

# BMJ Open Influence of ecological momentary assessment study design features on reported willingness to participate and perceptions of potential research studies: an experimental study

Joshua M Smyth <sup>1</sup>, Dusti R Jones <sup>1</sup>, Cheng K F Wen <sup>2</sup>,  
Frank T Matera <sup>1</sup>, Stefan Schneider <sup>2</sup>, Arthur Stone <sup>3</sup>

**To cite:** Smyth JM, Jones DR, Wen CKF, *et al.* Influence of ecological momentary assessment study design features on reported willingness to participate and perceptions of potential research studies: an experimental study. *BMJ Open* 2021;**11**:e049154. doi:10.1136/bmjopen-2021-049154

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-049154>).

Received 19 January 2021  
Accepted 08 July 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Behavioral Health and Medicine, Penn State University Park, University Park, Pennsylvania, USA

<sup>2</sup>Dornsife Center for Self-Report Science, University of Southern California, Los Angeles, California, USA

<sup>3</sup>Psychology and Dornsife Center for Self-Report Science, University of Southern California, Los Angeles, California, USA

## Correspondence to

Professor Joshua M Smyth;  
jms1187@psu.edu

## ABSTRACT

**Objective** Intensive ambulatory assessment, such as ecological momentary assessment (EMA), is increasingly used to capture naturalistic patient-reported outcomes. EMA design features (eg, study duration, prompt frequency) vary widely between studies, but it is not known if such design decisions influence potential subjects' willingness to participate in a study. We hypothesise that intentions to participate will be higher in studies that are less burdensome and have higher reward (eg, compensation).

**Design** This experimental study examined if four EMA study design features (study duration, prompt frequency, prompt length, compensation) affected intentions to participate in a hypothetical EMA study and participation appraisals (eg, participation effort). Participants were randomly assigned to conditions (reflecting a fully crossed design of the four features, each with two levels). Each condition presented a vignette describing a study (each a unique combination of design features) and asked them to report on likelihood of participating and study appraisals.

**Participants** A convenience sample of participants (n=600; 46% female, *Age*=40.39) were recruited using an online service.

**Primary and secondary outcome measures** Primary outcomes were willingness to participate (No/Yes) and reported participation likelihood (0–100 scale). Secondary outcomes included appraisals of interest, enjoyment, effort, and if the study makes a valuable contribution to science.

**Results** We examined main effects, and two-way interactions for participation likelihood, across study design features. Overall, reported willingness to participate and participation likelihood were high (89%, *M*=83.90, respectively). Shorter study duration, fewer prompts, shorter prompts and higher compensation increased willingness to participate and elicited higher participation likelihood (each associated with ~6%–8% increases). Findings suggested that more intensive studies were judged as somewhat less interesting and enjoyable, and requiring more effort.

**Conclusion** Hypotheses were generally supported. Design features influence behavioural intentions to participate in, and appraisals of, EMA studies. Implications for participant

## Strengths and limitations of this study

- To provide stronger evidence of causality, we tested the effects of ecological momentary assessment (EMA) design features using an experimental manipulation wherein participants were randomly assigned to 1 of 16 possible study design vignettes (ie, a fully crossed 2×2×2 factorial design).
- Four commonly altered EMA study design features were examined (duration of study involvement, frequency of EMA prompts per day, estimated length of time to complete each study prompt, and compensation), providing broad information regarding which EMA study design features may influence participation.
- Participants were recruited from Amazon's Mechanical Turk; the generalisability of data from MTurk convenience samples compared with population-based samples is still unclear.

recruitment and generalisability, and remaining research questions, are discussed.

## INTRODUCTION

Ecological momentary assessment (EMA) is a technique that assesses dynamic physiological, psychological and behavioural processes as they occur in real time and in natural contexts.<sup>1–3</sup> EMA is increasingly used in clinical and medical settings to obtain patient reported outcomes in everyday life. Because participants in EMA studies provide multiple assessments over a particular time (eg, six prompt responses per day for 3 days, typically on a smartphone or mobile device) as they go about their normal activities, data from EMA studies are particularly well suited to examine complex temporal processes, and also have the benefits of high ecological validity and reduced recall bias.<sup>2</sup> Although potentially of

great value, EMA studies are typically associated with relatively high participant burden, at least when compared with administration of questionnaires at a single point in time. As such, there is concern that self-selection bias in EMA studies may be high, although there is almost no empirical support for this contention, its magnitude, or for study design factors that may influence self-selection bias.<sup>4</sup>

To date, we know very little about what features of an EMA study may be related to willingness and acceptability to participate, or how design features influence self-selection into these studies (eg, related to participation burden).<sup>5–8</sup> There is certainly a history of evidence broadly consistent with the view that greater effort or burden required for study participation is related to less willingness to participate and/or lower study uptake,<sup>9–13</sup> but we are unaware of any work that has directly examined this for EMA studies (although several studies have examined associations between participant burden or study design with study compliance,<sup>3 14</sup> we are interested in factors that predict study uptake rather than compliance once enrolled).<sup>15</sup> There are several features of EMA studies that may influence participant burden—in particular, the length of a study (number of days of EMA), the frequency/intensity of sampling (the number of ‘beeps’ per day), and the time required to complete each momentary report. The potential effects of these features on participation rates may also be offset/balanced to a degree by other factors (eg, the amount of compensation, the perceived value of the research).<sup>16</sup> Indeed, researchers across disciplines have noted a dearth of literature examining how different design features of studies requiring the use of mobile and ambulatory technologies affect participants’ prospective acceptance and willingness to meaningfully engage in study requirements.<sup>17–23</sup>

Beyond the simple expectation that more demanding protocols would likely have lower uptake, it is unclear whether specific EMA design factors in fact influence the likelihood of someone participating and—if so—how they influence intended participation. For example, we do not know if potential participants are more sensitive to the length of the study, the number of prompts they would experience each day, or the length of those interruptions when they occur. Without such information, it becomes difficult to optimise the design of EMA studies; for example, is it more acceptable to have a shorter study with more frequent assessments or a longer duration study with fewer and briefer assessments (among the many possible combinations of even these few design features)?

A major roadblock to understanding factors impacting participant selectivity in EMA research is the high cost associated with each EMA study, making it prohibitive to systematically study these factors in an experimental fashion (ie, it is not feasible to run complex factorial EMA designs with sufficient sample sizes). We sought to provide information on several design features with some variation on the design features (rather than, eg,

conducting a more thorough parametric manipulation of one dimension, such as sampling density or compensation). Accordingly, we implement a vignette methodology as a cost-effective strategy for experimental research aiming to determine if several commonly adjusted study design factors are associated with uptake (ie, willingness to participate in a study)—beginning to provide data needed to optimise design decisions in the planning phase of applied EMA studies. We base the design on prior work that has examined factors that affect rates of study participation in survey research<sup>24</sup>; in this work, vignettes describing hypothetical details of the study were presented to each participant, followed by questions regarding likelihood of participation in the study.

