



HHS Public Access

Author manuscript

Proc ACM Interact Mob Wearable Ubiquitous Technol. Author manuscript; available in PMC 2018 September 07.

Published in final edited form as:

Proc ACM Interact Mob Wearable Ubiquitous Technol. 2017 September ; 1(3): . doi:10.1145/3130957.

Microinteraction Ecological Momentary Assessment Response Rates: Effect of Microinteractions or the Smartwatch?

ADITYA PONNADA,

Northeastern University, Boston MA

CAITLIN HAYNES,

Northeastern University, Boston MA

DHARAM MANIAR,

Northeastern University, Boston MA

JUSTIN MANJOURIDES, and

Northeastern University, Boston MA

STEPHEN INTILLE

Northeastern University, Boston MA

Abstract

Mobile-based ecological-momentary-assessment (EMA) is an in-situ measurement methodology where an electronic device prompts a person to answer questions of research interest. EMA has a key limitation: interruption burden. Microinteraction-EMA(μ EMA) may reduce burden without sacrificing high temporal density of measurement. In μ EMA, all EMA prompts can be answered with 'at a glance' microinteractions. In a prior 4-week pilot study comparing standard EMA delivered on a phone (phone-EMA) vs. μ EMA delivered on a smartwatch (watch- μ EMA), watch- μ EMA demonstrated higher response rates and lower perceived burden than phone-EMA, even when the watch- μ EMA interruption rate was 8 times more than phone-EMA. A new 4-week dataset was gathered on smartwatch-based EMA (i.e., watch-EMA with 6 back-to-back, multiple-choice questions on a watch) to compare whether the high response rates of watch- μ EMA previously observed were a result of using microinteractions, or due to the novelty and accessibility of the smartwatch. No statistically significant differences in compliance, completion, and first-prompt response rates were observed between phone-EMA and watch-EMA. However, watch- μ EMA response rates were significantly higher than watch-EMA. This pilot suggests that (1) the high compliance and low burden previously observed in watch- μ EMA is likely due to the microinteraction question technique, not simply the use of the watch versus the phone, and that (2) compliance with traditional EMA (with long surveys) may not improve simply by moving survey delivery from the phone to a smartwatch.

Keywords

Experience Sampling; Ecological Momentary Assessment; Smartwatch; Compliance; Empirical Studies; Microinteractions; Wearable computing

1 INTRODUCTION

Ecological momentary assessment (EMA) [40], also known as experience sampling [10], is an *in-situ* data collection method used to measure behavior in health, ubiquitous computing, and other domains where intensive longitudinal self-reported assessment of behavior in the real-world is important. Although, passive, continuous sensing devices can be used to measure some behaviors, such as physical activity from accelerometers, measurement of other behaviors or states or contexts will *require* self-report. Chronic pain, feelings of productivity, self-esteem, and dietary intake are four such examples where the behaviors/states/contexts change quickly, and passive, continuous sensing may never be able to measure them accurately. Some self-report may be required, and the more temporally dense that self-report measurement is, the more useful it may be for scientific research and for creating novel ubiquitous computing systems. In a typical EMA study, participants are prompted by a mobile device several times a day (often 6-10 times or more) with a set of multiple-choice questions relevant to a research construct of interest. EMA reduces recall biases resulting from retrospective recall, because these questions are asked amid everyday activities [37, 38], EMA can provide a more temporally-dense profile of the participant than retrospective recall, because the questions are asked repeatedly [40]. Finally, unlike diary studies, EMA can prevent back- or forward-dated responses by timestamping data entry [39].

One of the most prevalent challenges of EMA is to reduce interruption burden without sacrificing the high temporal density of self-report. Microinteraction-EMA, or micro-EMA (μ EMA), was introduced in a recent pilot study to address this challenge [21]. μ EMA is an *in-situ* data collection methodology that reduces the burden of each interruption so substantially that many more interruptions can take place – more than 30 a day – without accumulating burden, even in a multi-week study. Rather than multi-question, complex-answer, multiple-choice question sets, μ EMA only prompts single questions at a time – but at a high rate – and where the questions are of a yes/maybe/no style. The entire prompt-view-answer sequence should be accomplishable in a microinteraction [3], or ‘at a glance.’ The always-available nature of a smartwatch allows for instant accessibility and is therefore well-suited for micro-EMA, although μ EMA could also be used on other platforms, such as head-mounted computers. In the original pilot, μ EMA was implemented on a smartwatch (referred to as watch- μ EMA in this paper). Despite an interruption rate of ~6 times per hour – 8 times higher than the rate of interruption from standard EMA delivered on a phone (referred to as phone-EMA in this paper) – watch- μ EMA had higher response rates than traditional phone-EMA [21]. This original pilot study left unanswered, however, whether the high compliance rate was due to micro-interactions, *or* just due to the novelty and the improved accessibility of the smartwatch. This is an important distinction. If the improvement is simply due to the accessibility or novelty or tactile stimulus of the watch, then researchers using EMA on phones could simply re-implement the same relatively long EMA surveys on smartwatches and expect to improve compliance and lower participant burden. However, if the improvement is due to the use of the microinteraction questionstyle constraint, then new research is warranted to develop and validate μ EMA versions of existing EMA surveys in domains of interest. This paper describes a new 4-week pilot study.

A phone-EMA protocol was ported to the smartwatch (watch-EMA). Data were gathered that, combined with data from the prior study, allow comparison of response rates of watch-EMA, phone-EMA, and watch- μ EMA.

2 BACKGROUND: EMA AND COMPLIANCE

The technology used for EMA has evolved from personal digital assistants (PDAs) [4] to smartphones (e.g. [19]). So far, each new consumer device has made it more feasible and affordable to deploy EMA at large scale (e.g., [34]). The weakness of EMA as a methodology, however, is its interruption burden. Audio or tactile prompts interrupt everyday activity, distract people from what they are doing, and require time to respond to. Response compliance is likely to fall as perceived burden increases. The difficulty in accessing the smartphone contributes to the burden of answering questions. In fact, even though smartphones are pervasive, they are often not within a hand's reach [12]. The barrier to obtaining/reaching out to the phone and unlocking it to read the question(s) may make participants more hesitant to initiate responses to EMA prompts, and the further the phone is from the body, the less likely a person may notice the prompts. This could result in lower study compliance. One way to address such a burden could be to use a device that is easier to access, and harder to ignore, such as a smartwatch. Smartwatches are worn on wrist, permitting reliable delivery of tactile prompts, and thus, are substantially faster to access than a phone in a pocket or the bottom of a cluttered bag [3]. Smartwatches also contain advanced sensors that permit real-time monitoring of wrist motion and heart rate, which could be used for context-sensitive EMA [22].

