-pilot sessions.

Pre-pilot and Iterative Testing

After development and testing with the research team, three participants enrolled in a pre-pilot study to test the study procedures and software, two in the EMA condition and one in the μEMA condition. All three pre-pilot participants felt "mostly comfortable" with using technology. Using the remote monitoring tools, the research team checked data nightly and iteratively fixed problems as they arose. For example, participants reported a bug whereby results of answering surveys were not reflected immediately in the persistent status notification, demonstrating that they were paying attention to that information. Likewise, in μEMA condition, the watch was, intermittently, not prompting at correct times. This behavior was traced to a bug in the Android Wear alarms, which we worked around. At the conclusion of the pre-pilot, all three participants reported feeling that the system was "mostly" easy to use but also reported noticing inconsistencies in how they were prompted, consistent with what the research team had already uncovered. All three participants reported feeling that the questions interrupted them and slightly distracted them from the task they were doing, but none felt annoyed when prompted. After 28 days of testing with the three participants, all known software problems were resolved.

MEASURES

A question set is a set of questions asked back-to-back, at the same time. In the EMA condition, each participant was prompted with 6 question sets a day (with 1 re-prompt each time, if not answered), for a total of 168 expected question sets per participant over the 4-week study and 1,008 expected questions per participant. In the μEMA condition, question sets and questions are equivalent, because only single questions are asked at a time (with one possible reprompt). However, some questions were prompted only in response to answers from previous questions. Every participant, therefore, received a different number of questions each day over the 4-week period. Participants in the μEMA condition could receive a minimum of 1,008 and a maximum of 1,848 questions over the 4-week period.

In both conditions, question sets were occasionally not prompted due to either (1) an unanticipated bug in the software (for pre-pilot participants only), (2) the participant turning the phone/watch off, or (3) the device's battery draining, leading to the device turning off. Each question set was marked as *answered* only if all the questions in the set were answered. In addition to recording the number of question sets answered, the number of questions answered was also recorded. In the μEMA condition, the number of question sets answered and the number of questions answered are identical. Prompts for answering question sets might not be completed for two reasons. They might be ignored by the participant and time out, or they might be dismissed. Participants could dismiss a prompted question by swiping

it away on the watch, or by hitting the back or home button on the phone; these two conditions were programmatically indistinguishable.

Question Set Compliance

Percentage Question set (QS) compliance for the EMA and μEMA sample is defined as follows:

$$QSCompliance\% = (QSAnswered / QSScheduled) \times 100$$

By this definition, a participant who turns off a phone or watch will negatively impact compliance, because prompts will never be received. Participants were told to keep devices charged and nearby.

Question Set Completion

Question set completion is defined based on prompts actually delivered, not just expected.

$$QSCompletion\% = (QSAnswered / QSPrompted) \times 100$$

A participant who turns off a phone or watch will not negatively impact completion, because completion only considers prompts that were actually delivered. Similarly, if a bug in the software prevents a question from being prompted, that too will not negatively impact completion.

Initial Prompt Response Percentage

In both conditions, question sets prompted and unanswered were re-prompted once. Initial prompt response is defined based on *whether a question set was completed after the first prompt* (versus after the reprompt for that question):

$$InitialPromptResponse\% = \frac{QSCompletedAfterFirstPrompt}{QSPrompted} \times 100$$

If the initial prompt response percentage declines, it may signal that participants are tiring of answering them, thereby delaying responding (and making it more likely the second prompt will be ignored or missed as well).

Perceived Burden

Participants reported their perception of interruption, disturbance and annoyance four times during the 4-week study period via an online survey. The questions used to measure burden were: 1) “Did you feel the questions interrupted you?” 2) “Did you feel annoyed when you were prompted?” and 3) “The surveys distracted me from the task I was doing.” The 7 answer choices for these questions ranged from “Strongly disagree” to “Strongly agree.”

RESULTS

Box plots of the response rates for individuals randomized to each arm and separated by overall compliance, completion, and initial prompt response are presented in Figure 2. Based on these plots, we identified two subjects in the μEMA arm with response rates consistently more than 1.5 interquartile ranges below the 25^{th} percentile response rate. Because these subjects differ substantially from the bulk of the data, we followed common practice and considered them as outliers for further analysis [30]. Additionally, these outliers are known, based on exit interviews, to have not kept their equipment with them and functioning. Table 2 highlights response behavior with and without outliers in the μEMA condition.

Comparing Overall Compliance of μEMA and EMA

The objective of this analysis is to determine if subjects randomized to the μEMA condition are more likely to respond to prompts than those in the EMA condition, taking into account all prompts that should have been delivered. The intended number of prompts may differ from the received prompts due to factors such as the devices being powered off or software being disabled. Because each individual is prompted several times, we cannot simply compare the rate of responses between the treatment and control groups; we must account for the fact that the trials are repeated within each study subject. Let N_i be the number of prompts scheduled (or actually delivered, depending on the analysis) to each individual i, i=1, …, M. We define the number of responses for individual i as y_i, which is assumed to follow an over-dispersed Poisson model with mean λ. To relate the individual response rates to the treatment arms, we estimate the expected response rate for individual i as, $\lambda_i = y_i/N_i$ and use the following log-linear model: $\log(y_i) = \log(N_i) + \beta_0 + \beta_1 X_i$, where X_i, is the covariate of interest (μEMA assignment in this experiment) for individual i. Through this model, the log-rate of response is modeled linearly. To account for over-dispersion in the data (a violation of an assumption of the Poisson model that restricts the mean response and the variance of the response to be equal), we impose a negative-binomial distribution on the outcomes.