### Present study and hypotheses

The primary aim (aim 1) of the present study was to examine the extent to which four EMA design features (study duration, frequency of EMA prompts, estimated length of time to complete each study prompt, and compensation) influence participants’ stated likelihood of participating in an EMA study. Because we also seek to understand ambivalence/certainty towards intentions to participate, we examine two participation outcomes: (1) a yes/no variable, hereafter termed ‘willingness to participate’, which involves a simple binary judgement on participation; and (2) how likely participants think it is that they would actually participate in the study, hereafter termed ‘participation likelihood’. The latter measure is broadly based on theoretical frameworks from behavioural science (eg, Theory of Planned Behavior)<sup>25</sup> thought to reflect the degree of certainty that prospective participants will participate or not (understanding that such judgments are not fully accurate regarding future behaviour). That is, participants may indicate they will engage in endeavours they know they may later excuse themselves from—and the likelihood of participation variable may capture some of this uncertainty. Regarding study design features, we hypothesised that those studies with lower participant burden—that were of shorter duration (fewer days), with fewer prompts within each study day, with shorter prompt lengths (ie, less time required to complete a prompt)—and/or with higher compensation would be associated with being more willing to participate and with higher participation likelihood.

We also had several secondary study goals. First, because it is possible that certain study design features may interact with one another to influence participation likelihood (eg, shorter study prompts may have a stronger effect on participation likelihood when accompanied by fewer prompts per day), we examined two-way interactions among study design variables on an exploratory basis (aim 2a). Additionally, we explore appraisals about potential participation; that is, to what extent participating in the study would be rated as interesting, enjoyable, effortful, or making a contribution to science. We examine if study design features were associated with participation appraisals (Aim 2b), if participation

appraisals are linked with participation willingness and likelihood of participation (Aim 2c), and if study design features and participation appraisals interact to predict likelihood of participation (Aim 2d).

## METHODS

### Recruitment and study design

Other than participating in the study, it was not considered appropriate to involve patients or the public in the design, conduct, or reporting of this research. Participants were recruited for this online experimental study via Amazon Mechanical Turk (ie, MTurk), an online crowdsourcing website which allows participants to find paid online research study opportunities. The desired sample size of  $n=600$  was indicated on MTurk and payment for all participants' compensations was uploaded directly onto Amazon. Compensation for this study was \$10 per participant; the study took approximately 10 min to complete. Participants' survey data were verified by trained research assistants to ensure data quality, and successful study completion was indicated on MTurk for verified data. After verification, Amazon automatically and securely transferred compensation to participants based on their preferred payment method. Inclusion criteria were (1) being a current resident in the USA, (2) being over 18 years of age, (3) being fluent in English and (4) being a 'Master Qualified' MTurk worker on Amazon (a designation produced by Amazon reflecting a participant's previous completion of a high number of online tasks with exceptionally high-performance and reliability ratings in previous MTurk tasks).

MTurk workers were able to see the study description (ie, a brief explanation of the study, estimated time to take the survey, and compensation amount) on MTurk among a list of other research opportunities they were eligible for. Those interested selected the study's dedicated hyperlink which directed them to a web-based survey hosted by a secure Qualtrics account. After being screened for inclusion, eligible participants provided informed consent and were presented with a brief written introduction. Participants were told that in the survey a hypothetical research study would be described to them, and that they should imagine what it would be like to participate in that study. Participants then completed a demographic questionnaire after which they were automatically randomised (via Qualtrics) to 1 of 16 possible study design vignettes (ie, a fully crossed  $2 \times 2 \times 2 \times 2$  factorial design). Each vignette included a brief paragraph describing the hypothetical EMA study scenario and participation requirements that are typical in many EMA studies. These included four different study design features: duration of study involvement (subsequently referred to as 'study duration', conditions were participation for 3 days or 21 days); frequency of EMA prompts per day (subsequently referred to as 'prompt frequency', conditions were 3 prompts per day or 9 prompts per day); estimated length of time to complete each study prompt (subsequently referred to as 'prompt

length', conditions were 1 min or 4–5 min per prompt); and compensation (conditions were US\$50 or US\$150). All chosen study design features are well within the range of typical features for EMA studies (eg, most EMA studies have between 3 and 9 prompts per day).

### Patient and public involvement

There was no direct patient involvement for this work. Future work, especially that examining the association of EMA design features with study uptake in specific clinical populations would benefit from the inclusion of patient advisers.

### Measures

#### Participation

After reading the randomly assigned hypothetical EMA study vignette and its associated requirements, participants were asked to rate whether they would be willing to participate in the study (yes/no) and participation likelihood (ie, If you were offered participation in this study, how likely do you think it would be that you would actually participate, rated 0 'definitely would not participate'—100 'definitely would participate'). For the latter item, the starting value of the sliding scale was 0; we examined responses and observed no clustering around the starting default value.

#### Appraisals

Participation appraisals were chosen based on tenets from the Theory of Planned Behavior (TPB)<sup>25</sup> which posits that perceived behavioural intentions are influenced by attitudes (eg, initial appraisals of the study and its requirements) and are indicative (although imperfectly) of actual likelihood of engaging in a targeted behaviour (ie, self-selecting to actually participate in the study). Although we are not testing tenets of TPB, the TPB was used to identify on attitudes/appraisals that may (1) be influenced by study design factors, and (2) influence intentions to participate. Participants were asked to rate four dimensions related to their appraisals of the study: how *interesting* participating in the study would be, how *enjoyable* participating in the study would be, how much *effort* participation would require, and how much participating in the study would *contribute to science*. These participation appraisals were measured using a visual analogue slider scale ranging from 'Not at all' (0) to 'Very' (10), with a starting default of 0. Numeric values were hidden from participants (ie, participants could see 'Not at all' but not the '0' value). We again examined responses and observed no clustering around the starting value.

#### Careless responder and attention check

A single careless responder question (ie, What word is the same as ship? Correct answer: boat) was asked to ensure that participants were not randomly responding to questions. Additionally, a total of four attention check questions were asked, each corresponding to a study parameter (eg, How many days did participation in this study require?) to provide information as to whether



participants had adequately attended to the presented study design features.