Several studies have already used watch-type computers for EMA. For instance, Hawkey et al. used a programmable watch as a prompting device to study cardiovascular activity and psychosocial context (no compliance reported) [17]. Kikuchi et al. used a watch-type computer with a visual analog input to record momentary headache intensity (reported 97.5% compliance for a 7-day study) [24]. Timmerman et al. explored usable smartwatch input methods to record perceived exertion during physical activities (no compliance reported) [23]. Kim et al. captured depressive mood and locomotor dynamics using a watch-computer with physical joystick input (reported 94% response rate for 7 days) [26] and compared responses from EMA and the Day Reconstruction Method for fatigue assessment (reported 90.97% response rate for 3 days) [25]. Hernandez et al. recently compared EMA responses between a smartphone, smartwatch, and Google Glass, where participants wore all the devices simultaneously in a 5-day study with compliance-contingent compensation (reported 82.3% response rate for 5 days) [18]. In this prior work, it is unclear if the authors accounted for data losses (e.g., due to power drainage) when reporting compliance. Further, most of these studies used compliance-contingent compensation, which will influence study of compliance, and that may be impractical for intensive longitudinal (e.g., multi-week) studies. In all the watch-based EMA studies published to date, the watch-computers or smartwatches were loaned to the participants, and it is unclear if novelty impacted compliance and burden. Moreover, in these studies, the questions consisted of multi-question EMA surveys with more than one back to back question – not resulting in a microinteraction. The one exception was the Timmerman et al. study where watch-specific

interactions (e.g., swipe gestures and voice input) were used to measure perceived exertion [23].

Smartwatches have shorter access times than phones [3], and tactile prompts could be more reliable indicators than phone auditory prompts. In fact, recent empirical studies suggest that smartwatches are used ~50% of the time for glanceable interactions such as checking time and notifications [33]. Moreover, an in vivo study suggests that smartwatches are used as often as 6 times in an hour with each interaction lasting only 6-7s compared to smartphones (~38s) [33]. Thus, it is possible that users are more comfortable in using smartwatches for quicker access and short interactions as compared to smartphones. Moreover, today smartwatches are less common than smartphones [43], and the novelty of a smartphone-based EMA may impact compliance. Walsh and Brinker, in a 2-day EMA study, observed positive novelty effects of a loaned smartphone on participant response behavior [41]. The study's short duration of 2 days, however, prevents generalization to longer, more typical, EMA studies of a week or more. We are not aware of any other studies addressing the impact of device novelty, either phone or smartwatch, on EMA. While it is possible that smartwatches are considered quite novel when they are passed out to participants in a research study using EMA, and such novelty may boost compliance, smartwatches also have their own set of challenges that may reduce compliance. Unlike traditional watches, smartwatches need frequent charging [32]. Participants who are not used to smartwatches therefore need to learn how to use them and maintain them (charging them regularly, often daily or more); they many not yet have developed habits for maintenance and use as they have already done for their personal smartphones [29, 32]. Finally, interacting with a smartwatch's small screen interface could make answering multiple-choice EMA questions more challenging, and therefore more burdensome, than on the larger phone screen. Overall, although a smartwatch's ease-of-access, prompt salience, and novelty could improve compliance, factors such as the device's learning curve and small screen could offset those gains, resulting in no significant effect on study compliance.

3 METHODOLOGY: COMPARING WATCH-EMA, WATCH- μ EMA AND PHONE-EMA

In prior work, watch- μ EMA was compared with phone-EMA, and watch- μ EMA had significantly higher response rates, despite an ~8 times increase in interruption burden. Not tested was porting the phone EMA survey (6 questions prompted back-to-back) to a smartwatch (i.e., watch-EMA). Without this additional study condition, it is not clear if the effect observed in prior work is due to the microinteraction question style, or simply due to porting a phone-based questionnaire to a smartwatch. Therefore, in this study, we implemented watch-EMA, using the same question set from phone-EMA in the prior study and compared the phone-EMA (previous study), watch- μ EMA (previous study), and watch-EMA (present study) conditions. Like the phone-EMA condition in the previous study, participants in the new watch-EMA condition were prompted six times a day with a set of six back-to-back multiple choice questions for a period of four weeks; the prompting and reprompting frequency was also the same as phone-EMA. To remain consistent with the prior study, watch-EMA did not allow notifications from the other applications installed on

the phone. This was practical because participants were not smartwatch users with expectations about smartwatch notifications. Phone-EMA did permit additional notifications, because it was not practical to ask participants to change how they used their personal phones. Other than this, the differences between phone-EMA and watch-EMA were the device and the GUI changes necessitated by the device. The difference between watch-EMA and watch- μ EMA in the prior study was the use of standard, less-frequent question sets versus the use of microinteraction questions (more frequently prompted, single-question interactions with yes/maybe/no answers that can be answered with a glance/tap). In brief, watch-EMA has the same interruption burden as phone-EMA and has the same access time and prompt feedback as watch- μ EMA.

As in phone-EMA [21], the first five questions in watch-EMA were derived from the Positive and Negative Affect Schedule (PANAS) [42]. PANAS questions were of the form, “How {excited/nervous/stressed/upset/alert} are you?” with five choices: “Not at all,” “A little,” “Moderately,” “Quite a bit,” and “Extremely.” The sixth question was on physical activity and of the form, “Are you {sitting/walking/lying down}?” with three choices: “Yes,” “Sort of,” and “No.” Watch-EMA participants received six survey prompts randomly scheduled in two-hour slots between 8AM and 8PM (8-10AM, 10AM-Noon, Noon-2PM, 2-4PM, 4-6PM, 6-8PM). When the device prompted, questions appeared one at a time in a 6-question question set (QS). Prompts on the watch, which did not have audio output and therefore could not beep, used a vibration pattern lasting 11.5 s and gradually increasing in intensity. The watch-EMA condition used identical prompting as watch- μ EMA. If a QS was not fully answered within 5 min, another prompt was delivered. If that reprompt was not answered within the next 5 min, the QS disappeared from the screen.

For completeness in the current study, we also implemented the phone- μ EMA case. However, we anticipated that while μ EMA strategy may be viable on the watch, it was unlikely to be well-received on the phone, due to the difficulty of accessing the device. Common sense suggested hearing beeping 30 times a day, finding and pulling out and unlocking the phone, and then answering a single question each time, would become burdensome quickly. When we tested the phone- μ EMA condition with our first 5 participants, 4 dropped out of the study in less than 3 days, citing excessive burden of survey prompts. Deeming this burden unacceptable for continuing recruitment, we dropped the phone- μ EMA condition from further comparisons.

This work explores response behavior, not the validity of the responses, and so a standardized survey used in a prior large-scale EMA study [13] was replicated, as in the previous study. We have not examined the validity of responses for two reasons. First, establishing validity of a measurement instrument targeting a construct requires either ground truth data (for criterion validity) or measures of correlates (for convergent validity). Since this study was intended to examine compliance and burden for generalized EMA tasks, we did not include additional surveys to gather information from the participants that might provide ground truth or correlating data for the PANAS. Additional self-report surveys could induce unexplained burden during the study, confounding assessment of our EMA protocol. Second, establishing validity also depends on the domain of research interest as well as the purpose of the measurement. For instance, if a measure is intended to

differentiate between two groups of participants, its validity is established only when significant differences in its composite scores are observed across the groups. However, this study was designed primarily to evaluate the response rates and study burden. Therefore, establishing validity of PANAS using watch- μ EMA, watch-EMA or phone-EMA is beyond the scope of the present study. In fact, strongly validating a new EMA (or μ EMA) measure would first require sufficient data collection at a high response rate, the topic addressed by this and the prior study [21].

3.1 Hypotheses

Six hypotheses are tested in this study: *Hypothesis 1: Watch-EMA vs. phone-EMA compliance rate:* There is no true difference in compliance between watch-EMA and phone-EMA. *Compliance* is defined as the percentage of QS answered by the participant considering all the QS scheduled to be delivered (including questions not delivered due to the device being off for any reason). In other words, considering all sources of data loss and questions that are not responded to, moving the EMA survey to the smartwatch will not improve compliance over phone-EMA.