It appears as though subjects in the μEMA group tend to have higher response rates with less variability than those in the EMA group, for all prompts scheduled and delivered, as well as for initial prompt response. The average compliance rate of the μEMA arm was 81.2% compared to an average rate of 64.5% in the EMA arm (Table 3). Through our model fitting procedure, we also determined that demographic variables and self-reported comfort with technology did not significantly impact the individual completion rates. Therefore, we fit the simple Univariate Poisson regression model $\log(y_i) = \log(N_i) + \beta_0 + \beta_1(\mu EMA_i)$, where μEMA_i=1 for those individuals randomized to receive the smartwatch, and 0 otherwise. The resulting coefficient of the μEMA term, β_1, represents the log-relative rate of responding to a scheduled prompt for the μEMA group compared to the EMA group. The final model fit to our study data was $\log(y_i) = \log(N_i) - 0.44 + 0.23*(\mu EMA_i)$. Subjects in the μEMA group were $e^{0.23} = 1.25$ times more likely to respond to a scheduled prompt than subjects in the EMA group (95% CI: 1.10, 1.44). The corresponding p-value for this relative risk is <0.001, suggesting a significant difference in compliance at the α=0.05 level.

Comparing Overall Completion of µEMA and EMA

The same analysis was used to check the effect of µEMA on a participant's likelihood to respond, only considering the prompts that the participant definitely received because the device was on and working properly. This analysis resulted in the following model: $\log(y_i) = \log(N_i) - 0.39+0.30*(\mu EMA_i)$. Subjects in the µEMA group were $e^{0.30}= 1.35$ times more likely to respond to a delivered prompt than subjects in the EMA group (95% CI: 1.20, 1.51). The corresponding p-value for this relative risk is <0.001. Here we see that those individuals assigned to the µEMA group were significantly more likely than the EMA group to respond to prompts that were actually delivered.

Comparing Initial Prompt Response of µEMA and EMA

To investigate the effect of µEMA on a subject's likelihood to respond to the first prompt, we fit the same over dispersed log-linear model, except we replace y_i with z_i, the number of prompts that were answered on the first notification. The µEMA group responded to 88.3% of the first responses, compared to 53.3% in the EMA group. Through the same model discussed above, we found that no demographic information and self-reported comfort with technology improved the model to investigate the effect of μEMA_i Our final model for this analysis was $\log(z_i) = \log(N_i) - 0.63+ 0.50*(\mu EMA_i)$. Subjects in the µEMA group were $e^{0.50}= 1.65$ times more likely to respond to the first delivered prompt than those subjects in the EMA group (95% CI: 1.37, 1.99). Again, we see that the µEMA participants were significantly more likely to respond to the first delivered prompt than the EMA participants.

Comparing Final Reported Burden of µEMA and EMA

Table 4 summarizes the participant self-report data, obtained from the weekly surveys, and Figure 3 shows that self-reported burden increased over the 4 weeks. All participants were asked to estimate the average amount of time that it took them to answer a question. Those in the EMA condition reported taking an average of 4.0 s to answer a single question in the question set, and 22.8 s to answer all 6 questions in the set. Those in the µEMA condition reported taking 9.1 s to answer a watch question. Participants in the EMA condition *actually* took 38.2 s to answer the first question of the question set, and 54.6 s to complete responses for the full question set, measured from the beginning of the prompt sound/vibration. Participants in the µEMA condition took 6.8 s to answer each question from the start of the prompt. It usually takes ≈3 s or more to notice a prompt on the watch. Therefore, we estimate that most µEMA questions actually took ≈3-4 s to answer from noticing the prompt to resuming activity.

Lastly, participants in both conditions were given the opportunity to describe the most negative and positive aspects of the system they used. The negative aspects shared by µEMA participants were as follows. Three reported noticing prompts without questions. Three reported that they occasionally choose an unintended response due to the sensitivity of the watch face. Three reported the repetition of questions was bothersome. Three reported that the vibration pattern was too long and intense. Finally, three reported that the 8 AM start time on weekends was inconvenient. Other negative aspects reported by single µEMA participants included that "the watch vibrated regardless of if it was on mute or not," "the questions were not correlated with feeling," the answer choices were "worded awkwardly,"

the questions were “a distraction while studying,” “there was no option for silence vs. dismissal,” the “app drains phone battery,” “a bulky watch,” and “limited options for answers.” Positive aspects shared by μEMA participants are as follows. Seven reported the ease and quickness of answering questions on the wrist. Four reported the watch being easy to use. Three reported the questions forced introspection about behavior, and five reported that wearing a smartwatch was “cool” or “fun.” Other positive aspects reported by μEMA participants included that the watch was a “good talking point,” it was “good at alerting,” had “simple answer choices,” “it didn’t get in the way,” and that it had an “easy interface.” One appreciated that questions were only prompted for “12 hours a day.”

