



A population-based investigation of participation rate and self-selection bias in momentary data capture and survey studies

Arthur A. Stone^{1,2} · Stefan Schneider^{1,2} · Joshua M. Smyth³ · Doerte U. Junghaenel^{1,2} · Mick P. Couper⁴ · Cheng Wen² · Marilyn Mendez² · Sarah Velasco² · Sarah Goldstein²

Accepted: 16 February 2023 / Published online: 6 March 2023
© The Author(s) 2023

Abstract

Participant selection bias is of concern to researchers conducting surveys of all types. For momentary data capture studies, such as Ecological Momentary Assessment, the level of burden associated with these techniques and the possibility of low uptake rates makes the concerns especially salient. This study invited 3,000 individuals to participate in a study of health and mood and recorded the uptake rates at various points in the process. Respondents expressing interest in participating in general were randomized into a one-time survey, a low-burden momentary study, or a high-burden momentary study. Overall, 85.9% of the sample did not respond to the study invitation (including confirming non-interest); 6.9% of the sample expressed interest in the study by completing a brief survey; 2.1% agreed to participate in the study when the protocol specifics were described (none of the study protocols were actually run). Whites were more likely to complete the survey. Of those completing the survey, individuals who reported higher income, a more “open” personality, better typing skills, better computer skills, who viewed the research topic as important, and who expressed interest in research on daily feelings more likely consented to being enrolled in the experiment. The number of prior surveys taken had an inverted-U shaped association with participation in this study. Finally, all individuals randomized to the one-time survey group agreed to participate compared to two-thirds of individuals in the momentary groups. These results suggest that participant selection bias may affect both one-time survey and momentary data capture studies, with the caveat that the degree of such bias will be related to a study’s hypotheses.

Keywords EMA · Survey · Selection bias · Participation rate · General population

This paper concerns self-selection bias in survey research generally and particularly with studies using momentary assessment strategies, which are often referred to as Ecological Momentary Assessment (EMA, the term that will be used throughout the paper (Shiffman et al., 2008; Stone & Shiffman, 1994), the Experience Sampling Method (ESM) (Conner et al., 2009; Csikszentmihalyi & Hunter, 2003),

and Ambulatory Assessment (Ebner-Priemer & Trull, 2009; Wright & Zimmermann, 2019). Momentary studies are often based on a relatively small number of study participants and there may be concerns about making inferences from these studies to broader populations. As the EMA field matures, it is important to examine the feasibility of recruiting individuals from the general population – the focus of this paper – or from more specialized populations (e.g., patients) given our perception that many momentary researchers believe that only a small proportion of individuals approached for a momentary study actually participate (Stone et al., *in press*). The specific concern that this paper evaluates is that people with certain characteristics will not agree to participate; if so, this creates a threat to the external validity (Cook & Campbell, 1979) of the results.

The methodological phenomenon described above is long recognized and usually referred to as self-selection bias. It has received considerable attention in the social sciences and

✉ Arthur A. Stone
arthur.stone@usc.edu

¹ Department of Psychology, University of Southern California, Los Angeles, CA, USA

² Dornsife Center for Self-Report Science, University of Southern California, Los Angeles, CA, USA

³ Department of Biobehavioral Health, The Pennsylvania State University, State College, PA, USA

⁴ Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

the survey research literature (Bethlehem, 2010; Heckman, 2010). A number of studies have focused on motivational barriers of uptake into studies (in general, not focused on EMA) by examining the relevance of study content for prospective participants (Materia & Smyth, 2021), the availability of data collection devices (Jäckle et al., 2019), and by examining appropriate monetary and non-monetary incentives (Yu & Cooper, 1983).

There has been less attention to self-selection bias in the field of momentary data capture and virtually no research in general populations on the topic (Gabriel et al., 2019; Hektner et al., 2007; Scollon et al., 2009). A recent review of “pressing” issues for the field of EMA has identified participant self-selection as a major concern (Stone et al., 2023) and suggested possible ways of exploring the topic. The relevance of self-selection bias, defined as individuals deciding for themselves whether or not to participate in a study, lies in the possibility that the sample will not adequately represent the population from which it was drawn. That is, that those declining participation in an EMA study will be different in some ways from those agreeing to participate. Under the assumption of sound sampling strategies and excellent uptake rates, studies achieving high uptake reduce the threat of self-selection bias, whereas the threat likely increases with lower uptake rates. However, the definition of what constitutes “good” uptake rates is not a fixed value: it depends upon the associations being studied and how selection impacts the relevant variables. It is also the case that less than perfect uptake rates do not *necessarily* result in self-selection processes that will bias the external validity of the results, because such bias occurs when the characteristics of those not participating (measured or not measured) impact the associations under consideration.

For example, imagine an EMA study concerned with understanding whether momentary pain is associated with a greater likelihood of momentary social withdrawal using data from random prompts. Further, let’s assume that this relationship is moderated by trait extroversion such that pain relates to social withdrawal among introverts much more than among extroverts *and* that extroverts are more likely to participate in an EMA study. A study of this topic would be therefore likely to under-estimate (bias) the observed effect of momentary pain on social withdrawal, because extroverts were more likely to participate *and* because the personality characteristic is associated with the effect being studied. However, if the goal of the study was to examine the relationship between momentary pain and medication taking and this association was not related to extroversion, then the results would not be biased. Thus, these points need to be considered in the evaluation of self-selection bias and its impact on a study’s external validity.

We start by describing a typical momentary assessment protocol to provide the reader with a sense of what might

be asked of participants. EMA studies typically signal individuals several times a day over many days to complete brief surveys as they go about their daily routines¹. Unlike more traditional single-survey methods, EMA is usually viewed as relatively burdensome procedure given the somewhat intrusive nature of the assessments and the level of participant involvement required for participation or uptake into the study (the terms “uptake” and “uptake rates” are used throughout the paper). Of course, very long surveys may also be perceived as burdensome. Burdensome protocols are often less attractive to many people and may result in lower uptake rates, which in turn could lead to self-selection bias (Scollon et al., 2009). However, studies of highly motivated individuals, such as those where the collected data may influence their treatment, may yield higher uptake rates in spite of the burden. Furthermore, the requirements of answering EMA surveys on smartphones or other electronic devices may reduce the uptake of those with low technological competence (Keusch et al., 2019) or who do not have access to smartphones. Although we focus on influences for EMA studies, there are, undoubtedly, other factors that can also reduce uptake in standard surveys, such as questionnaire length, online only administration, etc., but these are not discussed herein.

As with most other methods, initial uptake rates in EMA studies are largely unknown, because studies typically rely on samples of convenience such as students, those who respond to flyers or advertisements, and other forms of sampling where the number of individuals “seeing” the invitation is not known. Thus, with some exceptions, such as a consecutive medical admission sample as a sampling frame, the denominator for calculating uptake rates is not known. Beyond not knowing the uptake rate, researchers usually do not have information about the *characteristics* of nonparticipants making it difficult to assess the representativeness of the sample relative to the population. These two conditions, unknown or low uptake rates (Scollon et al., 2009) and lack of information about nonparticipants, leave the question of possible self-selection bias largely unanswered with a strong possibility that such bias does exist.

The motivation for this study was to add to our knowledge of uptake rates and possible selection bias, with a focus on EMA studies.