Hypothesis 2: Watch-EMA vs. phone-EMA completion rate: There is no true difference in completion rates between watch-EMA and phone-EMA. *Completion* is defined as the percentage of QS answered by the participant considering only the delivered QS prompts. Considering only those prompts that were delivered (i.e., the watch or phone was functioning and not turned off), moving the EMA survey to the smartwatch will still not result in better question completion.

Compliance and completion rates differ based on the number of prompts received and scheduled for the participants. Received and scheduled prompts can differ based on device availability, which can change due to data loss from battery drainage or phone-watch disconnection. For instance, suppose a participant is scheduled to receive six prompts in a day, but the participant's device is switched off for six hours. As a result, the participant could receive only three prompts. If the participant answers all three prompts, the compliance rate will be 50% (3/6) and the completion rate will be 100% (3/3). It is not uncommon for EMA-based studies to report the latter statistic but to present it as overall compliance, which can therefore be high even if few questions are answered. Therefore, to avoid any misinterpretation of the quality of EMA response rates, we compare both compliance and completion rates in our studies. The completion rate is always higher than or equal to the compliance rate, with differences being caused by data loss due to devices not being functional or properly maintained/charged.

Hypothesis 3: Watch-EMA vs. phone-EMA first-prompt response rate: There is no true difference in the first prompt response rate between watch-EMA and phone-EMA. *First prompt response rate* is defined as the percentage of QS completed after the first prompt (without a reprompt) considering all delivered QS. Increased hesitation to initiate a response in watch-EMA could directly affect response rates to the first delivered prompts. If the primary burden from a QS is due to the time required to respond, any hesitancy to respond in phone-EMA would also be felt in watch-EMA.

If watch- μ EMA reduces a participant's hesitancy to engage in the short glanceable interactions at a high frequency, then watch- μ EMA should have a higher compliance, completion, and first-prompt response rate than watch-EMA, despite ~ 8 times more interruption, mirroring the watch- μ EMA to phone-EMA comparison in the prior work. This leads to three more hypotheses:

Hypothesis 4: Watch-EMA vs. watch- μ EMA compliance rate: Participants in the watch- μ EMA condition will have a higher compliance than participants answering the same questions using watch-EMA.

Hypothesis 5: Watch-EMA vs. watch- μ EMA completion rate: Participants in the watch- μ EMA condition will respond with a higher completion rate than participants answering the same questions using watch-EMA.

Hypothesis 6: Watch-EMA vs. watch- μ EMA first-prompt response rate: Participants in the watch- μ EMA condition will respond to the first of the prompted questions more often after being prompted than participants answering the same questions using watch-EMA.

3.2 New Watch-EMA Interface

In the prior study [21], watch- μ EMA was implemented on Moto 360 smartwatches that are compatible only with Android version 4.3+. The watch was paired with a smartphone, after which the μ EMA app was installed. All study participants carried their smartphones, because the Moto 360 must be paired with a phone to operate properly. In the watch- μ EMA condition, the participants did not need to do anything on their phones (our staff setup the pairing at the enrollment). An example of one question from the single-question microinteraction is shown in Fig. 1a. Watch- μ EMA questions require a single tap, with no scrolling, to answer. Fig. 1b shows one screen from the 6-question phone-EMA survey. When implementing that survey on the Moto 360's small screen (46 mm diameter) for watch-EMA, question answer options do not comfortably fit on the screen and require scrolling to view and select answer options (see Fig. 1c).

3.3 Study Participants

Following the same protocol and timeline used for watch- μ EMA and phone-EMA, the additional participants in this study were recruited from Northeastern University (students and staff members, ages 18 to 55 years old) using social media posts and campus flyers. Out of 28 respondents, 13 met the Android OS restrictions and were recruited. Three participants dropped out within a week due to frequent loss of connection of the watch with their phones (a previously unknown incompatibility of Samsung S7 devices with the Moto 360 smartwatches), leaving 10 watch-EMA participants for the 4-week pilot study (Table 1). In the prior study, 19 participants, also recruited from Northeastern University, were assigned to the watch- μ EMA condition, and 14 were assigned to the phone-EMA condition. For comparison, we have included demographic information from the phone-EMA and watch- μ EMA conditions from the previous study in Table 1. As compensation in all conditions, participants had the opportunity to try the smartwatch, as they saw fit, for four days after the study. No response-contingent financial incentives were provided.

3.4 Study Procedures

This study was approved by the Northeastern University IRB. **Day 1:** Researchers installed the watch-EMA app on each participant's personal smartphone. Participants were loaned a Moto 360 smartwatch and shown how to wear and charge it. **Day 7, 14, 21:** Participants were emailed a perceived burden questionnaire. **Day 28:** Participants were emailed a short survey with two open-ended questions about the study. **Day 32:** End of data collection. All participants could try the watch, using it as they saw fit, for up to 4 days as compensation.

4 RESULTS

Compliance was assessed quantitatively and perceived burden was measured via self-report. Table 2 & Fig. 2 (left) summarize response rates (in %) for the watch-EMA, phone-EMA, and watch- μ EMA conditions. We perform comparisons in two pairs – watch-EMA vs. phone-EMA, and watch- μ EMA vs. watch-EMA. Watch- μ EMA and phone-EMA have been compared in the previous study and are not reported here. In the watch-EMA condition, one outlier has been identified; that person reported keeping the watch charged for only six out of 28 days of the study resulting in very poor compliance and completion rates. A comparison of response rates with and without outliers, as well as the total scheduled, delivered, and completed prompts are presented in Table 2. Watch- μ EMA had higher compliance, completion, and first prompt response rates than phone- and watch-EMA. Previously, there were two outliers in watch- μ EMA. Figure 2 compares the response rates across all the three conditions including the outliers. Since the outliers in Figure 2 (left) are greater than 1.5 times the interquartilerange of the median, they have been removed from further analyses [30].

Participants were prompted multiple times and received a different number of prompts, due to differences in how well they managed to keep the devices charged. Even in the phone-EMA and watch-EMA conditions, the total number of scheduled prompts for participants could differ depending on when they started their first day of the study. For instance, a participant enrolled in the late evening would have fewer scheduled prompts than those enrolled in the morning. Even though this difference is small, we must adapt a generalized model that accounts for all the differences in prompting rates [7]. When comparing the response rates, let N_i be the number of prompts scheduled or delivered (depending on the comparison) for participant i (where $i = 1, 2, \dots, M$) and M is the number of participants. Then Y_i is the number of answered prompts by participant i , assumed to follow an over-dispersed Poisson distribution with mean λ_i , estimated for each subject as Y_i/N_i . We model the outcome using a log-linear model: $\log(Y_i) = \log(N_i) + \beta_0 + \beta_1^*(\text{watch-EMA})_i + \beta_2^*(\text{Age})_i + \beta_3^*(\text{Comfort with technology})_i + \beta_4^*(\text{Gender})_i$, where $(\text{watch-EMA})_i$, which is the covariate of interest, is equal to 1 if the individual is in the watch-EMA group and 0 for those in the phone-EMA group. β_0 is intercept of the model, and β_1 is the coefficient representing the log-relative rate of responding to a scheduled/delivered prompt in watch-EMA compared to phone-EMA. Likewise, the same log-linear model is used when comparing watch- μ EMA and watch-EMA, where covariate of interest $(\text{watch-}\mu\text{EMA})_i$ is 1 for watch- μ EMA and 0 for watch-EMA. R^2 for each model was computed using McFadden's pseudo R^2 formula [31]. This log-linear model allows us to accommodate

differential numbers of prompts received by the participants, by modeling the response rates. Further, each participant's response rates being independent of each other satisfies the independence assumption of this log-linear regression model. All model fits are assessed using the Pearson Goodness-of-Fit test [27] and by comparing deviances of the fitted full models to null models only including an intercept.