The two outliers in the μEMA condition were included in the qualitative analysis. While these participants did not consistently comply with persistent use of equipment and survey questions, they responded to the weekly qualitative surveys. Both participants found that the question sets prompted on the watch interrupted their task ‘somewhat,’ and were ‘somewhat’ annoying and distracting. When asked about the negative aspects of the study, one outlier participant replied that the follow up questions were most annoying, Bluetooth “being on” all day was inconvenient, the phone battery drained quickly, and the same questions became “tedious.” The other outlier participant found that the watch was “inconsistent” and made his phone freeze. When asked if they would participate in a similar study in the future, both participants declined. Despite having answered an average of 833 prompts over 4 weeks (and having endured even more prompts), 15 out of the 19 μEMA participants answered “Yes” to the question, “Would you be willing to participate for another four weeks in this study at a future date at your convenience?”

EMA participants expressed reservations as well. Four reported the negative effect of the software on the battery life of their phones, two reported they did not like that the phone cluttered their notification bar with messages that the software was running, and five reported that the audio alert was “annoying.” Other negative aspects reported by EMA participants included “the duration of time windows during work hours,” the “delayed time for some responses to register,” “the repetitiveness of the questions six times a day,” and the limited window of time given to respond.” Positive aspects shared by participants in the EMA group consisted of seven reporting the surveys being easy to answer and eight reporting that the surveys were not too time consuming. Other positive aspects reported by EMA participants included “easy interface,” “the back button was useful,” “being able to keep track of missed and completed [questions] was useful,” “the alert was attention grabbing,” and the “consistency of questions.” 13 out of the 15 EMA participants that completed the study expressed interest in participating for another 4 weeks, if needed.

DISCUSSION

Despite μEMA having ≈8x as many interruptions as standard EMA, participant response rates and feedback suggest μEMA compared favorably to EMA in the pilot, especially when two outliers who are known to have not kept their equipment with them and functioning are removed. Participants using μEMA were significantly more compliant than those in the standard EMA group, and they responded to significantly more received prompts. In fact, the initial prompt response was ≈35 percentage points higher among the μEMA participants

than EMA participants, suggesting there was less hesitation or friction from μEMA participants in responding. μEMA participants also initiated answering question sets more quickly than EMA participants, further implying a lower friction to answering questions. Although a slightly higher percentage of μEMA participants reported that the prompts interrupted them, they also reported the prompts to be less distracting. This trend might suggest that the fast nature of the microinteractions keeps disruption of everyday activity low.

Trends in Compliance Over 4 Weeks

Compliance drop across time in EMA studies is common, even with financial compensation (which we did not use), and most EMA studies in health and psychology run about a week – well under the four-week period of our study. Therefore, it is reasonable to expect compliance drop throughout this study for both the μEMA and EMA arms. To explore this further, we estimated the mean compliance on each day for both μEMA and EMA (Figure 4). EMA compliance drops towards the end of the study, which is consistent with previous studies. However, μEMA compliance is generally maintained at a high rate after a lower start. Participants may have perceived μEMA prompts as less distracting than EMA prompts, leading to more sustainability, even without financial compensation. Since compliance estimation takes all data losses into account, completion rate at a specific time will always be greater than or equal to compliance.

In the case of μEMA, the gap between compliance and completion rates could result from participants' lack of experience with the loaned smartwatch. Although interaction consisted solely of fast, single-tap interaction, maintaining a charged smartwatch requires development of a new charging habit. The Moto 360 battery only lasts a single day with heavy usage [25]. We suspect that some participants took a few days to get into a routine of charging and wearing the device. As a result of devices not being charged, participants would have received fewer prompts than scheduled for μEMA on certain days. In fact, ≈25% of the undelivered prompts in μEMA were due to battery drainage or phone/watch being powered off. Due to limitations of Android eco-system, it is difficult to distinguish other factors resulting in undelivered prompts. Figure 4 highlights how the compliance and completion converge for μEMA, likely as participants master the watch. Yet, high completion rate (≈90%) throughout the study suggests that participants responded to nearly all the prompts they received on their smartwatch.

Trends in Participant Attrition

In a study of compliance, it is important to note the number of people who dropped out of the study and the reasons why. One EMA participant dropped after a week, and that participant's data are included in these results for that week. That participant withdrew because s/he stated the prompts "disturbed me from the things I'm doing and it gets in the way of things I actually have to do." An additional 3 participants dropped within the first 24 h and their data are therefore not included in the results presented in this paper. One was in the μEMA condition and cited battery drain and inexplicable phone shutdowns. The two others were in the EMA condition. One said that "her schedule had changed and she would no longer have the time to participate."

LIMITATIONS AND FUTURE DIRECTIONS

This study has several limitations. The first was the relatively small sample size and distribution of the population. Due to time constraints and equipment limitations, such as requiring Android phones with version 4.3+ compatible the watch, only 38 out of the 88 interested individuals were eligible to participate. Today, smartphones are so customized and personal that it is challenging to loan devices, and we did not want participants to be required to carry an additional 'research phone.' Future studies could improve sample size by developing cross-platform μEMA software. A much larger number of smartwatches are now available, as well, used by many people on a daily basis.