Question 1. Uptake rates throughout recruitment Because little is known about uptake rates in general population

¹ There are many variants of EMA and some do not require signaling (e.g., when participants complete an assessment based on the occurrence of some event). The number of assessments per day, the length of each assessment, and the number of consecutive recording days vary widely according study goals.

samples, we examined uptake rates by mailing 3,000 individuals selected by a marketing firm an invitation to learn more about a study on health and mood. Uptake rates were able to be computed on the basis of positive replies and knowledge of the number of invitations mailed. We did not specifically hypothesize particular uptake rates for various stages in the recruitment process, but we speculated that overall interest in participating might be in the range of 20–25%. Furthermore, we did not know what the uptake rates would be for EMA. Prior work examined individuals who were members of a panel and were willing to complete surveys (Smyth et al., 2021). They were asked to indicate if they would participate in EMA studies and high, yet somewhat implausible, uptake rates were observed. Thus, to address this question we present uptake rates at multiple stages of the recruitment process.

Question 2. Were those who expressed interest in the study different from those who did not? We hypothesized demographic differences between those who expressed an interest in learning about the study *and* who completed a brief demographic survey (included with the initial, mailed invitation) versus those who did not. This may be considered a relatively “low” bar for uptake, because it does not go to the core issue of actual participation in the study. Nevertheless, it should provide insight into the characteristics of those who *may* decide to participate, and, as mentioned earlier, this information is usually not available for analysis. To examine variables that could be associated with possible self-selection bias, a marketing firm provided information about addressees’ gender, age, and race for everyone in the initial mailing. On the basis of prior work on factors relating to study uptake, we expected participants in the studies to be older and more likely to be white (Couper et al., 2007; Dunn et al., 2004; Kim et al., 2008).

Questions 3a and 3b. Did those who agreed to participate have different characteristics than those who only expressed interest in the study? This question may be thought of as emulating the typical situation for recruiting in many research studies, that is, a group of individuals express interest in the study after viewing or hearing an announcement for it and then only a portion go on to actually participate. Question 3a refers to analyses of this question employing the limited demographic data from the full sample. Question 3b refers to analyses with individuals who completed the demographic survey (as an indication of their interest in learning more about the study); we were able to examine a broad range of possible self-selection related variables with this survey. We had some specific predictions for these analyses. EMA studies often

depend upon participants completing questionnaires on smartphones, thus, we hypothesized that greater familiarity with smartphone and computer technology would be associated with uptake into the studies. Prior work has also shown that a personal interest in the topic or the stated aims of a study can influence uptake (Dillman et al., 2014) and we hypothesized that that would be the case here as well.

Question 4. Does the burden of an EMA protocol affect uptake and how does it compare to a one-time survey? Little is known about uptake rates in EMA protocols with different burden levels; for example, EMA protocols with longer questionnaires, with a higher frequency of daily questionnaires, and/or requiring a larger number of days of participation may be viewed as having greater burden. Therefore, we examined uptake rates for two protocols that differed in burden of completing the protocol: one with few daily prompts for a week and the other with a higher number of prompts for two weeks. We hypothesized that more burden would be associated with lower uptake. In order to provide a context for the EMA uptake rates, we added a third study arm that was a single administration of a one-time survey about health and mood. The survey was loosely matched in burden to the low-burden EMA condition, where the expected duration of study participation and the monetary compensation were the same in both conditions. We hypothesized that the EMA protocols would yield lower uptake rates than the one-time survey condition; however, we did not have a prediction about the magnitude of the difference. The aims of this section were achieved by randomizing individuals who expressed interest in participating in one of the three study arms.

A central feature of this study was the use of a procedure to ascertain “uptake” in EMA and survey studies without actually running the studies that were presented to respondents. The procedure was to follow all of the usual steps to enroll participants in a study, including exposing them to the consent procedure and scheduling appointments for the one-time survey or for the first day of EMA. In the end participants were not required to complete the one-time survey or the EMA assessment protocol, but were instead debriefed at the close of the protocol. Our goal in these procedures was to improve upon prior vignette-type work, including some of our own, that simply asked individuals about whether they would be willing to participate in hypothetical EMA studies (Smyth et al., 2021). That is, we sought to maximize the credibility of the recruitment process, without actually having to run the studies, which would have required substantial resources and collected data not directly pertinent to the goals of this study.

Methods

Invitees

The names and addresses of 3,000 potential participants for the current study were randomly selected by MSG, Inc. (<https://www.m-s-g.com/>). This firm created a file of names and addresses from a subset of the United States Postal Services (USPS) delivery sequence files, where the individuals included in the file are those whose name, gender, age, and ethnicity group were consistent (matched) across two other commercially available datasets (e.g., data obtained through validated external sources, for example, Experian; <https://www.experian.com/>) within the United States. This was our way of obtaining limited demographic information about all individuals who were sent invitations for the study.

Procedures

The study had several components that were designed to accomplish the goals mentioned above. The first part, called the Invitation Stage, allowed for the computation of response rates and analyses of characteristics of those who responded; it was comprised of the initial mailing and responses to the invitations. The second component allowed for the collection of more detailed personal information for subsequent determination of characteristics of those who agreed to be in the final study (the Demographic Survey Stage). It included administration of the demographic and individual differences questionnaire to interested individuals (called the “demographic survey”). The last component included procedures that randomized willing individuals into the three arms of the study, consented and scheduled them into the study, and, finally, provided a debriefing (the Screening/Consent/Debrief Stage). Each of these components is presented below.

Invitation stage Invitations were mailed to the list of 3,000 names and addresses in batches of 500 between June 2021 and September 2021. The invitation materials included a general introductory letter describing our interest in learning about everyday mood and health and described the ways that individuals could complete the demographic survey, with the promise of an additional \$10 Amazon gift card sent upon return of the completed survey. The materials also included a postcard for participants to indicate that they were not interested in participating and included a \$2 bill as an incentive. Individuals who wished to complete the demographic survey could access it online or could contact the research center staff by phone.

To encourage consideration and a response to the invitation, three weeks after the initial mailing, individuals who had not responded in any way were sent a reminder package. The reminder package included a similarly-worded letter with instructions on ways to complete the demographic survey, the same postcard for participants to indicate that they were not interested in participating, and a paper-and-pencil version of the demographic survey. A business reply envelope was also included for participants to return the paper-and-pencil demographic survey.

Demographic survey stage To encourage responding to the survey, invited participants had three ways to respond: online, by phone, or by returning the paper-and-pencil questionnaire (included only in the reminder package). The survey included demographic information (e.g., age, gender, ethnicity, income, education level), current mood and overall health, current electronic device use, prior survey participation (e.g., whether the participant had participated in surveys for other companies/organizations and whether they were paid for participation), Big Five personality questions (Costa & McCrae, *n.d.*), and whether the participant was interested in learning more about the study on mood and health advertised in the invitation letter. If the participant was not interested in continuing with the study, they were asked their reason for not being interested, whether they would like to participate in future studies, and how they would like to receive their compensation (either in the form of a physical gift card or electronically). Those who completed the questionnaire were compensated with a \$10 Amazon gift card, regardless of their interest in the study.

For those who completed the survey online, the survey was accessed by typing in a web link that was included in their invitation letter and that led to an online survey programmed in Qualtrics (www.qualtrics.com). Individuals who expressed interest in learning more about the study at the end of the survey were presented with a calendar where they were able to schedule an appointment to complete the next and the last stage of the study. For those who chose to complete the survey via phone, the same set of questions and response options included in the questionnaire was read to the participants by research staff using a standardized script. At the end of the phone interview, the research staff asked if the individual was interested in learning more about the study; if so, a staff member manually scheduled them in the online scheduling calendar. Those who returned the paper-and-pencil survey by mail were asked to provide their contact information (i.e., a phone number) and were subsequently contacted by a member of the research staff to schedule the next stage of the study.

There were times when a research staff member tried contacting a participant and the participant did not answer

(e.g., when responding to a participant's voice message expressing interest in completing the questionnaire). In this case, the staff member left a brief voice message. For each participant, the research center staff member conducted three follow-up contact attempts before ceasing contact.