### Analytic plan

Given our sample size of 600, three-way and four-way interactions would be statistically underpowered (eg, with fewer than 40 participants per cell in analyses for four-way interactions). As such, we opted not to conduct tests of three-way and four-way interactions, focusing on main effects and two-way interactions as these would provide robust evidence by leveraging power across cells (with approximately 300 participants per cell for main effects and 150 participants per cell for two-way interactions) to examine the influence of design features on intentions to participate. Power for regression (ie, using participation likelihood as a continuous variable) was therefore adequate (0.80) to detect small main effects ( $d=0.23$ ) and small-moderate two-way interactions ( $d=0.32$ ). Statistical power for our dichotomous outcome, particularly given the skewed response (ie, heavily weighted toward 'YES' responses at 0.89), was substantially lower; thus, we only examined main effects for this variable as even the two-way interactions would be statistically underpowered.

Data cleaning and modelling were performed in SAS V.9.4. As previously noted, Aim 1 of the present study was to examine whether study design features influence participation. All four study design features were entered as independent variables into regression models simultaneously (eg, in one full factorial model). To examine willingness to participate, a relative risk ratio model (using PROC GENMOD with robust standard errors) was specified to examine whether each study design parameter (higher study length, prompt frequency, prompt length, and lower compensation, were specified as referents) predicted willingness to participate (dummy coded: 0=no, 1=yes), (Given that distribution was heavily inflated by participants indicating that they would participate, we used a Poisson distribution rather than a binomial distribution in all analyses using risk ratios as these produce less biased results.) Multiple regression (using PROC GLM) was specified to examine if study design features predicted participation likelihood (0–100 ratings). To assess the sub-aim (aim 2a) of whether study design features interacted to predict participation likelihood, all two-way interactions were added to this multiple regression model. Importantly, because this was a secondary sub-aim, the results section first highlights the results from aim 1 rather than the higher order significant interactions. To address aim 2b, all four participation appraisals were simultaneously entered into relative risk and multiple regression models to examine whether appraisals influenced participation. Finally, multiple regressions were used to examine if study design features predicted appraisals (aim 2c), and whether study design and appraisals interacted to predict participation likelihood (aim 2d). Effect sizes were calculated for all associations; the risk ratio was used as an indicator of effect size for relative risk models and partial eta squared was used

**Table 1** Sample descriptive statistics

|                                    | Total N | Mean (or % frequency) | SD    | Range |
|------------------------------------|---------|-----------------------|-------|-------|
| Female                             | 599     | 46%                   | –     | –     |
| Married                            | 600     | 48%                   | –     | –     |
| Non-hispanic white                 | 596     | 78%                   | –     | –     |
| Hispanic                           | 595     | 6%                    | –     | –     |
| College degree                     | 599     | 58%                   | –     | –     |
| Income \$50k or higher             | 591     | 30%                   | –     | –     |
| Age                                | 600     | 40.39                 | 10.69 | 22–73 |
| Interest                           | 599     | 6.95                  | 2.73  | 0–10  |
| Enjoyable                          | 600     | 5.83                  | 2.82  | 0–10  |
| Effort                             | 599     | 6.11                  | 2.92  | 0–10  |
| Contribution to science            | 600     | 7.31                  | 2.13  | 0–10  |
| Estimated participation likelihood | 600     | 83.90                 | 24.92 | 0–100 |
| Willingness to participate         | 600     | 89%                   | –     | –     |

as an effect size for multiple regression models. Effect sizes for partial eta-squared are 0.01 for small effects, 0.06 for medium effects, and 0.14 for large effects.

## RESULTS

### Preliminary analyses

Two participants (0.33%) did not complete the survey; their data were removed from analyses. All but one participant answered the careless responder question correctly and most participants (87%) answered the majority of (ie, three or four of four) attention check questions correctly. Finally, although time spent reading the vignettes varied, participants spent an average of 68s reading the study vignette ( $SD=76.03$ , range 1.42–839.41) (results from aim 1 were conducted using all those randomised; these findings held and became slightly stronger when those who answered the careless responder item incorrectly, those who answered two or fewer attention checks correct, and those from the lowest 5% of reading time are removed from analyses).<sup>26–28</sup> Descriptive statistics are presented in table 1.

### Aim 1: main effects of study design features on intentions to participate

Our primary outcomes of interest were the two variables capturing reported intention to participate in the study. Across all conditions, 89% of participants indicated 'YES' (versus 'NO') they were willing to participate; estimates of participation likelihood (%) were similarly high ( $M=83.90$ ;  $SD=24.92$ ). As expected, the dichotomous response of willingness to participate was significantly correlated with