Our participants were not smartwatch users, for two reasons: (1) smartwatches were uncommon at the time of data collection, and (2) we wanted to ensure that the smartwatch prompted as intended, and so we used a single device model (even doing that, obtaining perfectly reliable prompting proved challenging, due to Android Wear bugs). Future work should recruit regular smartwatch users with a variety of models and compare their performance, especially the relationships between compliance and completion rates. Usage patterns on smartwatches will change, and they could impact the viability of μEMA as a method, as they impact the viability of EMA on phones.

Our implementation of EMA used 'persistent' notifications to incentivize compliance, as is common in EMA studies on Android devices; our implementation of μEMA had no equivalent. This would presumably bias better completion toward EMA, not μEMA. Nevertheless, despite having this notification in the EMA condition, compliance and completion for μEMA were significantly higher. Future research should explore the role persistent smartwatch notification could play on impacting compliance.

At the time of this study smartwatches were not common, and so novelty of using the smartwatch may bias the results. However, at one time EMA was used with (novel) PDAs and (novel) phones, and there is little evidence novelty affected behavior at that time. One prior 2-day study did find novelty impacting compliance [50], but it is unknown if those effects would be sustained for four weeks. Future work could estimate this bias using within-subject designs with larger sample sizes (to control for ordering effects, which are likely to be substantial), or alternatively by implementing standard EMA on a smartwatch and including it as an additional arm in this experiment to gauge any possible effects of the smartwatch alone.

Our pilot demonstrates the need for new functionality in future tests. For example, some μEMA participants indicated that they accidently tapped the wrong answer on the watch at times. We intentionally did not ask for confirmation, because that would require a non-microinteraction, double-tap process. However, a last-resort "undo" strategy for this situation could be considered.

Out of 3,068 missed first prompts by μEMA participants, 231 (7.5%) were prompts where an Android Wear bug led to a prompt without an answerable survey, and 150 (4.9%) were dismissed (a small percentage, because it is almost as easy to answer as to dismiss a microinteraction). The remaining 2,687 (87.8%) timed out. Based on testing with our team,

we suspect some prompts were actually not perceived; the tactile vibration pattern can be surprisingly easy to miss when someone is intensely engaged in physical or cognitive activity. This is both a problem and an opportunity for future work to address.

Ultimately, the scope of this pilot study was limited to measuring compliance, completion and perceived burden of the μEMA on participants. Future work should assess the validity of PANAS data collected, or the validity of other types of data collection specifically of interest to researchers (e.g., the eating/smoking system validation example mentioned in the introduction). Of particular interest, given the burden of temporally-intensive μEMA may be tolerable, is how it might be used to develop more sophisticated methods for modeling behavior, where a system incrementally learns and updates models by asking a large number of simple questions, throughout everyday life.

CONCLUSIONS

In this study, μEMA participants showed a significantly higher compliance rate, completion rate, and first prompt response rate, as compared to standard EMA participants, suggesting μEMA may be a viable method for measuring some aspects of behavior in everyday life at high temporal density. μEMA was not perceived as more burdensome than EMA on a similar measurement task. This is a surprising result because participants using μEMA were prompted ≈8 times more often than EMA, often several times an hour. This rate was sustained for 4-weeks, without financial compensation. μEMA may therefore create new opportunities to gather temporally dense data in ubicomp, health, and other domains where understanding the dynamic nature of behavior in natural settings at high temporal density is important.

ACKNOWLEDGEMENTS

This work was funded, in part, by a Google Glass Research Award. The phone-based EMA software system was made possible with funding from the NIH (R21 HL108018-01). The authors thank Dr. Donna Spruitj-Metz for thoughtful discussions on the utility of μEMA in behavioral science, and the anonymous reviewers for helpful feedback.

REFERENCES

1. Amft Oliver, Junker Holger, and Troster Gerhard. 2005 Detection of eating and drinking arm gestures using inertial body-worn sensors. In Proceedings of IEEE Intl Symp on Wearable Comp: 160–163.
2. Ashbrook Dan. L.. 2009 Enabling Mobile Microinteractions. Ph.D Dissertation. College of Computing, Georgia Institute of Technology, Atlanta, GA.
3. Barrett Lisa F. and Barrett Daniel J.. 2001 An introduction to computerized experience sampling in psychology. Soc Sci Comput Rev, 19: 175–185.
4. Baumeister Roy. F.. 2002 Yielding to temptation: Self-control failure, impulsive purchasing, and consumer behavior. Consumer Research 28: 670–676.
5. Collins Francis S. and Varmus Harold. 2015 A new initiative on precision medicine. New Eng. J. Med 372: 793–795. [PubMed: 25635347]
6. Collins Lorraine R., Kashdan Todd B., and Gollnisch Gernot. 2003 The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. Exp and Clin Psychopharmacology 11: 73–8.