4.1 Response Rates: Watch-EMA vs Phone-EMA

When comparing watch-EMA and phone-EMA, no significant effect of watch-EMA on compliance (Hypothesis 1) was found after fitting a univariate model at $\alpha=0.05$ level. For compliance rate, this model fitting procedure did not yield any significant impact of age, gender, and technology comfort. The resulting model: $\log(Y_i) = \log(N_i) - 0.42 - 0.105*(\text{watch-EMA})_i$ suggests that watch-EMA participants are only 0.9 ($e^{-0.105}$, 95% C.I.: 0.70, 1.16) times more likely to respond to a scheduled prompt compared to phone-EMA. Therefore, the smartwatch alone did not significantly affect participant response rates for scheduled prompts.

Using the same analysis for comparing watch-EMA and phone-EMA completion rate (Hypothesis 2), β_1 represents log-relative rate of responding to a delivered prompt in watch-EMA as compared to phone-EMA. No significant effect of watch-EMA on completion rate was found. For completion rate, this model fitting procedure yielded no significant effect of gender, age and self-reported comfort with technology. The resulting model: $\log(Y_i) = \log(N_i) - 0.37 + 0.108*(\text{watch-EMA})_i$ suggests that watch-EMA participants are only 1.11 ($e^{0.108}$, 95% C.I.: 0.92, 1.34) times more likely to respond to a delivered prompt. Thus, the smartwatch did not significantly change response rates for delivered prompts (i.e., completion rate).

Finally, using the same analysis for comparing watch-EMA and phone-EMA first-prompt completion rate, β_1 represents log-relative rate of responding to a first prompt delivered in watch-EMA as compared to phone-EMA. No significant effect (only marginally significant) of smartwatch was found on first-prompt response rates. For first prompt response rates, this model fitting procedure revealed no significant effect of gender, age and self-reported comfort with technology. The resulting model: $\log(Y_i) = \log(N_i) - 0.60 + 0.280*(\text{watch-EMA})_i$ suggests that watch-EMA participants are 1.32 ($e^{0.280}$, 95% C.I.: 1.00, 1.73) times more likely to respond to the first delivered prompts. Hence, the smartwatch alone did not significantly impact the first prompt response rate. None of the Watch-EMA vs Phone-EMA models showed any evidence of a lack-of-fit, at the $\alpha=0.05$ level. Table 3 summarizes these results.

4.2 Response Rates: Watch-EMA vs Watch- μ EMA

On comparing watch-EMA and watch- μ EMA, watch- μ EMA had significantly higher ($p < 0.05$) compliance than watch-EMA. Using the same model fitting procedure, no significant impact of gender, age, and technology comfort were found on compliance. The resulting model: $\log(Y_i) = \log(N_i) - 0.52 + 0.32*(\text{watch-}\mu\text{EMA})_i$ suggests that watch- μ EMA participants are 1.38 ($e^{0.32}$, 95% C.I.: 1.04, 1.84) times more likely to respond to a scheduled

prompt. Thus, watch- μ EMA resulted in 38% higher compliance rates than watch-EMA, which is more than 1 standard deviation (SD) of watch-EMA compliance rate.

Similarly, watch- μ EMA had significantly higher ($p < 0.05$) completion rate than watch-EMA. For completion rate, this model fitting procedure revealed no significant impact of age, gender, and technology comfort. In fact, the model: $\log(Y_i) = \log(N_i) - 0.27 + 0.17 * (\text{watch-}\mu\text{EMA})_i$ suggests that watch- μ EMA participants are 1.19 ($e^{0.17}$, 95% C.I.: 1.05, 1.36) times more likely to respond to a delivered prompt. Thus, watch- μ EMA resulted in 19% higher completion rates than phone-EMA, which is more than 1 SD of watch-EMA completion rate.

Finally, watch- μ EMA had significantly higher first prompt response rate than watch-EMA. For first prompt response rate, the same model fitting procedure yielded no significant impact of age, gender, and technology comfort. The resulting model: $\log(Y_i) = \log(N_i) - 0.32 + 0.20 * (\text{watch-}\mu\text{EMA})_i$ suggests that watch- μ EMA participants are 1.22 ($e^{0.20}$, 95% C.I.: 1.04, 1.44) times more likely to respond to a first delivered prompt on a watch. In other words, watch- μ EMA requires less number of re-prompts than watch-EMA, resulting in 22% higher response rate for the first delivered prompt (i.e. no reprompt), which is more than 1 SD of watch-EMA first prompt response rate.

None of these watch-EMA vs watch- μ EMA response-rate models showed any evidence of a lack-of-fit, at the $\alpha=0.05$ level. In fact, in the previous study, switching from phone-EMA to watch- μ EMA resulted in 25% increase in compliance (> 1 standard deviation of phone-EMA compliance), 35% increase in completion (> 2 standard deviation of phone-EMA completion), and 65% increase in first prompt response rate (> 2 standard deviation of phone-EMA first prompt response rate). Table 4 below summarizes these results.

4.3 Perceived Burden and Overall Experience

The perceived burden questionnaire had two, 5-point (1: Strongly Disagree to 5: Strongly Agree) questions: “Do you think prompts interrupted you?” and “Do you think prompts distracted you from what you were doing?” Responses “agree” and “strongly agree” were recorded as high distraction and interruption. Fig. 2 (right) summarizes weekly trends of perceived burden as percentage of participants reporting high distraction and interruption from the prompts. Watch-EMA is perceived as being more distracting and interrupting than phone-EMA. Nevertheless, perceived distraction in watch-EMA decreases towards the end of the study, which increases for phone-EMA. However, in the prior pilot study, watch- μ EMA was perceived as more interrupting but less distracting than phone-EMA. In fact, despite being highly interruptive (~ 8 times more than phone/watch-EMA), watch- μ EMA is still less distracting. Participants were also asked to describe their positive and negative experiences of participation and responding to the prompts. These questions were asked as part of the perceived burden survey sent via email on the last day of the study, and responses for the phone-EMA and watch- μ EMA conditions are in the prior work [21]. In terms of negative experiences, three watch-EMA participants reported that prompts “*interrupted while working or driving*” and “*vibrated inappropriately*.” Four participants mentioned frequent battery drainage of their smartphones due to the smartwatch. In addition, three participants said the smartwatch interface was “*complicated to use*,” inviting “*accidental*

responses.” More positively, watch-EMA participants reported the watch to be “*fun to use*” and “*a new experience.*”

4.4 Compliance and Completion Trends

The response rates drop for all the conditions for 4-weeks, which is consistent with the previous phone-EMA results. However, in watch- μ EMA, the rates remained higher throughout the study duration despite the intensive interruption of ~ 6 times per hour, for 12 hours per day, for 4 weeks. Watch-EMA had the lowest response rates throughout the study duration, except for the first seven days. The gap between watch-EMA and watch- μ EMA compliance and completion rates may have resulted from a combination of smartwatch batteries draining and watches occasionally un-pairing with the phone (an Android bug observed by the participants). Thus, watch-EMA participants received only 78% of 168 scheduled prompts per participant, watch- μ EMA participants received only 88% of ~ 1200 prompts per participant, and phone-EMA received 96% of 168 prompts per participant. Even though watch-EMA had higher average completion rate than phone-EMA, it dropped rapidly ten days into the study. Overall, it suggests that participants responding to watch- μ EMA could sustain the interruption for longer duration than watch-EMA and phone-EMA.