**Table 2** Participation responses for each study condition

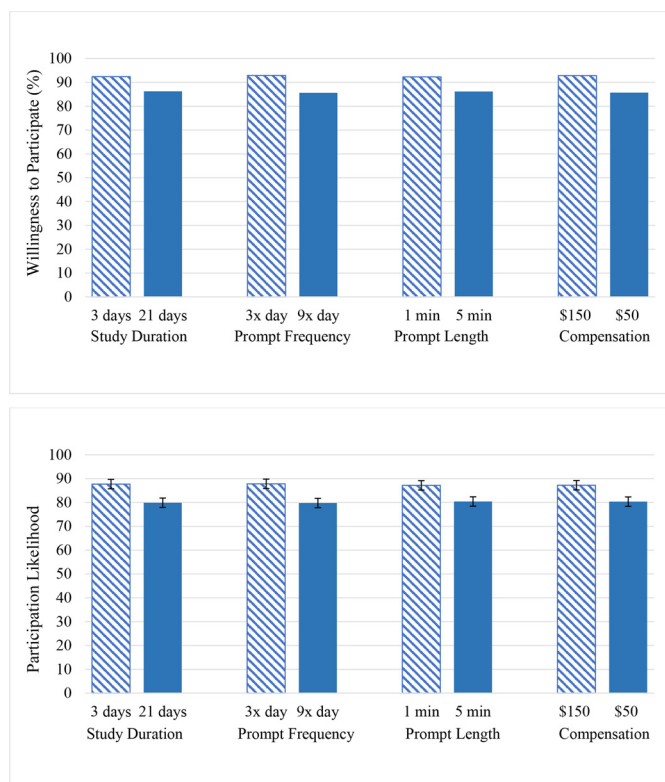
|   | Willingness to participate |      | Participation likelihood |
|---|----------------------------|------|--------------------------|
|   | % Yes                      | % No | M                        |
| Study duration=3 days, Prompt frequency=3 ×/day, Prompt length=1 min, Compensation = US\$50   | 92                         | 8    | 87.08                    |
| Study duration=3 days, Prompt frequency=3 ×/day, Prompt length=1 min; Compensation = US\$150  | 97                         | 3    | 92.74                    |
| Study duration=3 days, Prompt frequency=3 ×/day, Prompt length=5 min, Compensation = US\$50   | 91                         | 9    | 89.20                    |
| Study duration=3 days, Prompt frequency=3 ×/day, Prompt length=5 min, Compensation = US\$150  | 97                         | 3    | 91.87                    |
| Study duration=3 days, Prompt frequency=9 ×/day, Prompt length=1 min, Compensation = US\$50   | 87                         | 13   | 84.41                    |
| Study duration=3 days, Prompt frequency=9 ×/day, Prompt length=1 min, Compensation = US\$150  | 97                         | 3    | 90.46                    |
| Study duration=3 days, Prompt frequency=9 ×/day, Prompt length=5 min, Compensation = US\$50   | 88                         | 13   | 82.31                    |
| Study duration=3 days, Prompt frequency=9 ×/day, Prompt length=5 min, Compensation = US\$150  | 89                         | 11   | 84.22                    |
| Study duration=21 days, Prompt frequency=3 ×/day, Prompt length=1 min, Compensation = US\$50  | 95                         | 5    | 87.80                    |
| Study duration=21 days, Prompt frequency=3 ×/day, Prompt length=1 min, Compensation = US\$150 | 95                         | 5    | 89.97                    |
| Study duration=21 days, Prompt frequency=3 ×/day, Prompt length=5 mins, Compensation = US\$50 | 81                         | 19   | 75.03                    |
| Study duration=21 days, Prompt frequency=3 ×/day, Prompt length=5 min, Compensation = US\$150 | 94                         | 6    | 88.67                    |
| Study duration=21 days, Prompt frequency=9 ×/day, Prompt length=1 min, Compensation = US\$50  | 78                         | 22   | 72.83                    |
| Study duration=21 days, Prompt frequency=9 ×/day, Prompt length=1 min, Compensation = US\$150 | 97                         | 3    | 91.51                    |
| Study duration=21 days, Prompt frequency=9 ×/day, Prompt length=5 min, Compensation = US\$50  | 74                         | 26   | 62.92                    |
| Study duration=21 days, Prompt frequency=9 ×/day, Prompt length=5 min, Compensation = US\$150 | 75                         | 25   | 68.11                    |

Percentages may not sum to 100% due to rounding.

continuous participation likelihood ( $r_{pb}=0.81$ ), indicating that those who were unwilling to participate had a lower participation likelihood ( $M=25.59$ ,  $SD=20.59$ ) and those willing to participate had higher participation likelihood ( $M=90.86$ ,  $SD=13.80$ ). Participation responses for each of the sixteen conditions are available in [table 2](#).

All main effects of study design on willingness to participate were significant in the expected directions (see [figure 1](#)). Shorter study duration was associated with greater willingness to participate ( $RR=1.07$ ,  $SE=0.03$ ,  $p=0.016$ , CI 1.01 to 1.13). Of those in shorter study duration condition, 92% selected YES they were willing to participate, whereas in the longer study duration condition 86% selected YES they were willing to participate. Lower prompt frequency was associated with greater willingness to participate ( $RR=1.09$ ,  $SE=0.03$ ,  $p=0.004$ , CI 1.03 to 1.15). Of those in lower prompt frequency

condition, 93% selected YES they were willing to participate, whereas in the higher prompt frequency condition 86% selected YES they were willing to participate. Shorter prompt length was associated with greater willingness to participate ( $RR=1.07$ ,  $SE=0.03$ ,  $p=0.013$ , CI 1.02 to 1.13). Of those in shorter prompt length condition, 92% selected YES they were willing to participate, whereas in the longer prompt length condition 86% selected YES they were willing to participate. Higher compensation was associated with greater willingness to participate ( $RR=1.08$ ,  $SE=0.03$ ,  $p=0.005$ , CI 1.02 to 1.15). Of those in higher compensation condition, 93% selected YES they were willing to participate, whereas in the lower compensation condition 86% selected YES they were willing to participate. In the model overall, 11% of the variance in willingness to participate was explained by the four study design features main effects.

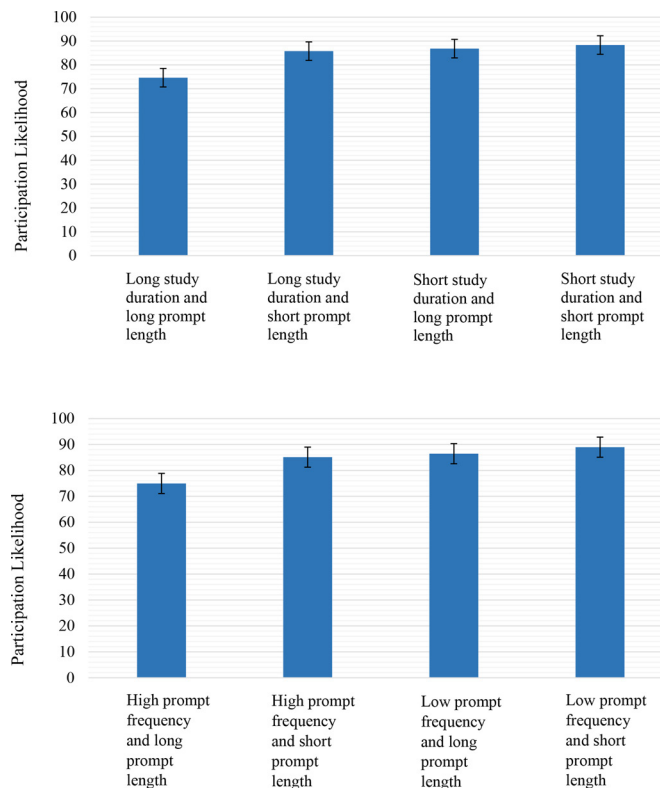


**Figure 1** Participation by study design factors.

All main effects for participation likelihood were in expected directions (see [figure 1](#)), demonstrating modest but statistically significant effects. Study duration ( $b=-7.82$ ,  $SE=1.97$ ,  $p<0.001$ ,  $\eta_p^2=0.026$ ), prompt frequency ( $b=-8.07$ ,  $SE=1.97$ ,  $p<0.001$ ,  $\eta_p^2=0.028$ ), prompt length ( $b=-6.78$ ,  $SE=1.97$ ,  $p=0.001$ ,  $\eta_p^2=0.020$ ), and compensation ( $b=6.88$ ,  $SE=1.97$ ,  $p=0.001$ ,  $\eta_p^2=0.021$ ) each significantly predicted participation likelihood such that those in the less burdensome and higher compensation conditions rated their participation likelihood higher than those in the more burdensome and lower compensation conditions.