7. Courvoisier Delphine S., Eid Michael, and Lischetzke Tanza. 2012 Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. Psychol Assess 24: 713–20. [PubMed: 22250597]
8. Estrin Deborah. 2014 Small data, where n=me. Comm. of the ACM 57: 32–34.
9. Dey Anind K., Wac Katarzyna, Ferreira Denzil, Tassini Kevin, Hong Jin-Hyuk, and Ramos Julian. 2011 Getting closer: An empirical investigation of the proximity of user to their smart phones. In Proceedings of Intl Conf on Ubiquitous Comput. (UbiComp '11), 163–172.
10. Dunton Genevieve F., Dzubur Eldin, Kawabata Keito, Yanez Brenda, Bo Bin, and Intille Stephen. 2014 Development of a smartphone application to measure physical activity using sensor-assisted self-report. Frontiers in Pub Health 2: 88–100. [PubMed: 25101256]
11. Dunton Genevieve F., Kawabata Keito, Intille Stephen, Wolch Jennifer, and Pentz Mary A.. 2002 Assessing the social and physical contexts of children's leisure-time physical activity: An ecological momentary assessment study. Am J. Health Promot 26: 135–42.
12. Ferreira Denzil, Goncalves Jorge, Kostakos Vassilis, Barkhuus Louise, and Dey Anind K.. 2014 Contextual experience sampling of mobile application microusage. In Proceedings of the 16th Intl Conf on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14), 91–100.
13. Fogarty James, Ko Andrew J., Htet Htet Aung Elspeth Golden, Tang Karen P., and Hudson Scott E.. 2005 Examining task engagement in sensor-based statistical models of human interruptibility. In Proceedings of the SIGCHI Conf on Human Factors in Comput Sys (CHI '05), 331–340.
14. Fuller-Tyszkiewicz Matthew, Skouteris Helen, Richardson Ben, Blore Jed, Holmes Millicent, and Mills Jacqueline. 2013 Does the burden of the experience sampling method undermine data quality in state body image research? Body Image 10: 607–13. [PubMed: 23856302]
15. Gillie Tony and Broadbent Donald. 1989 What makes interruptions disruptive A study of length, similarity, and complexity. Psychol Research 50: 243–50.
16. Hawkley Louise C., Burleson Mary H., Berntson Gary G., and Cacioppo John T.. 2003 Loneliness in everyday life: Cardiovascular activity, psychosocial context, and health behaviors. J Pers Soc Psychol 85: 105–120. [PubMed: 12872887]
17. Hekler Eric. B., Klasnja Predrag, Traver Vicente, and Hendriks Monique. 2013 Realizing effective behavioral management of health: The metamorphosis of behavioral science methods. IEEE Pulse 4: 29–34. [PubMed: 24056791]
18. Hektner Joel M., Schmidt Jennifer A., and Csikszentmihalyi Mihaly. 2007 Experience Sampling Method: Measuring the Quality of Everyday Life. SAGE Publications.
19. Hufford Michael R., Shields Alan L., Shiffman Saul, Paty Jean, and Balabanis Mark. 2002 Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. Psychol of Addictive Behav 16: 205–11.
20. IDC Forecasts Worldwide Shipments of Wearables to Surpass 200 Million in 2019. 2015 Retrieved January 10, 2016 from https://www.idc.com/getdoc.jsp?containerId=prUS40846515
21. Intille Stephen S.. 2007 Technological innovations enabling automatic, context-sensitive ecological momentary assessment in The Science of Real-Time Data Capture: Self-Report in Health Research, Stone AA, Shiffman S, Atienza AA, and Nebeling L, Eds: Oxford University Press: 308–337.
22. Kaplan Robert. M. and Stone Arthur A.. 2013 Bringing the laboratory and clinic to the community: Mobile technologies for health promotion and disease prevention. Ann Rev of Psychol 64: 471–98. [PubMed: 22994919]
23. Kapoor Ashish and Horvitz Eric. 2008 Experience sampling for building predictive user models: A comparative study. In Proceedings of the SIGCHI Conference on Human Factors in Comput Sys (CHI '08), 657–666.
24. Kim Jinhyuk, Nakamura Toru, Kikuchi Hiroe, Sasaki Tsukasa, and Yamamoto Yoshiharu. 2013 Co-variation of depressive mood and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. PloS One, 8 9: e74979. [PubMed: 24058642]
25. Lee Nicole. 2015 Moto 360 review: It's the best Android Wear watch, but that isn't saying much. Retrieved March 25, 2016 from https://www.engadget.com/2014/09/12/moto-360-review/
26. Lewis James R.. 1995 IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Intl J. Human-Computer Interaction 7, 1 (January 1995), 57–78.