For practical reasons, each batch of invitations had a respective close-out date when potential participants could no longer be enrolled in the study. Close-out dates occurred at the end of the 16th week after the initial invitation package was sent. If any questionnaires were received after the deadline and an individual expressed interest in the study, they were informed that the enrollment period for the study was closed, but that they would still be compensated with the \$10 Amazon gift card for completing the questionnaire.

Screening/Consent/Debrief stage Individuals who expressed interest in learning more about the study on health and mood were contacted with a phone call that informed the participant about the study. The phone call was also used to consent the participant (if they were interested in the study that was described to them) and to debrief them about the true purpose of the study (that the primary goal was on their interest in the proposed studies).

After completing the demographic survey, interested respondents were randomized to one of three study conditions. One condition was the low-burden EMA condition, which was described as a 7-day long study where the participant would answer three 1-minute EMA prompts about mood/health per day and would be paid \$40 for participating. The second condition was a high-burden EMA condition, which was described as a 14-day long study where the participant would answer six 1-minute EMA prompts about mood/health per day and would be paid \$160 for their participation. The third condition was a one-time survey condition, which was described as a single 40-minute online survey about mood and health with \$40 for compensation. The compensation rate for each study condition was roughly set as \$1 for each minute of expected (i.e., per description of the study design) participation in each condition. The interview script was the same for the three conditions except for the study description and eligibility sections, which were study specific. There were three parts to this section: screening/consent/debrief call.

Screening The screening introduced the study in a general way, asked about condition-specific eligibility, described the study conditions in more detail, and then asked the participant about their interest in the study. The eligibility criteria for all three study conditions were (1) at least 18 years of age, (2) no hearing impairments, (3) no vision impairments that could not be corrected with contact lenses or glasses, (4) have stable access to an active e-mail account,

(5) have daily access to wireless internet (Wi-Fi), and (6) fluency in English. For those randomized to either EMA study conditions, an additional eligibility condition was imposed (as it would be in most EMA protocols): they were asked whether they work a night shift or the equivalent that required them to be awake at night and asleep during the day.

Consent If the participant was not interested, the study was concluded, and the participant was thanked for their time. If the participant expressed interest in the study, they continued with the phone script and were emailed a consent form so that they could follow along as the staff member reviewed the form. Here the study was described in detail, including the condition-specific study description, risks and benefits of participating, confidentiality, participant rights in the study, and contact information. The staff member asked the participant for verbal consent to participate in the study. A person was considered consented for participation in the study if they successfully completed these steps; that is, they contributed positively to the uptake rate.

Debrief In this final step we notified the participant that the study was complete and disclosed that they would not actually be participating in the study described to them; we explained that the true purpose of the study was to learn about their interest in participating, because this was central for understanding EMA studies. Participants were told that they would be compensated for their participation with a \$20 Amazon gift card and that they could exclude their data from the study if desired (an Institutional Review Board [IRB] requirement); no requests for data exclusion were received. There were checks within the script to ensure that participants understood the study was complete and that they were not negatively affected by the deception. They were also given the option to ask any questions they had and share any concerns. The entire study procedure was approved by the USC IRB (ID: UP-20-00970).

Measures

Participant information

The study obtained information about the age, gender, and ethnicity of all 3,000 invitees from the marketing firm. Additionally, more comprehensive measures of demographic and individual differences were collected in the demographic survey. The variables in the survey are described below. We thought that these variables might distinguish between those who fully participated in the study versus those who did not, and they were broadly based on prior literature.

The demographic survey was comprised of the following sections:

Demographics. Age was coded as a continuous variable, gender was coded as male [1] or female [0], and race as white [1] or non-white [0]. Annual income was coded as greater than or equal to \$75,000 [1] versus less than \$75,000 [0]. Education was coded as bachelor's degree or higher [1] versus less than bachelor's degree [0].

Subjective Well-Being. Subjective well-being was measured by Cantril's ladder (Cantril, 1965), where individuals were asked to rate where on an imaginary ladder with 11 rungs they would personally feel they stand at this time. The bottom represents the "worst possible life for you" (coded 0) and the top rung represents the "best possible life for you" (coded 10); rungs in-between 0 and 10 were coded as 1 (next to the bottom of the ladder) through 9 (next to the top of the ladder).

Memory Problems and Self-Reported Poor Health. Survey respondents were asked to indicate whether or not they had talked to a health care provider about memory problems in the past 12 months (No [0], Yes [1]) and to rate their overall health as Excellent (1), Very Good (2), Good (3), Fair (4), or Poor (5). The variable was coded such that higher numbers represent worse overall health.

Poor Current Mood. Poor current mood was assessed with a single item as Excellent (1), Very Good (2), Good (3), Fair (4), or Poor (5), coded such that higher numbers represent worse mood.

Electronic Device Ownership, Device Skill Levels, and Internet Access. Participants were presented with four categories of electronic devices, including (1) a desktop, laptop, or tablet computer that has an attached keyboard; (2) a tablet device that does not have an attached keyboard; (3) a smartphone; and (4) fitness tracking watch or ring. They were asked to select all the categories of devices they currently own or use. From this question we extracted a single variable pertaining to owning a smartphone given its relevance to EMA, which was dichotomously coded. Two questions asked participants to rate their "computer skill level" and their "smartphone skill level" with the following response options: Beginner (1), Moderate (2), Competent (3), and Expert (4). Two additional questions asked "how confident are you in using a computer for writing tasks that involve typing on the computer keyboard, such as answering an email" and "how confident are you in using your smartphone for writing tasks that involve typing, such as answering email" with the following response options: Not confident at all (1), Somewhat confident (2), Very confident (3), and Completely confident (4).

Prior Survey Participation. Participants were asked whether they take surveys for any companies or organizations (No or Yes response options) and how often (if any) they participated (Never or 0 surveys (1), 1 survey

per year (2), 2–4 surveys per year (3), 5–7 surveys per year (4), and 8 or more surveys per year (5)).

Importance of Research Topic on Study Uptake. Participants were asked to rate "how important it is for you that you are interested in the topic of a study when you decide to participate in a study?" They were asked to choose from Not at all (1), A little bit (2), Somewhat (3), Quite a bit (4), or Very much (5).

Interest in Research Topic. Participants were asked to rate "how interested are you in research that tries to learn about people's experiences and feelings in daily life"? The same response options as for the previous question were provided.

Personality Assessment. The Big Five Personality Inventory (McCrae & Costa Jr., 1999) was used to assess the participants' personality traits. Participants were asked to rate whether they disagree strongly (1), disagree a little (2), neither agree nor disagree (3), agree a little (4), or agree strongly (5) to 44 statements. The responses for 16 statements were reverse-scored and all statements were summed according to the scoring instructions to create five scores representing the level of extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience.

Analysis plan

Question 1 To understand individuals' rates of uptake as they transitioned from the initial mailing to those who consented and participated in the randomized experiment, we present descriptive uptake rates at all stages of the recruitment process. Reasons for non-uptake at each stage are also presented in tabular form.

Question 2 This question was addressed by contrasting individuals who expressed interest in the study (by completing the demographic survey) versus those who did not complete the survey. Comparisons were based on the three variables from the marketing firm (percent male, percent white, and age), which were available for almost the entire sample. Although we did not advance any causal claims given the observational nature of the data, we designated the three demographic variables as predictors in logistic regression models with not participating versus participating as the dichotomous outcome. This is consistent with the idea that the demographics preceded in time any decisions about participating at various stages of the recruitment process. Thus, we present tables with uptake rates as the outcomes, for example, the proportion of uptake for men versus women. For the continuous predictor, age, the overall logistic regression result indicated whether or not there was a significant association between age and uptake. To report the magnitude of the effect, we generated least square estimates of the uptake rates for individuals 25, 50,

and 75 years of age. However, because the response options available for the variable on the frequency of prior survey participation could be considered ordinal rather than linear, we treated it as a categorical variable and presented least square means for each of the five levels.