#### Aim 2A: interactions between study design features

On an exploratory basis, we examined all possible two-way interactions between study features for their effects on participation likelihood (see [figure 2](#)). There were two small but statistically significant two-way interactions: study duration interacted with prompt length ( $b=-9.63$ ,  $SE=3.89$ ,  $p=0.014$ ,  $\eta_p^2=0.010$ ) such that those in both the longer study duration and longer prompt length conditions reported the lowest estimated participation likelihood compared with those in all other conditions. Prompt frequency similarly interacted with prompt length ( $b=-7.66$ ,  $SE=3.88$ ,  $p=0.049$ ,  $\eta_p^2=0.001$ ), such that those in both the more frequent prompts and longer prompt conditions reported the lowest estimated participation likelihood compared with those in all other conditions. The overall model including main effects and 2-way interactions, only very minimally increased variance explained



**Figure 2** Two-way study design interactions predicting participation likelihood.

(to slightly more than 11%) over the model with only main effects.

#### Aim 2B: main effects of study design features on participation appraisals

Results for associations between study design features and participation appraisals are shown in [table 3](#). When examining how interesting participating in the study would be, only prompt frequency was predictive of participants' ratings. Those in the fewer prompts condition expected participation to be more interesting ( $M=7.23$ ) than those with more frequent prompts condition ( $M=6.67$ ). When examining how enjoyable participating in the study would be, study duration and prompt frequency were predictive of participants' ratings. Those in the shorter study condition rated expected participation to be more enjoyable ( $M=6.08$ ) than those in the longer study condition ( $M=5.54$ ), and those in the fewer prompts condition rated expected participation to be more enjoyable ( $M=6.25$ ) than those in the more frequent prompts condition ( $M=5.36$ ). When examining how effortful participating in the study would be, study duration, prompt frequency, and prompt length were all predictive of participants' appraisals. Those in the shorter study condition rated the study as requiring less effort ( $M=5.50$ ) than those in the longer study condition ( $M=6.70$ ). Those in the fewer prompts condition rated the study as requiring less effort ( $M=5.80$ ) than those in the more frequent prompts condition ( $M=6.40$ ). Those in the shorter prompts condition rated the study as requiring less effort ( $M=5.60$ ) than

**Table 3** Main effects of study design on interest, enjoyment, study effort, and contribution to science

|                  | Interest |           |            | Enjoyment               |           |            |
|------------------|----------|-----------|------------|-------------------------|-----------|------------|
|                  | <i>b</i> | <i>SE</i> | $\eta_p^2$ | <i>b</i>                | <i>SE</i> | $\eta_p^2$ |
| Intercept        | 7.36***  | 0.25      | –          | 6.45***                 | 0.26      | –          |
| Study duration   | –0.31    | 0.22      | 0.003      | –0.54*                  | 0.23      | 0.010      |
| Prompt frequency | –0.56*   | 0.22      | 0.011      | –0.89***                | 0.23      | 0.025      |
| Prompt length    | –0.39    | 0.22      | 0.005      | –0.14                   | 0.23      | 0.001      |
| Compensation     | 0.43     | 0.22      | 0.006      | 0.28                    | 0.23      | 0.003      |
|                  | Effort   |           |            | Contribution to science |           |            |
|                  | <i>b</i> | <i>SE</i> | $\eta_p^2$ | <i>b</i>                | <i>SE</i> | $\eta_p^2$ |
| Intercept        | 4.73***  | 0.26      | –          | 7.32                    | 0.20      | –          |
| Study duration   | 1.19***  | 0.23      | 0.044      | 0.24                    | 0.18      | 0.003      |
| Prompt frequency | 0.60**   | 0.23      | 0.011      | –0.30                   | 0.18      | 0.005      |
| Prompt length    | 1.01***  | 0.23      | 0.031      | –0.07                   | 0.18      | <0.001     |
| Compensation     | –0.16    | 0.23      | 0.001      | 0.10                    | 0.18      | 0.012      |

\* $P \leq 0.05$ ; \*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ .

those in the longer prompts condition ( $M=6.60$ ). None of the study design features were significantly associated with ratings about the extent to which participation would contribute to science. Finally, all participation appraisals were positively correlated with one another (table 4). Bivariate correlations between participation appraisals and participation likelihood are shown in table 4.

#### Aim 2C: main effects of participation appraisals on intentions to participate

Next, we examined whether participation appraisals predicted willingness to participate and participation likelihood (see table 5). Given the high correlations between interest and enjoyment, we also combined these to account for potential multicollinearity. Results of the combined measure were broadly consistent with the results presented in table 5. There was a significant positive association between appraisals of interest and enjoyment on willingness to participate; those who indicated participation would be more interesting or enjoyable were more likely to indicate they would participate. There was a significant negative association between effort and willingness to participate; those who indicated participation would be more effortful were less likely to indicate they would participate. Finally, those who indicated the study would be more interesting, enjoyable, and make a

greater contribution to science reported greater participation likelihood; those who indicated participation would require more effort reported lower participation likelihood.

#### Aim 2D: interactions between study design features and participation appraisals

Finally, we explored all sixteen two-way interactions between each study design parameter (ie, study duration, prompt frequency, prompt length, and compensation) with each participation expectation (ie, study interest, enjoyment, effort and contribution to science) in regressions predicting participation likelihood. Of the 16 possible interactions, 8 interactions were significant, and all showed a similar pattern of association (simple slopes for significant interactions are shown in table 6; see figure 3). Results suggested that study duration interacted with interest ( $b=2.94$ ,  $SE=0.63$ ,  $p<0.0001$ ,  $\eta_p^2=0.036$ ), enjoyment ( $b=2.49$ ,  $SE=0.62$ ,  $p<0.0001$ ,  $\eta_p^2=0.027$ ), and contribution to science ( $b=4.32$ ,  $SE=0.85$ ,  $p<0.0001$ ,  $\eta_p^2=0.042$ ). More favourable participation appraisals were linked with higher participation likelihood and this association was stronger in the longer study duration condition. Prompt frequency interacted with both interest ( $b=2.35$ ,  $SE=0.63$ ,  $p=0.0002$ ,  $\eta_p^2=0.023$ ) and enjoyment ( $b=3.05$ ,  $SE=0.62$ ,  $p<0.0001$ ,  $\eta_p^2=0.040$ ). More favourable

**Table 4** Correlation table for interest, enjoyment, study effort, and contribution to science

|                          | Interest | Enjoyable | Effort   | Contribution to science |
|--------------------------|----------|-----------|----------|-------------------------|
| Enjoyable                | 0.82***  | –         | –        | –                       |
| Effort                   | 0.14***  | 0.16***   | –        | –                       |
| Contribution to science  | 0.61***  | 0.59***   | 0.20***  | –                       |
| Participation likelihood | 0.48***  | 0.46***   | –0.15*** | 0.37***                 |

\*\*\* $P \leq 0.001$ .