27. Lund Arnold M.. 2001 Measuring usability with the USE questionnaire. STC Usability SIG Newsletter, 8.
28. Nagel Kristine S., Hudson James M., and Abowd Gregory D.. 2004 Predictors of availability in home life context-mediated communication. In Proceedings of ACM Conf on Computer Supported Cooperative Work (CSCW '04), 497–506.
29. Nilsen Wendy J. and Pavel Misha. 2013 Moving behavioral theories into the 21st Century: Technological advancements for improving quality of life. IEEE Pulse 4: 25–28. [PubMed: 24056790]
30. Marcello Pagano and Gauvreau Kimberlee. 2000 Principles of Biostatistics. Vol. 2 Pacific Grove, CA.
31. Riley William T., Rivera Daniel E., Atienza Audie A., Nilsen Wendy, Allison Susannah M., and Mermelstein Robin. 2011 Health behavior models in the age of mobile interventions: Are our theories up to the task? Transl Behav Med 1: 53–71. [PubMed: 21796270]
32. Rivera Daniel E. and Jimison Holly B.. 2013 Systems modeling of behavior change: Two illustrations from optimized interventions for improved health outcomes. IEEE Pulse 4: 41–47. [PubMed: 24233191]
33. Saranummi Nillo, Spruijt-Metz Donna, Intille Stephen S., Korhonen IIkka, Nilsen Wendy J., and Pavel Misha. 2013 Moving the science of behavioral change into the 21st Century: Novel solutions to prevent disease and promote health. IEEE Pulse 4: 22–24.
34. Serre Fuschia, Fatseas Melina, Debrabant Romain, Alexandre Jean-Marc, Auriacombe Marc, and Swendsen Joel. 2012 Ecological momentary assessment in alcohol, tobacco, cannabis and opiate dependence: A comparison of feasibility and validity. Drug Alcohol Dependence 126: 118–23. [PubMed: 22647899]
35. Shiffman Saul and Stone Arthur A.. 1998 Ecological momentary assessment in health psychology. Health Psychol 17: 3–5.
36. Shiffman Saul, Stone Arthur A., and Hufford Michael R.. 2002 Ecological momentary assessment. Ann Rev Clinical Psychol, 4 1–32.
37. Smyth Joshua M. and Stone Arthur A.. 2003 Ecological momentary assessment research in behavioral medicine. Happiness Studies 4: 35–52.
38. Speier Cheri, Vessey Iris, and Valacich Joseph S.. 2003 The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. Decision Sciences 34: 771–797.
39. Speier Cheri, Valacich Joseph S., and Vessey Iris. 1999 The influence of task interruption on individual decision making: An information overload perspective. Decision Sciences 30: 337–360.
40. Spring Bonnie, Gotsis Marientina, Paiva Ana, and Spruijt-Metz Donna. 2013 Healthy apps: Mobile devices for continuous monitoring and intervention. IEEE Pulse 4: 34–40. [PubMed: 24233190]
41. Spruijt-Metz Donna and Nilsen Wendy. 2014 Dynamic models of behavior for just-in-time adaptive interventions. Pervasive Comp, IEEE 13: 13–17.
42. Donna Spruijt-Metz Eric Hekler, Saranummi Niilo, Intille Stephen, Korhonen IIkka, Nilsen Wendy, Rivera Daniel, Spring Bonnie, Michie S, Asch D, Sanna A, Salcedo V, Kukakfa R, and Pavel Misha. 2015 Building new computational models to support health behavior change and maintenance: New opportunities in behavioral research. Transl Beh Med: 1–12.
43. Donna Spruijt-Metz Wendy Nilsen, and Pavel Misha. 2015 mHealth for behavior change and monitoring In mHealth Multidisciplinary Verticals, Adibi S, Ed., Boca Raton, FL: CRC Press, 120–32.
44. Stone Arthur A. and Shiffman Saul. 1994 Ecological momentary assessment (EMA) in behavioral medicine. Annals of Beh Med 16: 199–202.
45. Stone Arthur A., Shiffman Saul, Atienza Audie A., and Nebeling Linda, Eds. 2007 The Science of Real-Time Data Capture: Self-Report in Health Research. New York, NY: Oxford University Press.
46. Stone Arthur A., Shiffman Saul, Atienza Audie A., and Nebeling Linda. 2007 Historical roots and rationale of ecological momentary assessment (EMA) The Science of Real-Time Data Capture: Self-Reports in Health Research. Stone AA, Shiffman S, Atienza AA, and Nebeling L, Eds., New York, NY: Oxford University Press: 3–10.

47. Tang Qu, Vidrine Damon J., Crowder Eric, and Intille Stephen S.. 2014 Automated detection of puffing and smoking with wrist accelerometers. In Proceedings of the 8th Intl Conf on Pervasive Comput Tech for Healthcare (PervasiveHealth '14), 80–87.

48. Timmermann Janko, Heuten Wilko, Boll Susanne. 2015 Input methods for Brog-RPE-scale on smartwatches. In Proceedings of Intl Conf on Pervasive Comp Tech for Healthcare (PervasiveHealth'15), 80–83.

49. Truong Khai N., Shihipar Thariq, and Wigdor Daniel J.. 2014 Slide to X: unlocking the potential of smartphone unlocking. In Proceedings of the SIGCHI Conf on Human Factors in Comput Sys (CHI '14). ACM, New York, NY, USA, 3635–3644.

50. Walsh Erin I. and Brinker Jay K.. 2016 Should participants be given a mobile phone, or use their own? Effects of novelty vs utility. Telematics and Informatics 33 1, 25–33.