Questions 3a and 3b We next examined individuals who consented in the last stage of the recruitment process (that is, who were actually going to participate in the protocol offered to them) versus those who did not. Like the preceding analyses, logistic regressions used the demographic variables to predict the outcome representing those who consented and agreed to participate in the study (those who we fully expected to start the study they were assigned [1]) versus those who did not consent [0]). Both MSG and demographic survey variables served as predictors in separate analyses. For Question 3a, the three demographic variables from MSG were used to predict the uptake outcome. An advantage of this approach is that almost the entire sample had the predictor variables. Question 3b addressed a similar question, but it focused only the 202 individuals who completed the demographic survey. Again, logistic regressions were used to predict those who consented versus those who did not. These analyses have the advantage of being able to utilize all of the variables from the demographic survey. A disadvantage of this approach is that the sample size is limited. For the many continuous predictors in the demographic survey, we present least square means for the 25th and for the 75th percentiles of a predictor's distribution (the exception was for age, where we used 25, 50, and 75 years of age instead of the percentiles). Both questions 3a and 3b were, then, focused on comparing those people that ultimately consented to be in one of the three protocols with either most of the people sent invitation letters to or with people who completed the demographic survey.

Question 4 For the comparison of people's willingness to participate in low-burden EMA, high-burden EMA, and one-time survey protocols in the randomized experiment, we first examined whether the three randomized groups differed on any of the variables from the demographic survey.

The primary analysis, though, was to determine if there were group differences (one-time survey, low-burden EMA, high-burden EMA) in uptake rates, and these were tested with Fisher Exact Tests. The first test included all three groups to determine if there were any differences; if so, tests among combinations of groups would be computed to determine the pattern of variation.

Statistical power For Questions 2 and 3a, the minimum effect size Cohen's w (the square root of the standardized chi-square statistic) that was detectable with 80% power ($\alpha = 0.05$) given an anticipated sample size of 3,000

subjects was $w = 0.05$. Values for w of 0.10, 0.30, and 0.50 are considered small, medium, large effect sizes, respectively (Cohen, 1988). The sample size for the remaining research questions 3b and 4 was determined by the number of individuals in each analysis step; for this reason, we report results from sensitivity power analyses for these questions throughout the "Results" section.

Results

Question 1. Overall response to initial mailing

Of the 3,000 invitations mailed, 62 responses were eliminated from consideration for one of three reasons: returned by the post office as undeliverable (52), sent to deceased individuals (9), and in one instance the addressee had moved out of the residence. This yielded a revised total invitation number of 2,938 (see Table 1 for the flowchart of the recruitment results). Of the revised total, 2,525 (85.9%) did not respond to the invitation in any manner. Also, three people expressed interest about the study after the response deadline. A total of 6.9% of the sample indicated interest in learning more about the study by completing the demographic survey. Only 2.8% expressed interest in participating and, ultimately, 2.1% of the sample entered the randomized portion of the study.

Question 2. Were those who expressed interest in the study different from those who did not?

Turning to the second question, we compared the demographic characteristics of the 202 individuals who completed the demographic survey with the rest of the sample that received the invitation, that is, excluding the 52 invitation packages that were undeliverable and the 10 invitation packages that were returned because the invitee was deceased or had moved away ($n = 2,736$), see Table 2. Of the 202 who completed the demographic survey, 121 completed the questionnaire online, 65 returned a printed survey, and 16 were completed by telephone.

Regarding the validity of the MSG data, we compared the marketing firm's information with the responses on the same measures obtained in the demographic survey. One hundred and ninety-eight individuals provided gender information on the demographic survey; there were two disagreements between the sources of information, yielding an agreement rate of 99%. There were more disagreements for race ($n = 196$), where 5 individuals were designated as non-white by the marketing firm, yet said they were white on the survey and 6 people were classified as non-white by the marketing firm, yet said they were white on the survey. This is an overall agreement rate of

Table 1 Flow of participants in the study

Action	N	Percent of initial mailing	Notes for this step
1. Initial mailing	3,000	100.0%	52 packages undeliverable (0.3%) 10 refused by recipient because addressee was deceased or had moved out of home
2. Delivered invitations	2,938	98.0%	2,525 did not respond in any way (85.9%) 204 refused by returning postcard indicating no interest in participating (6.9%) 7 other refusals or not able to contact (0.2%)
3. Completed demographic survey and screened for possible participation	202	6.9%	64 had no interest in being screened and participating (31.7%) 38 were not contactable (18.8%) 6 provided only partial information or we could not contact (3.0%) 7 were no longer contactable (3.4%) 3 could not review the consent form (1.5%)
4. Contacted for treatment	84	2.8%	10 did not meet EMA eligibility criteria (11.9%)
5. Experimental treatment	74	2.5%	Randomized into experimental arms: Arm 1: one-time survey, n=26 Arm 2: low-burden EMA, n=23 Arm 3: high-burden EMA, n=25
6. Final simulated uptake	61	2.1%	Uptake by experimental arm: Arm 1: one-time survey, 26 participated Arm 2: low-burden EMA, 18 participated Arm 3: high-burden EMA, 17 participated

Table 2 Differences on MSG demographic variables for those who completed the demographic survey (n=202) versus those who did not (n=2,736)

Demographic predictor	N	Percent participated	LR χ^2	p-value
White			16.65	<0.001***
No	683	3.7%		
Yes	2,241	7.9%		
Male			1.35	0.246
No	811	7.8%		
Yes	2,125	6.5%		
Age	2,938		0.24	0.624
25 years		7.4%		
50 years		7.0%		
75 years		6.6%		

Note. For age predictions, predicted means for ages 25, 50, and 75 are presented. LR = likelihood ratio.

94%. Finally, the correlation coefficient for age (n = 200) between the two sources of data was 0.964 (p < .001). The mean age for the 200 individuals with the marketing data was 57.2 whereas it was 56.8 with the survey data (p = .165 for difference in means).

The only difference in the variables available to us at this stage of recruitment was a higher proportion of whites completed the demographic survey (7.9%) than non-whites (3.7%). Gender and age were not significant predictors of completion of the demographic survey.

Question 3a. Were individuals who expressed interest in the study different from those who consented and participated (using the MSG variables)?

This question contrasted those who consented and participated in the study (n = 61, 2.1% of 2,938) versus all other individuals (n = 2,877; Table 3). Again, we look to the MSG demographic variables for these comparisons. Similar to the previously reported comparisons, those who participated were more likely to be white (2.45% versus

Table 3 Differences on MSG demographic variables for those who consented and participated versus those who did not

Demographic predictor	<i>N</i>	Percent participated	LR χ^2	<i>p</i> -value
White			7.63	0.006**
No	2,863	0.88%		
Yes	61	2.45%		
Male			3.99	0.046*
No	2,875	2.96%		
Yes	61	1.74%		
Age	2,938		1.43	0.231
25 years		1.48%		
50 years		1.90%		
75 years		2.43%		

Note. For age predictions, predicted means for ages 25, 50, and 75 are presented. LR = likelihood ratio

0.88%). Males were less likely to participate compared to females (1.74% versus 2.96%). There was no statistically significant difference in age by uptake status.

Question 3b. Were those who expressed interest in the study (by completing the demographic survey) different from those who consented and participated in the study (using the demographic survey variables)?