**Table 5** Main effects of participation appraisals on willingness to participate and participation likelihood

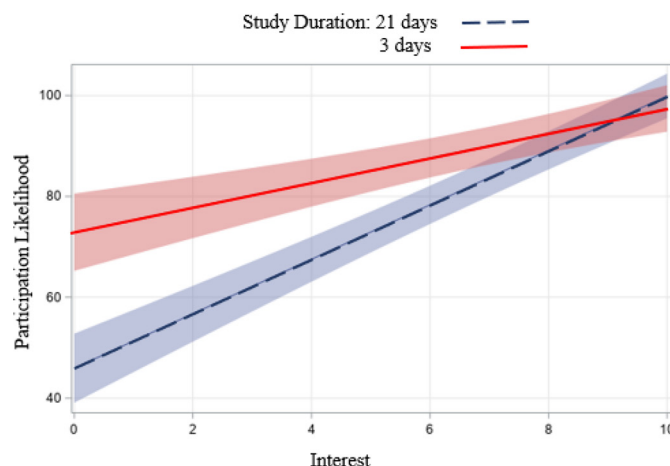
| Willingness to participate |          |       |              |
|----------------------------|----------|-------|--------------|
|                            | RR       | SE    | CI           |
| Interest                   | 1.03***  | 0.01  | 1.01 to 1.05 |
| Enjoyment                  | 1.02*    | 0.01  | 1.00 to 1.03 |
| Effort                     | 0.98***  | <0.01 | 0.97 to 0.99 |
| Contribution to science    | 1.02     | 0.01  | 1.00 to 1.04 |
| Participation likelihood   |          |       |              |
|                            | b        | SE    | $\eta_p^2$   |
| Intercept                  | 57.03*** | 3.35  | –            |
| Interest                   | 2.39***  | 0.57  | 0.029        |
| Enjoyment                  | 1.87***  | 0.54  | 0.020        |
| Effort                     | –2.08*** | 0.30  | 0.077        |
| Contribution to science    | 1.65***  | 0.52  | 0.017        |

\* $P \leq 0.05$ ; \*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ . Enjoyment lower bound is slightly above 1.00 but rounded for display.

participation appraisals were linked with higher participation likelihood and this association was stronger in the more frequent prompt condition. Prompt length interacted with interest ( $b=1.81$ ,  $SE=0.64$ ,  $p=0.0049$ ,  $\eta_p^2=0.014$ ), enjoyment ( $b=2.24$ ,  $SE=0.62$ ,  $p=0.0003$ ,  $\eta_p^2=0.022$ ), and contribution to science ( $b=2.43$ ,  $SE=0.87$ ,  $p=0.0052$ ,  $\eta_p^2=0.013$ ). More favourable participation appraisals were linked with higher participation likelihood and this association was stronger in the longer prompt length condition. There were no significant interactions with study compensation or with appraisals of effort.

## DISCUSSION

The overarching purpose of the present study was to experimentally manipulate hypothetical study design

**Figure 3** Representative simple slopes of two-way interactions between study design features and participation appraisals predicting estimated participation likelihood.

features of an EMA study to examine if study duration, prompt frequency, prompt length and compensation influenced willingness to participate and participation likelihood using design features common among EMA studies. We used two different participation outcomes: willingness to participate, indicating participants' initial yes/no response to the study, and participation likelihood, where participants judged the extent to which they thought they would actually participate (potentially reflecting any uncertainty in their initial yes/no decision). This was meant to emulate the early processes of participant recruitment into patient-reported outcome studies (eg, based on outreach regarding a potential study). Broadly reflecting the view that perceptions regarding participation burden influences willingness to participate,<sup>6 29</sup> our hypotheses were that shorter study duration, fewer prompts per day, shorter prompts (in terms of minutes to complete), and higher compensation would be associated with more willingness to

**Table 6** Simple slopes for significant interactions between study design features and participation appraisals predicting estimated participation likelihood

|                  | Interest |      | Enjoyment |        | Contribution |      |
|------------------|----------|------|-----------|--------|--------------|------|
|                  | b        | SE   | b         | SE     | b            | SE   |
| Study duration   |          |      |           |        |              |      |
| 3 days           | 2.45***  | 0.48 | 2.51***   | (0.46) | 1.96***      | 0.64 |
| 21 days          | 5.40***  | 0.41 | 5.00***   | (0.42) | 6.28***      | 0.56 |
| Prompt frequency |          |      |           |        |              |      |
| 3 × day          | 2.88***  | 0.46 | 2.35***   | 0.44   | –            | –    |
| 9 × day          | 5.23***  | 0.44 | 5.41***   | 0.44   | –            | –    |
| Prompt length    |          |      |           |        |              |      |
| 1 min per prompt | 3.11***  | 0.48 | 2.71***   | 0.45   | 3.02***      | 0.65 |
| 5 min per prompt | 4.92***  | 0.43 | 4.95***   | 0.43   | 5.45***      | 0.57 |

\*\*\* $P \leq 0.001$ . There were no significant interactions for compensation or study effort; we did not examine simple slopes for non-significant interactions.



participate and higher participation likelihood. These hypotheses were all supported. Moreover, interactions among study design factors indicated the importance of prompt length: longer prompt lengths coupled with either longer study duration or more frequent prompts were especially likely to have lower ratings of participation likelihood. Together, results from the present study provide proof of concept; study design features have long been assumed to be linked with intentions to participate. These results are consistent with this premise, indicating that study design features likely influence participants' motivation and stated willingness to enrol in EMA studies.

In the present study, most participants (89%) indicated that they were willing to participate in the hypothetical EMA studies. Actual study enrolment, however, is likely to be much lower. For example, other studies have found that initial reported willingness (as we used in this study) to participate in EMA studies was very high (eg, 66%–83% of invited participants provided indication of willingness to participate), but actual study enrolment was substantially lower (ranging from 16% to 45%).<sup>30</sup> The results presented here suggest that study design features have a modest but significant influence on *intentions* to participate, but this design does not yet provide evidence as to whether, and to what degree, these influence actual study enrolment. That said, consistent with evidence that participant burden and participant reward (eg, compensation) influence self-selection into studies,<sup>31 32</sup> study compliance,<sup>33–35</sup> and study attrition,<sup>36</sup> it seems plausible that study design features should exhibit effects on actual study enrolment. Based on these findings, it will be useful for future research to examine the effects of EMA study design on both intentions and actual study enrolment.