5 DISCUSSION

This study showed no statistically-significant difference between compliance, completion, and first-prompt response rates between watch-EMA and phone-EMA in a 4-week period. We have distinguished compliance and completion because the EMA literature is inconsistent in how “compliance” is reported, conflating these ideas; therefore, we include both. A study can have a high completion rate but low compliance if someone turns off the device – a practical issue in real, longer-term studies. Due to the ease of access and the use of wrist-worn vibro-tactile notification achievable with the smartwatch, we might expect compliance to improve when moving surveys from phone to watch. The lack of difference observed here could result from the burden of learning and maintaining the smartwatch, because our participants were not current smartwatch users. The qualitative responses support this possibility. Three participants reported that the watch interface was too small to interact with. If the primary burden accumulation when completing an EMA protocol is not from prompting interrupting the current activity, but rather from the time it requires to respond to any given prompt, then the difference between completing the same surveys on the phone and the watch when surveys consist of multiple question sets might be modest, especially if individuals keep their phones easily accessible.

Our results are consistent with those from Hernandez et al. [18], comparing watch-based EMA, phone-based EMA, and Google-Glass-based EMA. In a 5-day study, they too found no significant difference in the response rates of different devices, suggesting that translating a survey from the phone to the watch does not necessarily reduce burden or increase compliance. In that study, however, participants had to manage and wear three devices simultaneously, including Google Glass, in addition to other physiological sensors and narrative clip cameras, and the use of each device may have reinforced the use of the other. Hernandez et al. also do not appear to account for data losses due to battery or connectivity

loss between the devices to enable teasing apart compliance and completion. In our work, we stay consistent with most EMA work, where only a single device is used at a time, and for an extended time (~6 times as long as [18]), so the novelty can wear off.

In contrast with the watch-EMA vs. phone-EMA comparison, this study showed a statistical difference between compliance, completion, and first-prompt response rates in the watch-EMA vs. watch- μ EMA conditions, where watch- μ EMA prompted at a rate of ~8 times watch-EMA. Since the watch familiarity and maintenance issues would be similar in both cases, this change in compliance could be explained by our hypothesis that at-a-glance microinteraction may keep perceived burden manageable, even at high temporal density. In watch- μ EMA, responding to a prompt involves answering just one question with a yes/maybe/no answer; i.e., there is a momentary microinteraction designed to minimize attentional disruption with an interface designed to facilitate fast interaction (no scrolling, 3 answers, single tap). Because this interaction may result in comparatively little distraction from activities, and because every prompt is guaranteed to be of the same microinteraction style, this may result in less hesitation in initiating a response to a prompt on the part of a participant. Participants who know with certainty that responding to any prompt will take no longer than ~2s, with a single tap, may be more likely to do so, and to do so on the first prompt rather than waiting for a re-prompt. Prompt salience and ease-of-access to the device delivering EMA likely influence participant response behavior, but it may be that the complexity of interaction that is required following the prompt dominates the participant's decision to answer any given prompt, or skip it. Even the quick scroll required to answer each watch-EMA question, not required for watch- μ EMA, could slow a person down.

The total number of delivered prompts for phone-EMA was higher than watch-EMA and watch- μ EMA. However, lower prompt delivery due to the burden of maintaining an additional device (smartwatch) in watch- μ EMA did not discourage participants from initiating a response. On the other hand, additional burden of maintaining the device coupled with complex interaction in watch-EMA could have resulted in lower response rates. It is possible that this complexity could have demotivated the participants to keep the device charged. Nevertheless, future research should systematically explore the device maintenance behavior in EMA studies and its impact on compliance. Further, one might argue that 11.5s of vibration may be excessive on the watch compared with the phone. However, despite using the same vibration pattern for watch-EMA and watch- μ EMA, there were no significant differences in watch-EMA and phone-EMA response rates. It is possible that participants might have gotten used to the pattern, thereby rendering the vibration as less attention grabbing. This also presents an interesting research opportunity: exploring appropriate vibration patterns for smartwatches to ensure high compliance in watch- μ EMA.

We intentionally did not use any financial compensation in this work, unlike most EMA studies, which often financially compensate for high compliance to encourage the behavior. Compliance-contingent monetary compensation is hard to sustain affordably for large scale longer-term studies (e.g., [14]), especially for studies intended to last much longer than a month. Nevertheless, it is likely that financial compliance would improve overall compliance rates for all EMA methods, and it could slow compliance drop-off. We expect, however, that differences between μ EMA and non- μ EMA techniques would still be observed.

6 POTENTIAL APPLICATIONS OF μ EMA

This pilot study suggests that watch-based μ EMA might permit data collection of self-report data at a high temporal density with respectable compliance relative to standard EMA implemented on a phone or a watch, even after weeks of use. Here we discuss four ways that the μ EMA methodology might be used.

6.1 Fragmentation in time

A composite scale with multiple items used in traditional EMA could be broken into individual items delivered one at a time, where each item can be answered in a microinteraction. This is what we have done in the pilot study. In contrast to all six EMA questions appearing on phone-EMA and watch-EMA back-to-back after a single prompt, in watch- μ EMA, participants received one-prompt, one-question microinteractions. All six questions were administered within a two-hour window, just not back-to-back. After four weeks, μ EMA's low perceived burden may provide the same information for each two-hour window, but at a higher response rate. Similarly, many health-related questionnaires that require respondents to recall their health state for the past week (e.g., Neuropathic Pain Scale [15] and International Physical Activity Questionnaire (IPAQ) [8]) might be fragmented in time to lower perceived burden, even when administered for many weeks. Some specific questions may also require fragmentation, breaking them into parts, to ensure they can be answered in a glance, but then data may simply be aggregated into a single response.

6.2 Single Construct, Single Item

Composite scales/questionnaires that measure a single construct with multiple items can sometimes be reduced to single-item formats. These single-item scales could then be administered using μ EMA with high temporal density. Developing and validating such measures can be complex. Nevertheless, in domains such as marketing research (e.g., [5]) and organizational psychology (e.g., [16]), researchers have found no difference in predictive validity of some single- and multiple-item questionnaires measuring the same construct. Similarly, Robins et al. [35] have found strong convergent validity between a single-item self-esteem scale and Rosenberg's self-esteem scale (7-items) [36]. Finally, a single-item health-related quality of life questionnaire has been found to significantly correlate with the overall score of a standard health-related quality of life questionnaire [9]. In some cases, converting these single-items to microinteraction single items may require additional validation if they have many, or complex, answer items, but μ EMA may be relatively easy to deploy in measurement domains that already have single-item construct measurement tools.