51. Watson David, Clark Lee A., and Tellegen Auke. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers Soc Psychol 54: 1063–70.

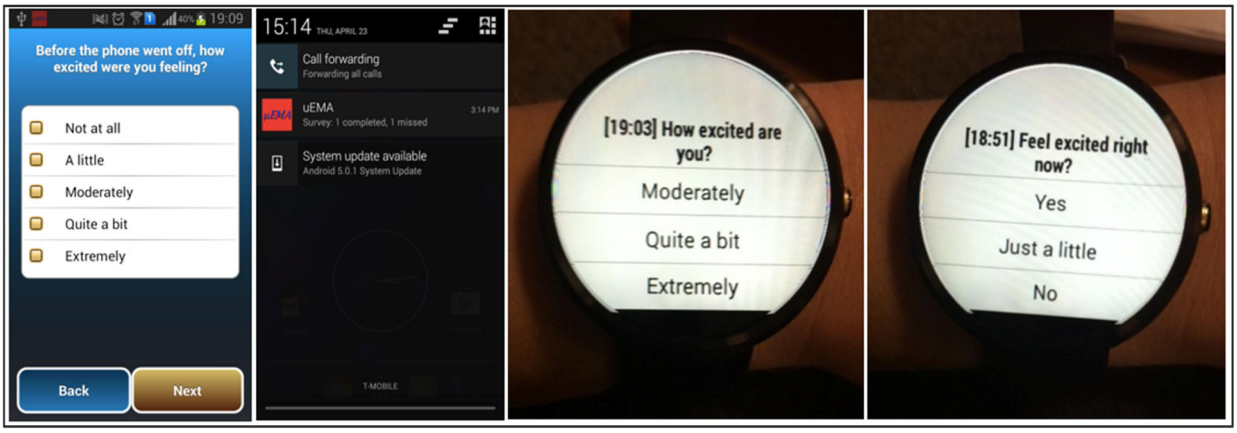

Figure 1:
(left) EMA on the phone – one of six questions in a question set, (middle left) EMA persistent notification screen, (middle right) µEMA on the watch – single question, (right) follow-up question.