The second way of addressing these differences was by limiting the analyses to the 202 individuals who completed the demographic survey and contrasting those who consented ($n = 61$) versus those who did not ($n = 141$). This strategy had the advantage of allowing us to examine the variables in the demographic survey that were specifically chosen because they might be associated with selection bias. The minimum detectable effect size given this sample size (power = 0.80, $\alpha = 0.05$) was $w = 0.20$, a small to medium effect. Uptake rates are higher in these analyses compared to the previous ones, because the sample is limited to those who expressed interest in the study.

Table 4 presents the prediction of uptake rates from the demographic survey variables. Individuals with incomes greater or equal to \$75,000 were just over 60% more likely to consent (39.6% versus 24.3%) and those with better computer skills and smartphone typing skills were 40% (33.8% versus 24.2%) and 46% (39.5% versus 27.0%) more likely to consent, respectively. Individuals' views on the importance of the research were 47% more likely to consent (36.7% versus 25.1%) and those who had an interest in learning about feelings were 53% more likely (38.4% versus 25.0%). Prior survey participation also mattered: those reporting prior participation in surveys were over 50% more likely to consent (40.3% versus 26.3%).

The association was curvilinear in that lower participation rates and higher participation rates in prior surveys were associated with relatively lower participation rates in the current study whereas having completed 1 or 2–4 survey were associated with higher participation rates. Finally, high levels of openness to experience were associated with a 66% higher rate of consent (39.7% versus 23.9%)

Question 4. Does the burden of an EMA protocol affect uptake and how does it compare to a one-time survey?

Once participants were randomized in the three study arms, they had to pass additional eligibility criteria for the conditions (for example, not being a shift worker for the EMA conditions). Ten individuals were eliminated following randomization based on eligibility requirements specific to the conditions: 4 were dropped in the one-time survey group, 4 in the low-burden EMA group, and 2 in the high-burden EMA group.

A total of 74 individuals were included in the three experimental study arms: 26 into the one-time survey group, 23 into the low-burden EMA group, and 25 into the high-burden EMA group. Prior to conducting analyses of uptake rates by group, we examined the demographic and individual differences (based on the demographic survey questions) among the groups; these comparisons are shown in Appendix Table 5. No significant differences were detected. However, we recognize that the minimum detectable effect given the modest sample size was relatively large at $w = 0.44$.

The rate of consent and uptake was computed for each of the three groups: for the one-time survey group, the rate was 100% (26/26); for the low-burden EMA group, the rate was 78.3% (18/26); and, for the high-burden EMA group, the rate was 68.0% (17/25). A Fisher Exact test including all three groups indicated significant differences among the groups ($p = .003$). Subsequent post hoc tests combining groups showed that the combined EMA groups differed significantly from the one-time survey group ($p = .003$) and that there was no difference between the two EMA groups ($p = .523$).

When interpreting the randomized group uptake rates, one should keep in mind that only 2.5% of those sent invitations qualified for the randomized trial. Uptake rates reported above for the randomized groups are based on the number of individuals who qualified for being randomized into the study (a total of 74 people; Table 1), but another perspective on the rates, which is addressed in the Discussion, takes into account the overall uptake rates. These rates are very low (under 1% across groups) given the large number of individuals who did not reach the randomization stage of the study.

A final possibility that we wished to address was that, although person characteristics were not associated with

Table 4 Differences on the demographic survey variables for those who consented and participated versus those who did not

Demographic variable	<i>N</i>	Percent participated	LR χ^2	<i>p</i> -value
Age (years)	200		2.16	0.141
25 years		21.9%		
50 years		28.4%		
75 years		35.8%		
Male			2.24	0.135
No	61	37.7%		
Yes	137	27.0%		
White			0.0	0.915
No	24	29.2		
Yes	172	30.2		
With bachelor's degree or greater	200		3.60	0.058
No		15.4%		
Yes		32.8%		
Income greater than or equal to \$75,000			4.56	0.033*
No	70	24.3%		
Yes	106	39.6%		
Married			2.73	0.099
No	119	37.2%		
Yes	78	26.1%		
Subjective Well-being	198		0.19	0.662
25th		29.6%		
75th		31.5%		
Memory Prob			2.08	0.149
No	187	29.4%		
Yes	12	50.0%		
Poor health	199		0.00	0.949
25th		30.5%		
75th		30.8%		
Poor Current mood	199		0.24	0.625
25th		29.6%		
75th		31.3%		
Computer skill	181		4.22	0.040*
25th		24.2%		
75th		33.8%		
Typing skill	181		7.25	0.007**
25th		27.0%		
75th		39.5%		
Smartphone skill ¹	172		0.29	0.593
Smartphone confidence	171		2.61	0.106
25th		32.5%		
75th		39.4%		
Own smartphone	199		1.63	0.202
No		17.7%		
Yes		30.7%		
Importance research topic	197		13.26	<0.001***
25th		25.1%		
75th		36.7%		
Interest in feelings	196		15.70	<0.001***
25th		25.0%		
75th		38.4%		
Prior survey participation	199		3.87	0.049*
25th		26.3%		
75th		40.3%		

Table 4 (continued)

Demographic variable	<i>N</i>	Percent participated	LR χ^2	<i>p</i> -value
Frequency of Prior surveys	198		17.7	0.001***
0 surveys		18.1%		
1 survey		39.5%		
2–4 surveys		51.1%		
5–7 surveys		27.3%		
8 or more surveys		20.0%		
Personality: Extraversion	193		1.19	0.275
25th		29.3%		
75th		34.4%		
Personality: Agreeable	193		1.47	0.226
25th		29.1%		
75th		34.7%		
Personality: Conscientious	193		0.13	0.717
25th		32.5%		
75th		30.5%		
Personality: Neuroticism	190		0.66	0.415
25th		34.1%		
75th		30.2%		
Personality: Openness	195		9.55	0.002**
25th		23.9%		
75th		39.7%		

Notes.

¹ Over half of the responses in Smartphone Confidence were 3s, so the 25th and 75th percentiles were equivalent. Therefore, these statistics are not presented, but overall test did not approach significance.

Memory Prob: In the past 12 months, have you talked to a health care provider about memory problems?

Poor Health: In general, would you say your health is:

Poor Current Mood: How is your mood right now?

Computer skill: If you own or use a computer, how would you rate your computer skill level?

Typing skill: In general, how confident are you in using a computer for writing tasks that involve typing on the computer keyboard, such as answering an e-mail?

Smartphone skill: If you own or use a smartphone, how would you rate your smartphone skill level?

Own smartphone: Do you have a smartphone?

Importance research topic: How important is it for you that you are interested in the topic of the survey when you decide to participate in a study?

Prior surveys: Do you take surveys for any other companies or organizations?

Freq Prior Surveys: How often do you participate in marketing or other research surveys (paid or non-paid)?

Interest in project: How much are you interested in research that tries to learn more about people's experiences and feelings in daily life?

uptake for the entire group of 74 who were randomized, it could be that those characteristics were differentially associated with uptake by group. A problem arises for such analyses, because everyone in the one-time survey group agreed to participate, so there is no variation in uptake for that cell of the design. Logistic regression cannot estimate a model evaluating interactions (here, group by a survey variable) in cases when there is no variation in a cell (only main effects are possible). Thus, we were not able to evaluate this question.