6.3 Event Markers

μ EMA could be used to measure behaviors/states that are already known to change or occur with high frequency throughout the day, but where the patterns of those changes relative to other contextual factors (some of which could be measured passively) are not well understood. Examples are chronic pain experiences [20], posttraumatic stress disorder (PTSD) episodes [28], happiness [1] or cognitive alertness [2]. In this case, μ EMA could be

used to mark events as “occurring” or “not occurring” at unprecedented temporal density. For instance, changes in chronic pain intensity (e.g., for lower back pain or fibromyalgia) can be captured using single-item pain intensity scales such as numeric rating scales, verbal rating scale or a visual analogue scale [20]. Likewise, a recent n-of-1 case study with a PTSD patient demonstrated a single button, wrist-worn interface could be used to mark precursors of episodes of hyperarousal at high temporal density both for scientific study and possible intervention [28]. Such markers could prove valuable for measurement of rapidly changing phenomena, to develop person-specific models of distributions of events that might drive delivery of interventions, validate other more sophisticated measures, or even to help train or validate machine learning systems that need data that are densely labeled.

6.4 Context-sensitive, Computer-adaptive Modelling

Lastly, μ EMA could be integrated with other mobile sensing capabilities enabled by mobile devices, such as always-on sensing of motion, location, device use, and social interaction, to drive computer-adaptive sampling of behavior, state, and/or contexts. In this case, rather than repeatedly asking the same questions, a mobile device such as a phone, smartwatch, or head-mounted computer use context-sensitive μ EMA [22]. Sensors such as heart rate sensors or galvanic skin response sensors might provide data that trigger selective use of μ EMA self-report, incrementally building up models based on item-response theory. Just as computer-adaptive tests can reduce a long self-report instrument to just a few questions [11], sensor-driven computer-adaptive sensing may make it possible to distill all required self-report information desired at a given moment in time down to a single, carefully-timed μ EMA survey. If μ EMA surveys can be presented many times per hour, one can imagine continuously and incrementally building up sophisticated models of individual-level behaviors, and only asking for self-report input when predictions must be confirmed.

7 LIMITATIONS

Although providing preliminary evidence that μ EMA is a technique worth further exploration, this pilot study has several limitations. The 4-week watch-EMA data collection was limited to only 10 participants due to the availability of smartwatches and participants with phones running Android 4.3+. The watch- μ EMA (19 participants) and phone-EMA (14 participants) had similarly small sample sizes as well. Moreover, although the 4-week study is substantially longer than all prior work with smartwatches, future work could replicate this study with a larger study population for a longer study duration. To keep the experiment as controlled as possible, only people who did not already use smartwatches were recruited. At the time of this study, Android wear software had alarm-related bugs and inconsistencies (observed in our pre-pilot testing), and so to ensure that the watch prompted accurately, we used the same watch model for all participants (which was the same as used in the watch- μ EMA condition in the prior study). While we could be certain of our data using this method, it prevented us from recruiting people who were already comfortable using a smartwatch daily. Recruiting enough such participants, all with the identical smartwatch model, would have been, and still would be, challenging given the scarcity of smartwatch wearers and the large number of models available. It is not uncommon in EMA studies in some research domains (e.g., health) to give out dedicated hardware to participants (e.g.,

[18]). Our study was limited to Android watches, because iOS devices do not provide the same level of control over prompting and use of continuous logging.

Our study did not gather information on each participant's specific educational background, although all participants are known to have at least a high school level of education. It is possible that individuals with certain domain experience (e.g., computer science, behavioral science, and statistics) may be more familiar with EMA and related data collection methodologies. This could influence compliance. However, it is also possible that a study duration of four weeks could be long enough to counter the influences of any such prior experience with EMA.

In this study, watch-EMA did not allow notifications from other applications on the smartphone, but notifications were not restricted for phone-EMA. It is likely that compliance rates are impacted by how someone uses his or her smartphone, although we are aware of no studies that have studied this impact in a systematic way. It is possible that allowing additional notifications on the watch from any app could desensitize participants towards our watch- μ EMA or watch-EMA prompts, but it is also possible such notifications could incentivize participants to check watch notifications more often. Future work could explore how use of personal apps, on both phones and smartwatches, influences EMA or watch- μ EMA compliance. Evolving usage patterns must be kept in mind going forward with both techniques as researchers exploit devices that participants already own and use for personal communication.

This work is focused only on response rates, not the validity of data obtained using watch-EMA, watch- μ EMA, or phone-EMA. Establishing EMA data validity is dependent on the research construct and experiment goals – a different validation with additional data collection on ground truth (or equivalent) is required for each construct. Such studies are complex, even for phone-EMA, where a validated paper survey is converted to EMA on the phone. A new data gathering method that changes the way questions are asked from sets of back-to-back, multiple-choice questions prompted (relatively) infrequently, to single, one-tap answers prompted at an intensive rate, but answered 'at a glance,' clearly will require extensive validation studies in each domain of interest. Nevertheless, we have checked the distribution of responses we obtained across the three conditions and did not find any pattern that would indicate that watch- μ EMA influences participants' responses in a specific way. This pilot study, however, suggests that despite what might appear to be an unusually high interruption rate, the technique may have acceptable compliance, at least in a 4-week deployment. This may motivate additional research on methods to exploit the μ EMA strategy, on watches or other devices, to support research in health, ubiquitous computing and other domains where *in situ*, self-report measurement with high temporal density may be important.

8 CONCLUSIONS

The results from this pilot study suggest that a mere adaption of phone-EMA to watch-EMA may not dramatically impact study compliance, completion and first prompt response rates. However, watch- μ EMA, may permit collection self-report with approximately eight times as

much temporal density as traditional EMA, but still at an acceptable level of burden. Validating that the information collected using microinteractions is of value in any domain requires domain-dependent studies. However, the pilot experiment demonstrates that μ EMA strategy may offer a new type of self-report data collection that warrants further study, especially for behaviors, states, or contexts known to require self-report and change frequently in our everyday lives, and especially when combined with passive data collection that could drive context-sensitive EMA and μ EMA.

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 μ EMA in behavioural science, the anonymous reviewers for helpful feedback that has significantly increased manuscript clarity, and Maciej Kos, and Anmol Sakarda for valuable writing suggestions.

This work was made possible, in part, by a Google Glass Research Award and funding from the NIH's National Heart, Lung and Blood Institute (R21 HL108018-01). Authors' addresses: A. Ponnada, D. Maniar, C. Haynes, and S. Intille, 910-177, 360 Huntington Avenue, Northeastern University; J. Manjourides, 312 Robinson Hall, 360 Huntington Avenue, Northeastern University, Boston, MA, 02115, USA.

REFERENCES

- [1]. Ahmed M Abdel Khalek. 2006 Measuring happiness with a single-item scale. *Social Behavior and Personality: An International Journal* 34 2: 139–150.
- [2]. Abdullah Saeed, Murnane Elizabeth L., Matthews Mark, Kay Matthew, Kientz Julie A., Gay Geri, and Choudhury Tanzeem. 2016 Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)* ACM, New York, NY, USA, 178–189.
- [3]. Ashbrook Daniel L., Clawson James R., Lyons Kent, Starner Thad E., and Patel Nirmal. 2008 Quickdraw: The impact of mobility and on-body placement on device access time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)* ACM, New York, NY, USA, 219–222.
- [4]. Beal Daniel J. and Weiss Howard M.. 2003 Methods of ecological momentary assessment in organizational research. *Organizational Research Methods* 6, 4:440–464.
- [5]. Bergkvist Lars and Rossiter John R.. 2007 The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research* 44 2: 175–184.
- [6]. Berlyne Daniel E.. 1950 Novelty and curiosity as determinants of exploratory behavior. *British J of Psychology* 41, 1-2:68–80.
- [7]. Consul Prem C., and Jain Gaurav C.. 1973 A generalization of the Poisson distribution. *Technometrics* 154: 791–799.
- [8]. Craig Cora L., Marshall Alison L., Sjöström Michael, Bauman Adrian E., Booth Michael L., Ainsworth Barbara E., and Pratt Michael 2003 International Physical Activity Questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise* 35 8: 1381–1395. [PubMed: 12900694]
- [9]. Cunny Kelley A. and Perri Matthew, III. 1991 Single-item vs multiple-item measures of health-related quality of life. *Psychological Reports* 69 1: 127–130. [PubMed: 1961779]
- [10]. Curran Shelly A., Beacham Abbie O., and Andrykowski Micheal A.. 2004 Ecological momentary assessment of fatigue following breast cancer treatment. *Journal of Behavioral Medicine* 27, 5:425–444. [PubMed: 15675633]
- [11]. Cella David, Gershon Richard, Lai Jin-Shei, and Choi Seung. 2007 The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research* 16 1: 133–141. [PubMed: 17401637]