Although the effect sizes for the influence on intended participation for any one study design feature individually were generally of modest magnitude, these can have meaningful cumulative impact on likelihood and intentions to participate in an EMA study, as can be seen in [table 2](#). For example, of those in the most 'desirable' condition (lowest study duration, prompt frequency, and prompt length with the highest compensation), 97% indicated they were willing to participate and ratings of participation likelihood were ~93%. In contrast, of those in the least desirable condition (highest study duration, prompt frequency, and prompt length with the lowest compensation), only 74% indicated they were willing to participate and ratings of participation likelihood were ~63%. Indeed, these findings highlight the importance of carefully considering study design features as participants seem to be substantially less likely to participate in very burdensome designs. Clearly it will be essential to design studies that can adequately test (eg, with sufficient statistical power) the 3- and 4-way interactive effects of multiple study design features (although we note that such studies would require several thousand participants to do so).

## Participation appraisals

It was expected that study design features requiring less burden and higher reward would be associated with increased interest and enjoyment and perceived as less effortful. Our results generally supported these hypotheses. In these analyses, prompt frequency seemed to play an important role: Those in the less frequent prompt condition indicated that they thought the study would be more interesting, enjoyable, and require less effort. Shorter prompt lengths, however, elicited weaker associations between participation appraisals and participation likelihood; it may be the case that participants believe that shorter prompt lengths imply cruder assessments that are less helpful towards achieving study goals.

The results of the present study generally indicated that more intensive and, presumably, burdensome studies (eg, longer study duration, more frequent prompts) may be interpreted as being less interesting and enjoyable, and requiring more effort—possibly limiting participant interest to self-select into such EMA trials. We also found that participant appraisals exhibited largely consistent associations with intentions to participate, suggesting that how participants expect to feel during the study, and the potential benefits they see from the study (eg, contribution to science) can play a role in whether they decide to participate. Future research is needed to further explore the influence of appraisals on self-selection into EMA trials and to extend this to other factors not explored in the present study (eg, motivation to enrol in studies of personal import). This type of work is necessary to determine the extent to which individuals who agree to participate in these studies differ from those who do not agree to participate. For instance, those who participate may have relatively high levels of motivation, interest, and a sense of being able to complete the required reporting tasks. This may skew EMA participant samples toward individuals who find meaning in their participation (eg, patients hoping the research will help others with the same illness), who are more familiar with electronic devices, or those with certain personality characteristics (eg, high conscientiousness, openness to experience). Moreover, there may be important interactions between individual factors that influence participation in EMA studies (such as motivation or personality) and study design features (eg, those with less motivation may be especially unlikely to participate in studies with burdensome designs or more strongly influenced by greater compensation).

## Interactions between study design features and participation appraisals

Interestingly, there were several significant interactions between study design features and participation appraisals in consistent directions. These suggested that participation appraisals played a stronger role in participation likelihood when study design features were more intensive or burdensome (eg, longer study duration, more frequent prompts, longer prompts). Although speculative, this could suggest that studies that are more

interesting, enjoyable, or are thought to provide an important contribution to science act as a sort of buffer against more burdensome study design features. Alternatively, one could view this as suggesting that more intensive studies need to highlight the potential benefits to offset the burden of participating (eg, during recruitment or study training procedures, highlighting to participants the significance of their contributions to science and improving health research). One recent study,<sup>6</sup> for example, found that simply providing participants with a highly descriptive study overview session before participation significantly changed perceived willingness to enrol in studies requiring the use of mobile devices for ambulatory data capture.

It is also potentially important to note that two variables, compensation and expected effort to participate, did not interact with any other variables; in both cases, the main effects of these variables on participation likelihood were consistent. This may suggest that compensation and expected effort are particularly important and/or have a uniform role in influencing intentions to participate, highlighting that researchers should pay careful consideration to these when designing their studies (eg, providing maximal compensation without introducing coercion; providing participants with 24/7 technical assistance to ease the burden of interacting with mobile technologies).<sup>13</sup>

### Limitations

There are several limitations to the present study. First, although MTurk has been used widely in social and behavioural science in recent years due to its capacity for efficiently recruiting and compensating large samples of research subjects,<sup>37</sup> the generalisability of data from MTurk convenience samples compared with population-based samples is still unclear.<sup>38 39</sup> The MTurk sample here was research savvy (given their 'Master Qualified' status) and perhaps more likely inclined to participate in EMA (or any) studies. Thus, the results in the present study may not generalise to 'research naïve' samples, and perhaps represent a conservative estimate of the influence of design features on intentions to participate. Second, all study design features examined in the present study were extreme in nature (ie, 3 vs 21 days duration) and do not provide fine-grained information regarding if there are 'dosage' effects for design features or thresholds at which a design feature becomes too burdensome (eg, 1–9 EMA prompts per day is acceptable but 10+ is too burdensome). As a result, we cannot yet make specific suggestions as to what composition of study design features future EMA studies should utilise to maximise the likelihood of participation. Rather, our results suggest that study design features are an important aspect of research participation that EMA researchers should carefully consider. Third, the present study was not powered to examine higher order interactions. Future research with larger samples are needed to explore how various combinations of design features influence participation.

Fourth, the present study focused on behavioural intentions to participate in hypothetical EMA studies and additional factors thought to influence behavioural intentions (eg, interest, effort). Yet, behavioural intentions are not always strongly related to behavioural engagement<sup>40</sup> and initial willingness to participate in EMA studies seems to be much higher than actual study enrolment.<sup>30</sup> Thus, studies that measure enrolment based on EMA design features are needed to determine which, and for whom, study design features matter most. Finally, given the exploratory nature of this work, we conducted multiple statistical tests without correction (although the consistency in the patterning and effect sizes of results reduces concern about potential type I error).

### CONCLUSION

It is a widely accepted notion that study design features influence who decides to participate in EMA studies, yet there is limited empirical evidence exploring this presupposition. The purpose of the present study was to examine whether study design features (study duration, prompt frequency, prompt length, and compensation) influenced intentions to participate in a hypothetical EMA study. Consistent with theories that higher participant burden should be linked with lower study uptake, studies with longer study duration, more frequent prompts, longer prompts, and lower compensation produced less reported willingness to participate and lower participation likelihood. Results of the present study provide a proof of concept that EMA design features influence behavioural intentions to participate in EMA studies and provide initial evidence of the impact of a subset of specific design features.