Discussion

The broad goals of this study were to increase knowledge about uptake rates in population-based research employing momentary data capture and to explore the potential for self-selection bias. The study incorporated several noteworthy methodological features including: the use of a national sample for recruitment where selected demographic information was available for the entire

sample; the use of a recruitment protocol that simulated actual recruitment into a study; and a randomized experiment to determine uptake rates for a one-time survey, a low-burden EMA study, and a high-burden EMA study. These design features yielded estimates of uptake rates to the three protocols and allowed us to examine demographic and individual difference variables at different stages of recruitment, ones that are likely to be pertinent to the evaluation of selection bias in survey and momentary studies. As mentioned earlier, confirming that self-selection translates into biased study results depends on the extent to which factors underlying the nonresponse relate to specific associations being investigated and, therefore, could not directly be addressed here. We could, though, estimate how the recruitment methods altered the composition of the final, randomized sample compared to the population from which it was drawn and draw inferences from those rates.

We first consider the response rate to the mailed recruitment letter. Of the 3,000 mailed, only about 14% responded in any manner. There are many plausible reasons for these nonresponses including the obvious explanation that the targeted individuals simply discarded the invitation letter or read the invitation and decided not to respond. Another plausible explanation is that individuals no longer lived at the mailing address, yet we did not receive notification of this from the postal service. In any case, this rate should be viewed in the context of what we believe to be the good recruitment practices employed in this study (Dillman et al., 2014), consisting of individually addressed envelopes and letters, an immediate monetary incentive, a follow-up reminder, and the availability of multiple modes for responding to the solicitation. This response rate seems particularly low considering that those who responded saying they had no interest in the study were counted as positive responses. Perhaps a more relevant indicator of responsiveness to the solicitation is the proportion of the sample that completed the demographic survey, keeping in mind that doing so did not yet commit individuals to any additional participation, but did indicate some level of willingness and interest; this rate was 6.9% of those included in the original sample. Thus, only about 7 of 100 mailed invitations yielded a response indicative of interest in learning more about the study, albeit with the requirement of the completion of a brief demographic survey.

These rates need to be considered in light of previous survey research employing similar methods. Prior work has shown uptake rates comparable to those observed here. The American Family Health Study used addressed-based sampling for completion of a screening questionnaire and had a 5.3% uptake rate with a \$2 incentive (<https://afhs.isr.umich.edu/about-the-study/afhs-methodology/>). Using similar methods, the AmeriSpeak panel had a response rate of 5.8% (Bilgen et al., 2020). A slightly

lower rate was achieved in a credit bureau sample, 3.9% (Bucks et al., 2020), 3.7–5.2% in a sample from taxpayer records (Koenig et al., 2021) (there were variations in the procedures), and yet another study using address-based sampling found a rate of 7.0% (Winneg et al., 2021). On the other hand, some research groups have reported higher uptake rates. The Health Information National Trends study (Westat, 2021) conducted an addressed-based survey in 2019 and reported uptake rates as high as 30%, whereas a 2020 survey in Nebraska observed a 15.8% rate. Clearly there is considerable variation in address-based recruitment rates, and the rate observed in this study falls within the range of previous findings, in line with the many studies at the lower end of the spectrum.

The next uptake rate we consider required that prospective participants progressed to the stage of recruitment where they were randomized in the study groups. Of those who completed the demographic survey (202), 84 agreed to be randomized into an arm of the study (2.8% of total sample; 41.6% of those who completed the demographic survey). Notably, at this point in the recruitment process the specifics of the research design (i.e., the protocols of the three study arms) were not known by prospective participants. Thus, the decision to participate or not could not have been based on the specifics of the study protocols. A few additional respondents were deemed not eligible at this point, dropping the uptake rate for the experiment to 2.1% of the total sample ($n = 74$). That about 2 of every 100 individuals to whom recruitment letters were mailed agreed to be randomized in the study appears very low. Again, we wondered if there was a shortcoming in the recruitment protocol that reduced interest in the study. We reviewed the protocol yet could not identify any shortcomings in the materials or procedures that deviate from accepted survey recruitment practice.

The observation that only 2% entered randomization does, we believe, increase the likelihood of self-selection bias given the high probability that there are at least some characteristics of the final sample that differ from the population. As mentioned earlier, the final determination of whether these differences result in selection bias is dependent upon the associations studied. In any case, the results of this study support the contention of many researchers that uptake rates are low when drawn from the general population.

Furthermore, the phenomenon of low uptake was not unique to momentary studies, because the exact nature of the research protocol had not yet been disclosed to those who opted out prior to randomization. Potential participants could have imagined that data would be collected with any kind of procedure; in fact, it seems likely that they would have imagined a conventional questionnaire assessment considering that momentary methods are not well publicized to the general population.

Our next major objective was to examine demographic and individual differences variables contrasting those who opted out at various stages in the recruitment process with those who did not. Observed differences could be indicative of selection bias. We first consider comparisons between those who completed the demographic survey (less than 7% of the sample) with those who did not (Question 2). When designing the study, we hoped that the uptake rate in the screening would have been higher, which would have made the group comparisons more balanced. The only demographic difference we found based on the (limited) data from MSG was that whites were more than twice as likely to complete the demographic survey compared to non-whites, which was a prerequisite for subsequent participation. Age and gender did not distinguish the participants in completion of the demographic survey. Thus, even at this stage of the process, sample differences emerged and these could be signs that self-selection bias is plausible.

We now turn to an important set of comparisons: the ones between those who expressed interest in the study versus those who consented and participated in the study (Questions 3a and 3b). This goes to the core theme of self-selection: are there differences in those who ultimately participate in studies versus those who do not. One strong difference observed with the MSG data was that whites were about two and a half times more likely to have consented and participated than non-whites. A second difference was that females were almost twice as likely to participate than were males, although this finding barely reached statistical significance. Considering that these results are based on only three variables available from the marketing company, we suspect that there are many other differences that would have been detected if we had access to more comprehensive information about the entire sample. As previously discussed, the observed race and gender differences are consistent with prior work on self-selection into survey studies. Those studies have shown that race can be linked to uptake, where whites were more likely to do so (Couper et al., 2007; Kim et al., 2008). Regarding gender, several studies have found that women were more likely to participate in studies than men (Abraham et al., 2006; Andreeva et al., 2015; Burg et al., 1997; Galea & Tracy, 2007; Keeble et al., 2016), consistent with the results reported here.

Prior to discussing the screening survey results (Question 3b), we remind readers that these 202 individuals were themselves different from the overall sample: as just shown, they were more likely to be white and female. That is, interpretations of comparisons using this sample should take into account that this is already a self-selected sample. Many sample characteristics were associated with consent and uptake. Those with higher income were about 60% more likely to consent/participate, those with better computer skills (comparing the 75th and 25th percentile of participants) were 40% more likely, those more experienced using smartphones were 45% more likely, those who viewed the research topic

important were 46% more likely, those reporting an interest in peoples' feelings were 53% more likely, those who participated in prior surveys were 53% more likely, and, finally, those reporting higher levels of openness to experience were 66% more likely to consent/participate. Regarding the number of prior surveys taken, we found a curvilinear association such that low uptake into the current study was associated with having taken no prior surveys and with taking 8 or more surveys (18–20%), whereas those who took between 1 and 7 prior surveys had higher response rates (27–51%).

These results are generally consistent with prior work on factors associated with uptake, though some previous results have been mixed regarding demographic correlates. Higher socioeconomic status, being employed, being married, and having higher education have also been shown to be positively associated with uptake rates (Abraham et al., 2006; Andreeva et al., 2015; Galea & Tracy, 2007; Keeble et al., 2016; Partin et al., 2003). However, we only observed differences on income and not on education. Consistent with some prior work (Abraham et al., 2006; Andreeva et al., 2015; Galea & Tracy, 2007; Keeble et al., 2016; Partin et al., 2003), we observed that those who participated were older than those who did not; however, other research has not observed this effect (Couper et al., 2007; Matías-Guiu et al., 2014; O'Neil, 1979). We can only speculate about the curvilinear association with uptake and number of prior surveys taken. Perhaps those who did not take prior surveys were generally inclined not to participate, but nevertheless had an interest in hearing what the study entailed. And possibly those who completed many prior surveys were simply disinclined (tired) to do more. However, individuals in these groups *did* complete the demographic survey, which undermines these speculations.