- [12]. 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 the 13th International Conference on Ubiquitous Computing (UbiComp '11). ACM, New York, NY, USA, 163–172.
- [13]. Dunton Genevieve F., Kawabata Keito, Intille Stephen S., Wolch Jennifer and Pentz Marry A.. 2011 Physical and social contextual influences on children's leisure-time physical activity: An ecological momentary assessment study. *Journal of Physical Activity and Health* 26, 3: 135–142.
- [14]. Collins Francis S. and Varmus Harold. 2015 A new initiative on precision medicine. *New Eng. J. Med* 372: 793–795. [PubMed: 25635347]
- [15]. Galer Bradley S. and Jensen Mark P.. 1997 Development and preliminary validation of a pain measure specific to neuropathic pain: The Neuropathic Pain Scale. *Neurology* 48 2: 332 – 338. [PubMed: 9040716]
- [16]. Gardner Donald G., Cummings Larry L., Dunham Randall B., and Pierce Jon L.. 1998 Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement* 58 6: 898–915.
- [17]. Hawkey Louise C.. 2003 Loneliness in everyday life: Cardiovascular activity, psychosocial context, and health behaviors. *Journal of Personality and Social Psychology* 85, 1:105–120. [PubMed: 12872887]
- [18]. Hernandez Javier, Daniel McDuff Christian Infante, Maes Pattie, Quigley Karen, and Picard Rosalind. 2016 Wearable ESM: Differences in the experience sampling method across wearable devices. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Sendeeds (MobileHCI '16). ACM, New York, NY, USA, 195–205.
- [19]. Heron Kristin E. and Smyth Joshua M.. 2010 Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behavior treatments. *British Journal of Health Psychology* 15: 1–39. [PubMed: 19646331]
- [20]. Hjermsstad Marianne J., Fayers Peter M., Haugen Dagny F., Caraceni Augusto, Hanks Geoffrey W., Loge Jon H., Fainsinger Robin, Aass Nina, and Kaasa Stein. 2011 Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: A systematic literature review. *Journal of Pain and Symptom Management* 41 6: 1073–1093. [PubMed: 21621130]
- [21]. Intille Stephen, Haynes Caitlin, Maniar Dharam, Ponnada Aditya and Manjourides Justin. 2016 μ EMA: Micro-interactions based ecological momentary assessments (EMA) using a smartwatch. In Proceedings of the ACM International Conference on Ubiquitous and Pervasive Computing (UbiComp' 16). ACM, New York, NY, USA.
- [22]. Intille SS. 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.
- [23]. Timmermann Janko, Heuten Wilko, and Boll Susanne. 2015 Input methods for the Borg-RPE-scale on smartwatches. In 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'15): 80–83.
- [24]. Kikuchi Hiroe, Yoshiuchi Kazuhiro, Yamamoto Yoshiharu, Komaki Gen and Akabayashi Akira. 2011 Does sleep aggravate tension-type headache?: An investigation using computerized ecological momentary assessment and actigraphy. *Biopsychosoc Med* 5,10. [PubMed: 21835045]
- [25]. Kim Jinhyuk, Kikuchi Hiroe, and Yamamoto Yoshiharu. 2013 Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *British J of Health Psych* 18:155–67.
- [26]. Kim Jinhyuk, Nakamura Torn, Kikuchi Hiroe, Saaki Tsukasa, and Yamamoto Yoshiharu. 2013 Co-variation of depressive mood and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. *PloS One* 8: e74979. [PubMed: 24058642]
- [27]. Kotz Samuel, Balakrishnan Narayanswamy, and Johnson Norman L.. 2004 Continuous multivariate distributions, models and applications. John Wiley & Sons, New Jersey, USA.
- [28]. Eg Jacob, Larsen Kasper Eskelund, and Thomas B. Christiansen. 2011 Active self-tracking of subjective experience with a one-button wearable: A case study in military PTSD. 2nd Symposia

- on Computing and Mental Health. in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver CO.
- [29]. Lyons Kent. 2015 Using digital watch practices to inform smartwatch design. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15). ACM, New York, NY, USA, 2199–2204.
- [30]. Marcello Pagano and Gauvreau Kimberlee. 2000 Principles of Biostatistics. Vol. 2 Pacific Grove, CA.
- [31]. McFadden Daniel. 1973 Conditional logit analysis of qualitative choice behavior, Berkeley, USA.
- [32]. Min Chulhong, Kang Seungwoo, Yoo Chungkuk, Cha Jeehoon, Choi Sangwon, Oh Youngha, and Song Juneha. 2015 Exploring current practices for battery use and management of smartwatches. In Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15). ACM, New York, NY, USA, 11–18.
- [33]. Pizza Stefania, Brown Barry, Donald McMillan, and Airi Lampinen. 2016 Smartwatch in vivo. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 5456–5469.
- [34]. Proudfoot Judith. 2013 The future is in our hands: The role of mobile phones in the prevention and management of mental disorders. Australian and New Zealand Journal of Psychiatry, 47 2: 111–113. [PubMed: 23382507]
- [35]. Robins Richard W., Hendin Holly M., and Trzesniewski Kali H.. 2001 Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. Personality and Social Psychology Bulletin 27 2: 151–161.
- [36]. Rosenberg Morris. 1965 Society and the adolescent self-image. Vol. 11 Princeton University Press Princeton, NJ.
- [37]. Shiffman Saul, Stone Arthur A., and Hufford Michael R.. 2008 Ecological momentary assessment. Annual Review Clinical Psychology 4:1–32.
- [38]. Shiffman Saul. 2000 Real-time self-report of momentary states in the natural environment: Computerized ecological momentary assessment. The Science of Self-report: Implications for Research and Practice: 277–296.
- [39]. Stone Arthur A. and Shiffman Saul. 1994 Ecological momentary assessment (EMA) in behavioral medicine. Annals of Behavioral Medicine 16, 3:199–202.
- [40]. Stone Arthur A., Shiffman Saul, Atienza Audie A. and Nebeling Linda. 2007 Historical roots and rationale of ecological momentary assessment (EMA) In The Science of Real-Time Data Capture: Self-Reports in Health Research, Stone AA, Shiffman S, Antienza A, and Nebeling L, Editors. Oxford University Press: New York, NY:3–10.
- [41]. 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.
- [42]. Watson David, Clark Lee A., and Tellegen Auke. 1988 Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal Personality and Social Psychology 54, 6:1063–70.
- [43]. Worldwide Wearables Market Increases 67.2% Amid Seasonal Retrenchment, According to IDC. 2016 Retrieved September 6, 2016 from <http://www.idc.com/getdoc.jsp?containerId=prUS41284516>.