**Contributors** JMS contributed to conceptualisation, methodology, investigation, resources, writing – original draft preparation, supervision, project administration, and funding acquisition. DRJ contributed to formal analysis, data curation, visualisation, and writing – original draft preparation. CKFW contributed to conceptualisation, methodology, writing – reviewing and editing. FTM contributed to investigation, resources, and writing – reviewing and editing. SS contributed to conceptualisation, methodology, investigation, data curation, writing – review and editing, and funding acquisition. AS contributed to conceptualisation, methodology, investigation, resources, writing – reviewing and editing, project administration, and funding acquisition.

**Funding** This work was supported by the National Institute on Aging, grant number R37-AG057685.

**Competing interests** AAS is a Senior Scientist with the Gallup Organization and a consultant with Adelphi Values, Inc. Other authors have no competing interests to declare.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not required.

**Ethics approval** Ethical approval was obtained by The Pennsylvania State University Human Research Ethics Committee (2020; approval #00014248).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Data and statistical code are available upon request from the corresponding author for reasonable, non-commercial use.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Joshua M Smyth <http://orcid.org/0000-0002-0904-5390>

Dusti R Jones <http://orcid.org/0000-0002-9060-7046>

Cheng K F Wen <http://orcid.org/0000-0002-2538-0439>

Frank T Materia <http://orcid.org/0000-0002-8015-1743>

Stefan Schneider <http://orcid.org/0000-0002-4562-0524>

Arthur Stone <http://orcid.org/0000-0001-7490-1758>

## REFERENCES

- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1–32.
- Smyth JM, Stone AA. Ecological Momentary assessment research in behavioral medicine. *J Happiness Stud* 2003;4:35–52.
- Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. *J Abnorm Psychol* 2020;129:56–63.
- Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br J Health Psychol* 2010;15:1–39.
- Adell E, Várhelyi A, Nilsson L. The definition of acceptance and acceptability. In: . Farnham, Surrey, England: Ashgate, 2014: 11–21.
- Downs DS, Smyth JM, Heron KE, et al. Beliefs about using Smartphones for health behavior change: an elicitation study with overweight and obese rural women. *J Technol Behav Sci* 2019;4:33–41.
- Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Serv Res* 2017;17:88.
- Berry N, Lobban F, Emsley R, et al. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. *J Med Internet Res* 2016;18:e121.
- Lingler JH, Schmidt KL, Gentry AL, et al. A new measure of research participant burden: brief report. *J Empir Res Hum Res Ethics* 2014;9:46–9.
- Sharp LM, Frankel J. Respondent burden: a test of some common assumptions. *Public Opin Q* 1983;47:36–53.
- Cox K, McGarry J. Why patients don't take part in cancer clinical trials: an overview of the literature. *Eur J Cancer Care* 2003;12:114–22.
- Groves RM, Cialdini RB, Couper MP. Understanding the decision to participate in a survey. *Public Opin Q* 1992;56:475.
- Wenz A, Jäckle A, Couper MP. Willingness to use mobile technologies for data collection in a probability household panel. *Surv Res Methods* 2019;13.
- Jones A, Remmerswaal D, Verveer I, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* 2019;114:609–19.
- Napa Scollon C, Prieto C-K, Diener E. Experience sampling: promises and pitfalls, strength and weaknesses. In: *Assessing well-being: the collected works of ED Diener*. The Netherlands: Springer, 2009: 39. 157–80.
- Pratap A, Allred R, Duffy J, et al. Contemporary views of research participant willingness to participate and share digital data in biomedical research. *JAMA Netw Open* 2019;2:e1915717.
- Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018;52:446–62.
- Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. *Australas Phys Eng Sci Med* 2015;38:23–38.
- Steele R, Lo A, Secombe C, et al. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. *Int J Med Inform* 2009;78:788–801.
- Zhao J, Freeman B, Li M. Can mobile phone Apps influence people's health behavior change? an evidence review. *J Med Internet Res* 2016;18:e287.
- Chow CK, Redfern J, Hillis GS, et al. Effect of Lifestyle-Focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. *JAMA* 2015;314:1255–63.
- Chau JP-C, Lee DT-F, Yu DS-F, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. *Int J Med Inform* 2012;81:674–82.
- Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. *Clin Psychol Psychother* 2014;21:427–36.
- Couper MP, Singer E, Conrad FG, et al. Experimental studies of disclosure risk, disclosure harm, topic sensitivity, and survey participation. *J Off Stat* 2010;26:287–300.
- Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process* 1991;50:179–211.
- Chen W, Qian L, Shi J, et al. Comparing performance between log-binomial and robust poisson regression models for estimating risk ratios under model misspecification. *BMC Med Res Methodol* 2018;18:63.
- Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702–6.
- McNutt L-A, Wu C, Xue X, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;157:940–3.
- Matthew-Maich N, Harris L, Ploeg J, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. *JMIR Mhealth Uhealth* 2016;4:e29.
- Dornsife U. Real-Time sample composition statistics. understanding America study n.d. Available: <https://uasdata.usc.edu/index.php> [Accessed 31 Dec 2020].
- Frankel J. Measurement of Respondent burden: study design and early findings. In: *Applied statistics*. Washington, 1980.
- Abeler J, Nosenzo D. Self-Selection into laboratory experiments: pro-social motives versus monetary incentives. *Exp Econ* 2015;18:195–214.
- Merrill PD, McClure LA. Dichotomizing partial compliance and increased participant burden in factorial designs: the performance of four noncompliance methods. *Trials* 2015;16:523.
- Eisele G, Vachon H, Lafit G, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment* 2020;107319112095710.
- Rintala A, Wampers M, Myin-Germeyns I, et al. Response compliance and predictors thereof in studies using the experience sampling method. *Psychol Assess* 2019;31:226–35.
- Pirotta S, Joham A, Hochberg L, et al. Strategies to reduce attrition in weight loss interventions: a systematic review and meta-analysis. *Obes Rev* 2019;20:1400–12.
- Chan C, Holosko MJ. An overview of the use of mechanical Turk in behavioral sciences: implications for social work. *Res Soc Work Pract* 2015;26:441–8.
- Mullinix KJ, Leeper TJ, Druckman JN, et al. The generalizability of survey experiments. *J Exp Polit Sci* 2015;2:109–38.
- Newman A, Bavik YL, Mount M. Data collection via online platforms: challenges and recommendations for future research. *Applied Psychology* 2020.
- Ajzen I, Brown TC, Carvajal F. Explaining the discrepancy between intentions and actions: the case of hypothetical bias in contingent valuation. *Pers Soc Psychol Bull* 2004;30:1108–21.