An important observation is that those who consented and participated were more interested in the research topic and, relatedly, had an interest in studies trying to understand more about people's feelings. Although the topic of this study, as introduced in the introductory letter, was certainly broad and seemingly noncontroversial, we speculate that different uptake rates could be observed in studies stating other objectives. For instance, surveys on political topics might engage (or dissuade) segments of the population, resulting in either lower or higher uptake rates and impacting the chance of self-selection bias. The point is that our findings are linked to the various characteristics of the study design, which must be viewed as limiting generalizability of these findings. Uncovering person characteristics that are most strongly associated with self-selection, both broadly across different study designs and specific to EMA studies, is an important agenda for future research. The identification of such characteristics would inform improved strategies to compensate for self-selection, such as targeted oversampling of participant subgroups and the identification of variables that need to be included in inverse probability weights.

We now turn to the results of the randomized experiment (Question 4). Seventy-four individuals qualified for the final protocol to test how data collection modality (one-time survey, EMA) and burden (low- versus high-burden EMA) affected uptake. One way to interpret these results is to view this sample as analogous to those that express interest in participating in studies using advertisements (newspapers, radio, social media, posters, and flyers, for example). In this case a researcher has no knowledge of the number exposed or the characteristics of those exposed, because people who were not interested have already self-selected out. Thus, the uptake rates we report can be viewed as representative of those offered participation subsequent to aforementioned self-selection processes.

In any case, the results were that all respondents (100%) in the one-time survey condition agreed to do the survey whereas 73% of the EMA conditions agreed to enter the study (both EMA groups combined), a significant difference despite the relatively small number of participants. No significant difference in uptake rates was detected between the two momentary groups. A complementary way of viewing the results is compute the uptake rates with on the original sample of 2,938 as the denominator. These uptake rates are now all under 1%: 0.88% for the one-time survey group, 0.61% for the low-burden EMA group, and 0.58% for the high-burden EMA group, a perspective that highlights that group differences in uptake are quite low in an absolute sense. We advise caution for interpreting the specific magnitude of the group differences given the relatively small sample used for the estimates, and replication would certainly be desirable. We also note a limitation for the interpretation of these findings in that there was an additional layer of eligibility screening for those assigned to the EMA groups. Some respondents were eliminated from the EMA groups by this process, which may be interpreted as undermining the strict randomization.

For momentary researchers an important question concerns the interpretation of the difference in uptake between the survey and momentary methods. The one-time survey uptake showed a relative 27% advantage (100% agreement for the one-time survey compared to 73% for EMA), but the absolute rates were low in both groups overall (i.e., 0.88% for one-time survey participants versus about 0.75% for EMA participants based on the full sample). One plausible takeaway is that momentary researchers can take comfort in the fact that a large majority of those who would agree to complete the one-time survey would also be likely to agree to enter the momentary data collection protocols. To our knowledge, this is the first study to compare uptake rates across survey and momentary studies. Furthermore, although the low-burden EMA group was matched on time commitment and monetary incentive to the one-time survey group (whereas the high-burden EMA group faced a much greater time commitment), we observed only a small difference between the uptake rates for the two EMA groups. Thus, an even more burdensome momentary protocol fared reasonably

well in terms of uptake in comparison to the traditional one-time survey group.

Notably, prior work has employed other methods for assessing uptake differences among EMA protocols varying in the number of daily prompts, length of prompts, number of days in the study, and other design features (Eisele et al., 2022; Hasselhorn et al., 2022; Smyth et al., 2021). The method used in one of our prior studies was to simply ask potential participants if they would consent to enroll in various hypothetical EMA protocols. Given the greater verisimilitude of the simulated procedure employed here, we recommend that prior results on uptake rates for EMA design factors be replicated with the simulation procedures or in studies that vary design factors and run subjects through the full EMA protocol. Though costly, the latter design also has the advantage of yielding additional data about compliance rates and how they may vary by design features.

There are general limitations to results presented here. First, it is possible to challenge the veracity of the simulation, although we believe that the methods closely emulated standard practice for recruiting survey participants. Second, as mentioned earlier, the generalizability of the results may be limited in some respects, including the focus on the topic of everyday health and mood, the use of a marketing firm's sample, and the particulars of the recruitment methods. Third, the protocol involved deception and required thorough review by the Institutional Review Board. We only encountered slight disappointment in a few individuals who had consented and believed they would be starting their assigned study. Nevertheless, this could be an impediment to future research employing these methods.

In summary, this study compared selected demographic characteristics between those who responded to letters of invitation to participate in a study of everyday health and mood versus those who declined at various stages of the recruitment process. Only a small proportion of the sample responded to the invitation and smaller proportion decided to participate in the study (all of which occurred invitees were informed of the data collection method). The differences in those who responded versus those who did not were generally consistent with prior work and, in our opinion, the uptake rates suggest that selection bias is a plausible possibility. The embedded experiment with those respondents who expressed interest showed that uptake to a one-time survey was very high amongst those who were randomized and uptake to the more complex momentary protocols was lower, but was still moderate to high. However, these uptake rates were based on only a very small part of the original sample. These findings support the view that participant self-selection bias is plausible, if not likely, in the momentary and one-time survey designs we investigated, understanding that the potential for bias will be related to a particular study's goals.

Appendix 1

Table 5 Sample characteristics of the three randomization groups

Variable	Of those randomized, after eligibility screening (<i>n</i> =74)		
	One-time survey (<i>n</i> =26)	Low-burden EMA (<i>n</i> =25)	High-burden EMA (<i>n</i> =23)
	Average or Percent	Average or Percent	Average or Percent
Age (years)	57.0	60.0	63.1
Male	64.0%	70.8%	56.5%
White	80.0%	92.0%	95.5%
With bachelor's degree or greater	92.3%	88.0%	100.0%
Income greater than or equal to \$75,000	64.0%	59.1%	81.8%
Married			
No			
Yes			
Subjective Well-being	7.46	7.56	7.65
Memory Prob	11.5%	4.0%	13.0%
No			
Yes			
General health	3.46	3.64	3.56
Current mood	3.62	3.80	3.61
Computer skill	3.08	2.88	2.91
Typing skill	3.69	3.42	3.68
Smartphone skill	2.88	2.96	2.90
Smartphone confidence	3.32	3.13	3.38
Own smartphone	96.2%	92.0%	95.7%
Importance research topic	2.77	2.64	2.78
Interest in feelings	2.46	2.76	3.09
Prior survey participation	0.27	0.44	0.49
Frequency of Prior surveys	1.23	1.16	1.17
25 th			
75 th			
Personality: Extraversion	25.5	26.7	26.0
25 th			
75 th			
Personality: Agreeable	33.9	36.4	37.1
Personality: Conscientious	34.7	37.2	37.4
Personality: Neuroticism	21.3	18.6	19.8
Personality: Openness	39.0	38.6	38.6

Notes.

Memory Prob: In the past 12 months, have you talked to a health care provider about memory problems?

Health: In general, would you say your health is:

Current Mood: How is your mood right now?

Computer skill: If you own or use a computer, how would you rate your computer skill level?

Typing skill: In general, how confident are you in using a computer for writing tasks that involve typing on the computer keyboard, such as answering an e-mail?

Smartphone skill: If you own or use a smartphone, how would you rate your smartphone skill level?

Own smartphone: Do you have a smartphone?

Importance research topic: How important is it for you that you are interested in the topic of the survey when you decide to participate in a study?

Prior surveys: Do you take surveys for any other companies or organizations?

Freq Prior Surveys: How often do you participate in marketing or other research surveys (paid or non-paid)?

Interest in project: How much are you interested in research that tries to learn more about people's experiences and feelings in daily life?

Acknowledgements This research was supported by grant National Institute on Aging, R37AG057685 (PI: AAS).

Funding Open access funding provided by SCEL, Statewide California Electronic Library Consortium

Data Availability A redacted data set will be available upon request.

Declarations

Open science statement The data collected for this study are available at xxx. The study was not preregistered.

Competing interests None of the authors has any competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the american time use survey: who is missing from the data and how much does it matter? *International Journal of Public Opinion Quarterly*, 70(5), 676–703.
- Andreeva, V. A., Salanave, B., Castetbon, K., Deschamps, V., Vernay, M., Kesse-Guyot, E., & Hercberg, S. (2015). Comparison of the sociodemographic characteristics of the large NutriNet-Santé e-cohort with French Census data: the issue of volunteer bias revisited. *Journal of Epidemiology and Community Health*, 69(9), 893–898.
- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78, 161–188.
- Bilgen, I., Wiese, C. J., Liebert, L., Hendarwan, E., Singh, R., Xia, K., Pineau, V. J., & Dennis, J. M. (2020, February). *Examination of recruitment mailing design decisions for a probability-based Research Panel*. CIPHER.
- Bucks, B., Couper, M. P., & Fulford, S. L. (2020). A mixed-mode and incentive experiment using administrative data. *Journal of Survey Statistics and Methodology*, 8, 352–369.
- Burg, J. A. R., Allred, S. L., & Sapp, J. H. (1997). The potential for bias due to attrition in the national exposure registry: an examination of reasons for nonresponse, nonrespondent characteristics, and the response rate. *Toxicology and Industrial Health*, 13(1), 1–13.
- Cantril, H. (1965). *Pattern of human concerns*. Rutgers University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Erlbaum Associates.
- Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: a modern idiographic approach to personality research. *Social and Personality Psychology Compass*, 3(3), 292–313. <https://doi.org/10.1111/j.1751-9004.2009.00170.x>
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design & analysis issues for field settings*. Rand McNally College Pub. Co.
- Costa, P. T., & McCrae, R. R. (n.d.) *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual* Odessa, FL Psychological Assessment Resources 1992.
- Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an internet survey. *Social Science Research*, 36(1), 131–148.
- Csikszentmihalyi, M., & Hunter, J. (2003). Happiness in everyday life: the uses of the experience sampling. *Journal of Happiness Studies*, 4, 185–199.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method* (4th Edition). Wiley.
- Dunn, K. M., Jordan, K., Lacey, R. J., Shapley, M., & Jinks, C. (2004). Patterns of consent in epidemiologic research: evidence from over 25,000 responders. *American Journal of Epidemiology*, 159, 1087–1094.
- Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment: an innovative and promising approach for clinical psychology. *European Psychologist*, 14, 109–119.
- Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, 29(2), 136–151. <https://doi.org/10.1177/1073191120957102>
- Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: a discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods*, 22, 969–1006.
- Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of Epidemiology*, 17(9), 643–653.
- Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2022). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods*, 54(4), 1541–1558. <https://doi.org/10.3758/s13428-021-01683-6>
- Heckman, J. J. (2010). Selection bias and self-selection. In S. N. Durlauf, & L. E. Blume (Eds.), *Microeconometrics*. The New Palgrave Economics Collection.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: measuring the quality of Everyday Life*. Sage.
- Jäckle, A., Burton, J., Couper, M., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: coverage and participation rates and biases. *Survey Research Methods*, 13(1), 23–44.
- Keeble, C., Baxter, P., Barber, S., & Law, G. (2016). Participation rates in epidemiology studies and surveys: a review 2005–2007. *The Internet Journal of Epidemiology*, 14(1), 1–14.
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, 83(Suppl 1), 210–235. <https://doi.org/10.1093/poq/nfz007>
- Kim, S. Y., Tucker, M., Danielson, M., Johnson, C. H., Snesrud, P., & Shulman, H. (2008). How can PRAMS survey response rates be improved among american indian mothers? Data from 10 states. *Maternal and Child Health Journal*, 12(Suppl 1), 119–125. <https://doi.org/10.1007/s10995-008-0334-4>
- Koenig, T., Bennett-Harper, S., Stapleton, M., Newsome, J., Levin, K., Leary, S., Javaid, R., Langetieg, P., & Schafer, B. (2021, May). *Will web response ever catch up to Paper? Comparing response rates in Standard and web push experimental conditions*

- in a Household Survey. American Association for Public Opinion Research.
- Materia, F. T., & Smyth, J. M. (2021). Acceptability of intervention design factors in mHealth intervention research: experimental factorial study. *JMIR mHealth and uHealth*, *9*(7), e23303. <https://doi.org/10.2196/23303>
- Matías-Guiu, J., Serrano-Castro, P. J., Mauri-Llerda, J., Hernández-Ramos, F. J., Sánchez-Alvarez, J. C., & Sanz, M. (2014). Analysis of factors influencing telephone call response rate in an epidemiological study. *The Scientific World Journal*, *2014*, 17937. <https://doi.org/10.1155/2014/179375>
- McCrae, R. R., & Costa, P. T. Jr. (1999). A five-factor theory of personality. *Handbook of personality: theory and research* (2nd ed., pp. 139–153). Guilford Press.
- O’Neil, M. J. (1979). Estimating the nonresponse Bias Due to refusals in telephone surveys. *Public Opinion Quarterly*, *43*(2), 218–232.
- Partin, M. R., Malone, M., Winnett, M., Slater, J., Bar-Cohen, A., & Caplan, L. (2003). The impact of survey nonresponse bias on conclusions drawn from a mammography intervention trial. *Journal of Clinical Epidemiology*, *56*(9), 867–873.
- Scollon, C. N., Kim-Prieto, C., & Diener, E. (2009). Experience sampling: promises and pitfalls, strengths and weaknesses. *Social indicators research series: Vol.39. Assessing well-being: the collected works of Ed Diener* (39 vol., pp. 157–180). Springer Science + Business Media.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.
- Smyth, J. M., Jones, D. R., Wen, C. K. F., Materia, F. T., Schneider, S., & Stone, A. A. (2021). Influence of ecological momentary assessment study design features on reported willingness to participate and perceptions of potential research studies: an experimental study. *BMJ Open*, *11*(7), e049154. British Medical Journal Publishing Group.
- Stone, A. A., Schneider, S., & Smyth, J. M. (2023). Evaluation of pressing issues in ecological momentary assessment. *Annual Review of Clinical Psychology*, *19*.
- Stone, A. A., Schneider, S., Smyth, J. M., Junghaenel, D. U., Wen, C. K. F., Couper, M.P., & Goldstein, S. (in press). Shedding light on participant selection bias in Ecological Momentary Assessment (EMA) studies: Findings from an Internet panel study. *PLoS-One*.
- Stone, A. A., & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*, 199–202.
- Westat (2021). *Health Information National Trends Survey 5 (HINTS 5): Web pilot results report*. <https://hints.cancer.gov/data/methodology-reports.aspx>
- Winneg, K., Pasek, J., Turakhia, C., Su, J., & Jamieson, K. H. (2021). *Over-recruitment and selective retention: examining a strategy for increasing panel selection efficiency*. American Association for Public Opinion Research.
- Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. <https://doi.org/10.1037/pas0000685>
- Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, *20*, 36–44.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.