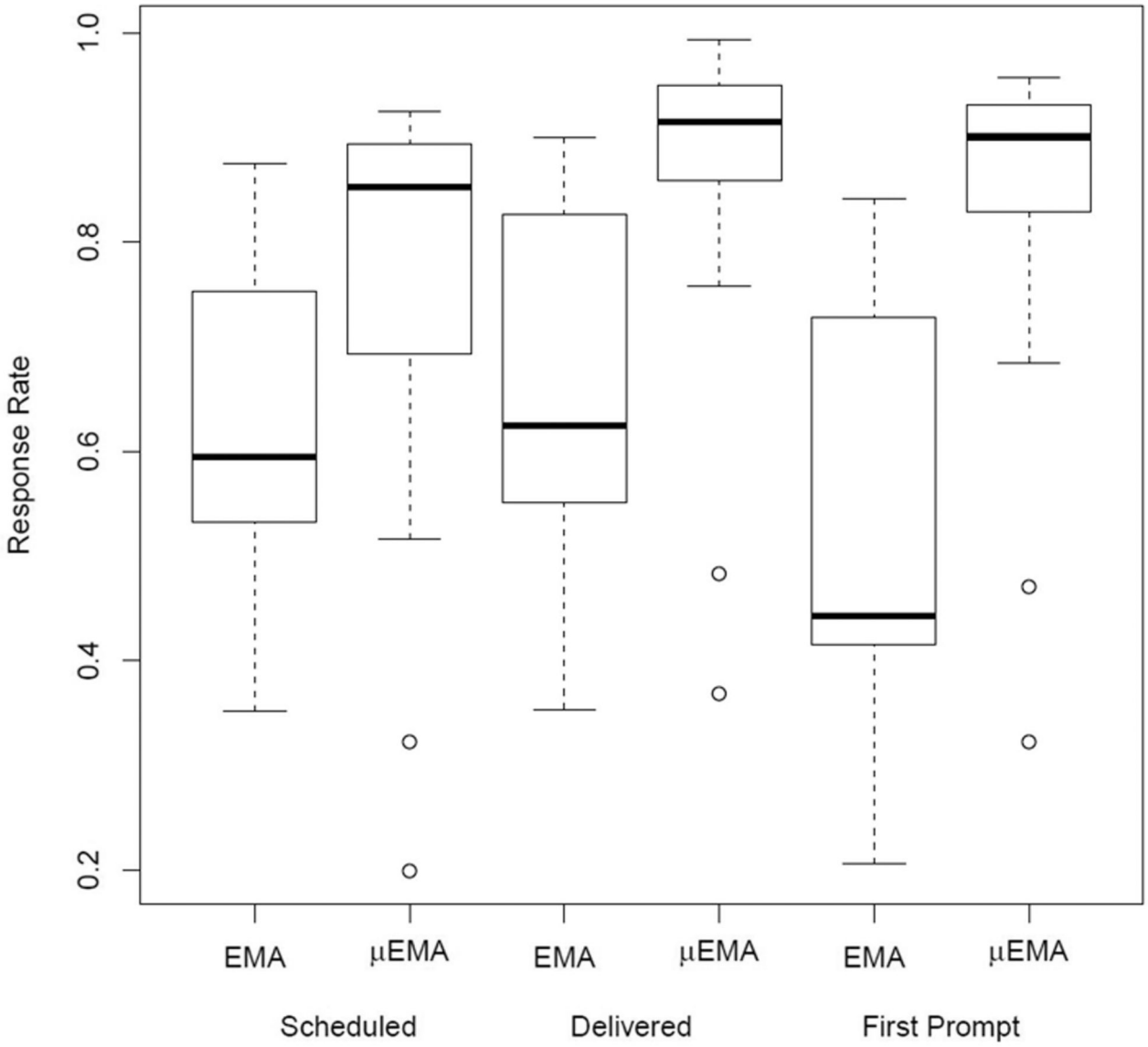

Figure 2.
Box plot showing results from compliance (scheduled), completion (delivered), and first prompt analysis

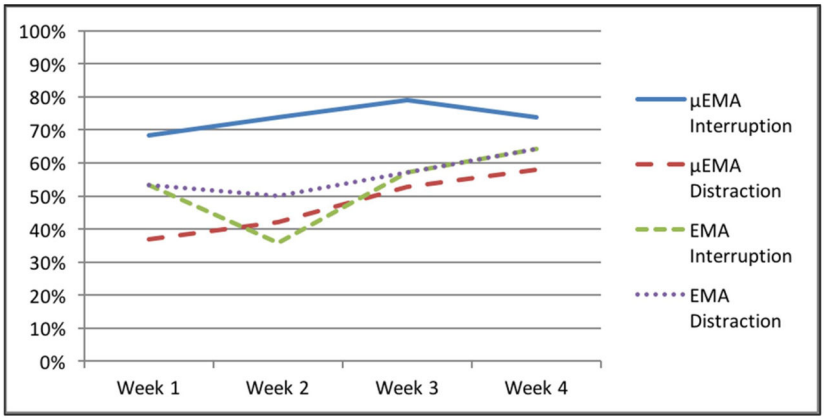

Figure 3.
Perceived burden responses by week

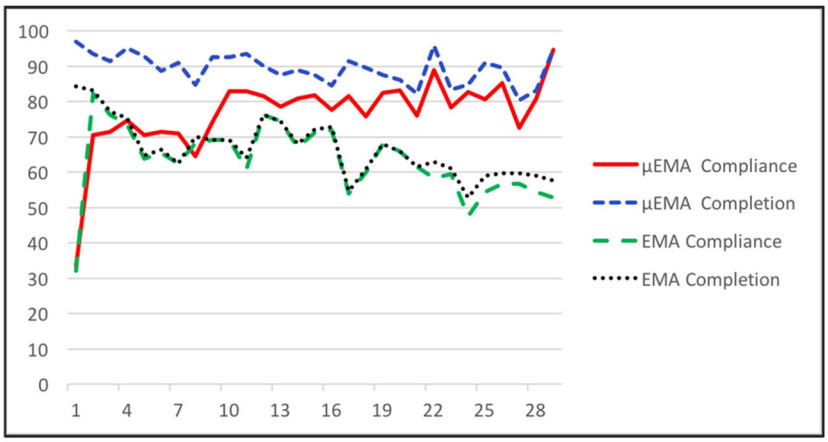

Figure 4.
µEMA and EMA compliance & completion (%) change by days in study

Table 1.

Demographic characteristics of sample that completed the study (N=33)

	Total	**EMA**	**μEMA**
All	33 (100%)	14 (42%)	19 (58%)
Male	15 (45%)	6 (43%)	9 (47%)
Female	21 (55%)	8 (57%)	10 (53%)
Age(min,max)	26.7 (18,55)	29.4 (18,55)	24.6 (18, 49)
Education Level			
HS Diploma	2 (6%)	1 (7%)	1 (5%)
Bachelors	19 (58%)	7 (50%)	12 (63%)
Masters	12 (36%)	6 (43%)	6 (32%)
Self-Reported Comfort Level Using Technology			
Strongly	20 (61 %)	7 (50%)	13 (68.4%)
Mostly	11 (33%)	6 (42.9%)	5 (26.3%)
Somewhat	2 (6%)	1 (7.1%)	1 (5.3%)
Neutral/Not	0	0	0

Table 2.

µEMA response rate with and without outliers. Asterisk (*) indicates significant ($p<0.05$) difference with EMA

	With Outliers	Without Outliers
Compliance %	75.64 %	81.21 % *
Completion %	87.81% *	91.81 % *
First Prompt %	84.95 % *	88.33 % *

Table 3.

Overall response behavior

	EMA	μEMA
Mean Question Set Compliance	64.54%	81.21%
Mean Question Set Completion	67.36%	91.81%
Question Sets Answered	1546	15278
Questions Answered	9270	15278
Mean Question Sets Completed After first prompt	53.28%	88.33%

Table 4.

Summary of weekly survey responses

Report that:	**EMA**	**μEMA**
Questions interrupt	51.1%	63.2%
Frequency of prompts is distracting	49.2%	34.21%