CCS Concepts: → **Human-centered-computing** → **Ubiquitous and mobile computing** → **Empirical studies in mobile and ubiquitous computing**; • **Human-Computer-Interaction** → **HCI design and evaluation methods**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

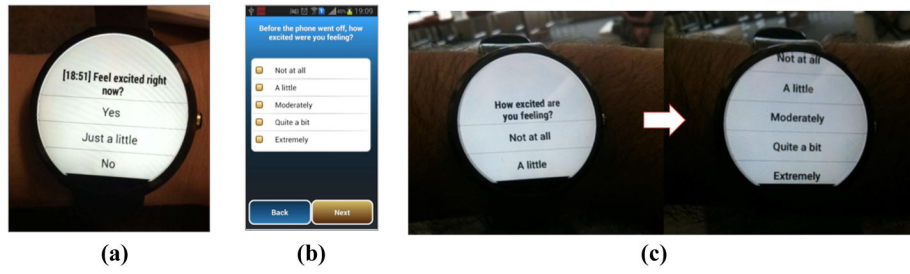


Fig. 1. (a) Single question with yes/just-a-little/no answer on watch- μ EMA (from [21]), (b) One of the 6 questions in each phone-EMA question set (from [21]), and (c) One of the 6 questions in the watch-EMA question set, showing required scrolling to view all the answer options (from this study).

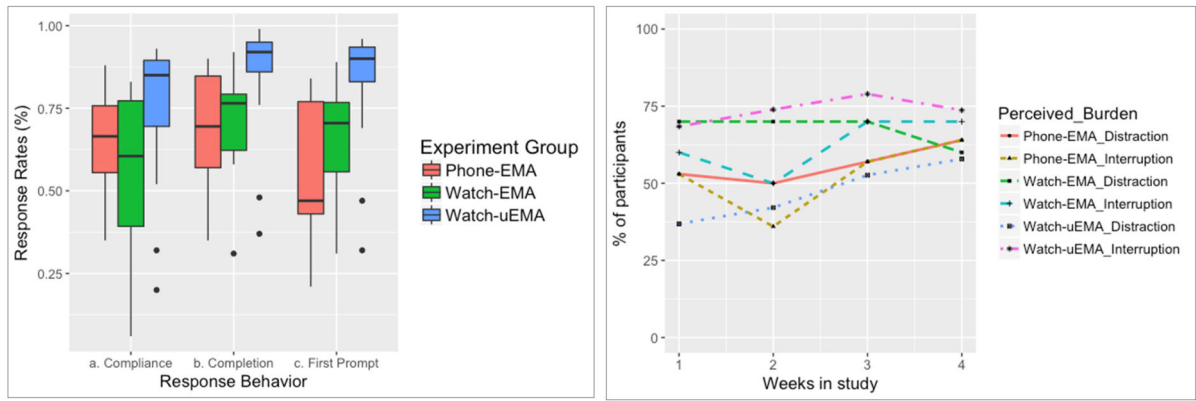


Fig. 2. (Left) Response behavior of watch-EMA compared with phone-EMA and watch- μ EMA (from the previous study) including outliers. (Right) Perceived burden (% participants reporting high interruption and high distraction) of watch-EMA compared with phone-EMA and watch- μ EMA (from the previous study) over 4-weeks.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

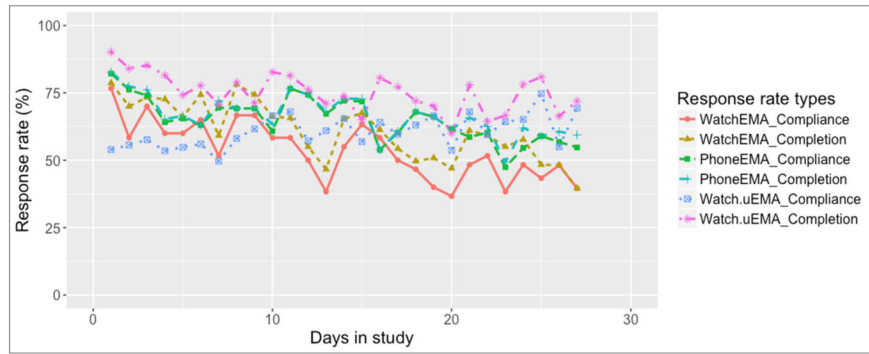


Fig. 3. Trends in compliance and completion rates of watch-EMA compared with phone-EMA and watch- μ EMA (from the previous study) for a period of 4-weeks.

Table 1.

Demographics of watch- μ EMA (previous study), phone-EMA (previous study), and watch-EMA groups. The (highlighted) watch-EMA data was newly-collected for this study.

	Watch- μ EMA	Phone-EMA	Watch-EMA
All	19(100%)	14(100%)	10(100%)
Male	9(47.4%)	6(42.9%)	7(70%)
Female	10(52.6%)	8(57.1%)	3(30%)
Age (min, max)	24.6(18,55)	28.3(18, 52)	23.7(19,32)
Comfort with Technology (1-item questionnaire)			
Strongly	13(68.4%)	7(50%)	8(80%)
Mostly	5(26.3%)	6(42.9%)	2(20%)
Somewhat	1(5.3%)	1(7.1%)	0(0%)
Neutral/Not	0(0%)	0(0%)	0(0%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Summary of response rates in the phone-EMA and watch- μ EMA conditions (prior study), and the watch-EMA condition from this study. Watch- μ EMA response rates highlighted for emphases.

	Number of prompts and responses					
	With outliers			Without outliers		
	Scheduled	Delivered (% Delivered)	Completed	Scheduled	Delivered (% Delivered)	Completed
Phone-EMA	2395	2295(96%)	1546	2395	2295(96%)	1546
Watch- μ EMA	20929	18028(86%)	15831	18813	16641(88%)	15278
Watch-EMA	1680	1200(71%)	905	1512	1168(77%)	895

	Response rates (%)					
	With outliers			Without outliers		
	Compliance	Completion	First prompt	Compliance	Completion	First prompt
Phone-EMA	64.54%	67.36%	53.28%	64.54%	67.36%	53.28%
Watch- μ EMA	75.54%	87.81%	84.95%	81.21%	91.81%	88.33%
Watch-EMA	54.9%	75.4%	71.3%	59.91%	76.62%	72.43%

Table 3.

Model parameters for watch-EMA vs phone-EMA comparison

Response rate	Coeff. (β_1)	Std. Error	95% C.I. (β_1)	RR (e^{β_1})	95% (RR)	C.I.	p -value
Compliance	-0.105	0.13	-0.36, 0.15	0.90	0.70, 1.16		0.42
Completion	0.108	0.095	-0.08, 0.29	1.11	0.92, 1.34		0.27
First prompt	0.280	0.14	0.009, 0.55	1.32	1.00, 1.73		0.053

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.Model parameters for watch- μ EMA vs watch-EMA comparisons

Response rate	Coeff. (β_1)	Std. Error	95% C.I. (β_1)	95% (RR)	C.I	RR (e^{β_1})	p -value
Compliance	0.32	0.14	0.047, 0.609	1.04, 1.84	1.38	0.034	
Completion	0.17	0.07	0.045, 0.309	1.05, 1.36	1.19	0.016	
First prompt	0.20	0.081	0.043, 0.361	1.04, 1.44	1.22	0.020	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript