Understanding Society

Working Paper Series

No. 2023 – 01

March 2023

Participation of household panel members in daily burst measurement using a mobile app: Effects of position of the invitation, bonus incentives, and number of daily questions

Annette Jäckle[1], Jonathan Burton[1], Mick P. Couper[2], and Brienna Perelli-Harris[3]

[1]University of Essex, [2]University of Michigan, [3]University of Southampton

Understanding Society

Economic and Social Research Council

Non-technical summary

Respondents in wave 13 of the *Understanding Society* Innovation Panel (IP13) were asked to install an app measuring wellbeing, and use it at the end of every day for 14 days to answer a set of questions. We examine which protocols work best to implement this type of data collection. Who agrees to participate in these additional activities and who does not? What barriers are there to participation? Do those who participate differ from those who do not in ways that may affect the conclusions drawn from such intensive measurement studies? We experimentally vary three features of the study protocol: (1) at what point within the annual interview we asked respondents to participate in the wellbeing study (early vs late), (2) how long the daily questionnaire was (2 vs 10 mins), and (3) the incentives for completing the app study.

Overall, 44.6% of the IP13 respondents downloaded and used the app at least once, while 11.1% completed all 14 days of the daily app survey. Including the invitation to download the app earlier in the survey had a significant positive effect on participation, while the incentive and questionnaire length treatments did not have significant effects.

A number of barriers were reported by respondents who tried but failed to install the app, including being unable to find the app (41.5%), difficulty logging in (14.8%), and not having a compatible device (10.4%). Among those who did not download the app, more than half (58.8%) cited lack of interest or time, while 32.1% selected technical or ability related reasons. Privacy and/or data security concerns were reported by 11.0% of respondents.

Finally, we examined how the group of participants differed from the initial group of IP13 respondents at each stage of the process (downloading the app, using it at least once, and fully complying with the protocol) in terms of socio-demographic and health-related variables. For most characteristics the patterns of biases are consistent, with younger age groups, those with a degree, in work, and who use health apps over-represented, and those who are divorced, separated or widowed, with a long-standing illness or disability, who use only a tablet or neither a tablet nor a smartphone, and who never install and use mobile apps under-represented across all steps of the protocol.

These findings point to the importance of developing methods to increase participation in intensive measurement studies and, in particular, to paying attention to differences in participation across key subgroups of the population.

Participation of household panel members in daily burst measurement using a mobile app: Effects of position of the invitation, bonus incentives, and number of daily questions

Annette Jäckle (University of Essex)

Jonathan Burton (University of Essex)

Mick P. Couper (University of Michigan)

Brienna Perelli-Harris (University of Southampton)

Abstract

Smartphone apps are increasingly being used for a variety of studies involving intensive measurement. Many of these studies are conducted among small groups of volunteers where selection bias is of less concern and ways to increase participation have been largely ignored. As these app-based studies are being included in large population-based surveys to permit broader inference and extend measurement across a broad range of topics, concerns over participation (or otherwise) in such studies and the implications for selection bias are becoming more salient. Given this, we conducted an experiment to test the effect of various features of an app study protocol on uptake or participation. Respondents in a large nationally-representative longitudinal survey in Great Britain were asked to download an app and use it every evening for 14 days to answer questions about their experiences and wellbeing that day. We experimentally varied three features of the study protocols: Inviting participants to the app study earlier in the annual interview increased participation rates, while a shorter daily questionnaire and bonus incentives for completing all 14 days had no significant effect. We also examine the effects of study protocols on non-participation bias, as well as the problems respondents had with installing the app and reasons for not participating.

Keywords: experiment, electronic diary, smartphone, response rates, non-response bias

JEL classification: C81, C83

Acknowledgements: This research was funded by the Economic and Social Research Council (ESRC) grant for the *Understanding Society* survey (ES/N00812X/1). The mobile application for this study was developed by Connect Internet Solutions Ltd.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom, aejack@essex.ac.uk.

1 Introduction

Survey respondents are increasingly being asked to perform a number of tasks in addition to answering the survey questions. These additional tasks include the provision of biospecimens (blood, saliva, etc.), undergoing additional physical tests (grip strength, height and weight measurement, etc.), consent to administrative data linkage, participating in ecological momentary assessments (EMAs), wearing activity trackers, downloading passive geolocation trackers on their devices, and using mobile devices to answer additional survey questions. There is a growing body of research (reviewed below) trying to understand who complies with these additional requests and who does not, and why; what survey design features could improve compliance with these additional requests and participation in these added tasks; what biases (if any) are produced by differential selection into these supplemental data collection activities; and so on.

As these new measurement tools are increasingly being incorporated into survey protocols, it is important to understand the error properties of these additional measures, in the same way that the total survey error perspective does for the surveys themselves (see Groves 2009). Further, understanding the similarities and differences between the various additional tasks (e.g., whether they require active engagement from participants or capture data passively after installation; whether they are one-time or ongoing, etc.; see Wenz, Jäckle and Couper 2018) will help us learn to understand why survey respondents participate in some of these activities but not others. It is also important to understand what methods and protocols work best for studies such as this. Developing optimal protocols for maximizing participation and minimizing the potential for selection bias is a critical step in wider adoption of these new methods.

The focus of this paper is on one particular activity embedded in a nationally-representative longitudinal probability sample in Great Britain. Respondents in wave 13 of the *Understanding Society* Innovation Panel were asked to download and use an app to answer a set of questions to capture daily changes in assessment of mood, happiness, relationship quality over a 14-day period. We experimentally vary three features of the study protocol and examine the following research questions:

RQ1: What are the overall rates of participation or compliance at each stage of the app study? Given that it is relatively rare to embed electronic daily diary studies in an ongoing longitudinal panel of the general population, our first question is to examine uptake at each of the stages in the process, from being willing to participate, to downloading and using the app at least once, to completing all 14 daily surveys.

RQ2: Which features of the request increase participation in a daily app study? We experimentally manipulated three features designed to improve participation rates in the daily app study: the offer of a bonus incentive, the stated length of the daily survey, and the position of the invitation to the app study in the interview. We examine the effects of each of these treatments on the different stages of participation.

RQ3: What are the barriers to participation in an app-based study? This research question is largely exploratory. Given that the study requires the use of an app, we are interested in reasons why those who expressed willingness to participate in the study may have failed to download and use the app. This could guide the design of specific protocols to increase the success of this stage of the protocol.

RQ4: What selection biases (if any) are introduced at different stages of the app study? Do study protocols affect biases? This research question focuses on changes in sample composition across different stages of the study protocol. We expect sample loss (drop out) at each stage (see RQ1). Of potentially greater concern is *differential* sample loss, i.e., increased selectivity of the sample of compliers in the app study relative to the full sample of survey respondents. That is, we examine selection bias conditional on participating in the main survey, not relative to the general population. We also examine whether the extent and nature of biases varies between the experimental treatment groups.

2 Background

In intensive measurement studies such as this, (non)response can take many forms, including agreeing to participate in the study and download the app (in some studies formal consent may also be required), actually downloading and installing the app, and completing the daily surveys in accordance with the protocol. We use the terms *initial willingness*, *compliance*[1] with downloading the app and *ongoing compliance* with the daily protocol respectively to describe these steps. We further define ongoing compliance as the number of survey days submitted, conditional on completing at least one day (i.e., successfully downloading the app). That is, we do not examine missing data within the daily surveys.

Our study involves one of a family of methods variously called measurement burst designs (Patrick, Maggs and Lefkowitz 2015), intensive longitudinal designs (Bolger and Laurenceau 2013), or "slice of life" designs (Smyth et al. 2017). This includes approaches such as ecological momentary assessment (EMA); Shiffman, Stone and Hufford 2008), experience sampling method (ESM; Csikszentmihalyi and Larson 1987), the day reconstruction method (DRM; Diener and Tay 2014; Kahneman et al. 2004), diary methods (Bolger, Davis and Rafaeli 2003), and so on. Early versions of intensive measurement or designs were conducted by telephone (e.g., Sliwinski et al. 2009) or IVR (e.g., Lee et al. 2020), but increasingly web- or app-based approaches (see de Vries et al. 2020) or SMS (e.g., Cardenas and Stormshak 2019) are being used.

These methods are used to measure a wide variety of topics, including subjective states (mood, personality), behaviour (alcohol use, nutrition, sleep), health conditions (pain, fatigue), social relations, and so on. With a few exceptions, these studies are based on relatively small samples of volunteers, with little attention to whether the sample is representative of the broader population. Most papers using these methods focus on the substantive outcomes. The literature on issues of implementation are scarce. Given the focus on volunteer participants, there has been little focus on initial cooperation rates and selection biases. The focus is on compliance rates or compliance bias, defined in various ways based on levels of participation in the intensive measurement protocol, following initial recruitment and (in many cases) completion of a set of baseline measurements, as well as (in some cases) an in-person training session.

[1] The term "compliance" has been criticized in the medical literature as reflecting a paternalistic view of patients following doctors' instruction, and "adherence" or "concordance" have been proposed as alternatives (e.g., Bissonnette 2008; De las Cuevas 2011; Lutfey and Wishner 1999; Tilson 2004). While these latter terms may better reflect the voluntary nature of participation in these activities and "adherence" is often used in the context of medical treatment regimens, the term "compliance" is pervasive in the intensive measurement literature, and we use it here for convenience.

Compliance rates are often not reported in the literature and (when reported) often not clearly defined. For example, Morren et al. (2009) conducted a systematic review of 62 studies using electronic pain diaries, focusing on response, attrition, and compliance rates. They defined the response rate as the number of individuals who were willing and able to participate and conformed to the study's inclusion criteria, divided by the total number of individuals identified as potential participants. Attrition was defined as the percentage of respondents who were not included in the final analyses, for instance because they withdrew, made insufficient diary entries, or when technical malfunction occurred. Compliance was reported in different ways, and was defined as the percentage of diary entries, or the percentage of completed items. The response rate was reported in less than a third of the studies (mean rate = 53%), almost three in four studies reported the attrition rate (mean rate = 19%), and only nine studies (21%) reported both attrition and response rates. About 75% of the studies reported the compliance rate; the mean rate across studies was 83%. In a later systematic review of 104 publications based on 62 unique quantitative EMA studies on pain, May and colleagues (2018) found that nearly one third of the studies (31.7%) did not report completion rates. Similarly, de Vries, Baselmans and Bartels (2021) conducted a systematic review of 53 smartphone-based EMA studies focusing on wellbeing. They report a median sample size of 97, but a wide range, including two very large public-facing studies, the Mappiness app (n=2,250) and Track your Happiness app (n=26,700). Only 25 of the studies (47.2%) reported information on compliance with the EMA, with an additional 7 providing enough information to calculate a compliance rate. Based on the 32 studies, participants completed on average 71.6% (SD = 14.1%) of all EMAs with a range of 43–95%. See van Berkel et al. (2020), Wen et al. (2017) and Williams et al. (2021) for additional reviews of compliance rates in intensive measurement studies.

Generally, compliance is defined as the percent of assessment or surveys responded to in line with the protocol (see, e.g., Jones et al. 2019; Morren et al. 2009). For signal-contingent methods, compliance is sometimes defined as responses received within a specified time frame around the signal (e.g., Liddle et al. 2017; Williams et al. 2021). In an early paper, Stone and Shiffman (2002) laid out guidelines for reporting compliance in EMA and ESM studies. However, as Green et al. (2006, p. 102) note, "Different designs call for different definitions of adequate compliance and response rates." Given the voluntary nature of many of these studies, there is little research on differences in compliance rates by socio-demographic characteristics, and even less on compliance bias.

In one recent exception, Stone and colleagues (2022b) used an address-based general population sample in the U.S. to invite people directly to a low-burden EMA study (three 1-minute prompts per day over 7 days), a high-burden EMA study (six 2-minute prompts per day over 14 days) or a one-time 40- minute web survey. Invitees were asked to complete a short survey to measure covariates and identify eligibility. Of those invited, 85.9% did not respond to the invitation in any manner. Only 2.8% expressed interest in participating and, ultimately, 2.1% of the sample entered the randomized portion of the study. Given that only 74 persons were randomized the three conditions, differences between the low- and high-burden EMA studies were not detected (both around 73%), but all who were invited to complete the web survey did so. They concluded that the low "uptake rates suggest that selection bias is a plausible possibility."

There is a growing body of research in the survey methodology literature on survey respondents' stated willingness to participate or on actual participation in additional app-based activities involving intensive measurement.

Several papers have focused on hypothetical willingness (e.g., Keusch et al. 2019; Revilla, Couper and Ochoa 2019; Struminskaya, Toepoel et al. 2020; Wenz, Jäckle and Couper 2019). Other papers have focused on passive measurement (i.e., downloading an app to collect data in the background) (e.g., Keusch et al. 2019; Kreuter et al. 2020; Struminskaya, Lugtig et al. 2021). But there has been relatively little research to date on active participation using a mobile app embedded within a large-scale survey. In one exception, Jäckle and colleagues (2019) invited participants in the *Understanding Society* innovation Panel to download an app and record their expenditures (including scanning receipts) daily for one month. They found that 13% overall or 16% of mobile device owners installed and used the app at least once. Among those who did use the app, continued compliance across the course of the month was high (81.5% remaining for at least 29 days). In a follow-up study, Wenz et al. (2020) found that offering personalized feedback in the spending app did not increase initial participation or ongoing adherence. However, participants did react positively to the feedback. In the same study, Jäckle et al. (2022a), tested whether (1) inviting sample members to a mobile app study within an interview rather than by post after the interview, and (2) offering a browser-based follow-up to the mobile app increased participation. They found that the in-interview invitation increased participation rates without affecting the sample composition of participants, while the browser-based alternative increased participation rates and reduce participation biases.

In another example, McCool and colleagues (2021) conducted a field test of a travel study in which participants were asked to download an app and provide seven days of time-location sensor data along with annotations of trips. Half the sample came from the Dutch population register, while the balance had previously participated in a travel diary survey. App registration rates were 26.5% for the fresh sample and 44.4% for previous respondents. Their paper focused more on gaps in the passive (GPS) data than the active (annotation) data; nonetheless, they concluded that respondents "proved willing to provide annotative data to the passive traces."

Our particular design asks survey participants to download an app and complete a short daily wellbeing survey every evening for 14 consecutive days. Using Bolger, Davis and Rafaeli's (2003) terminology, this can be described as an interval-contingent electronic daily diary. Given the literature reviewed above, our focus is on how best to implement app-based data collection protocols to maximise participation and minimise compliance bias within the context of a large nationally-representative survey. We test three experimental manipulations, informed by the survey methods literature. We review relevant literature related to each of these below.

2.1 Incentives

There is a large literature on the use of incentives in surveys. Focusing on studies directly relevant to ours, Haas and colleagues (2021) experimented with different levels of monetary incentives in an app study involving passive measurement among Android smartphone users in a national labor force panel (PASS) in Germany (see also Keusch et al., 2022). Sample members were randomly allocated to an incentive of €10 vs. €20 for installing the study app, and to an incentive of €5 vs. €10 for sharing the full range of sensor data collected passively by the app. While the higher incentive for app installation led to a statistically significant but modest increase in participation rates, the higher incentive for data sharing did not significantly increase participation rates. Similarly, Jäckle and colleagues (2019) experimented with different levels of incentives for app download (£2 vs. £6) in an active spending diary app study, but did not find a significant effect on participation rates. McCool and colleagues (2021) experimented with three incentive conditions in their 7-day travel app study. One third was promised €5 conditional on registration and €5 conditional on seven days of recorded travel data, another third was promised €10 conditional on seven days of travel data and the third was promised €20 conditional on seven days of recorded travel data. Response rates across the three incentive conditions were 30.1% for €5+€5, 36.4% for €0+€10 and 39.7% for €0+€20. In two studies of hypothetical willingness to participate in app-based data collection, Keusch et al. (2019) and Pinter (2015) found positive effects of incentives. In sum, the effects of respondent incentives on participation in mobile app-based data collection vary across studies and seem to depend on how the incentive payment is structured.

2.2 Stated length of daily survey

We are not aware of studies looking at the length of the daily surveys. However, other measures of respondent burden or effort have been examined. Using hypothetical vignettes, Keusch et al. (2019) found that a shorter experimental period (one month as opposed to six months) and monetary incentives increased willingness to participate in an app-based study. In a similar factorial vignette study, Ságvári, Gulyás and Koltai (2021) also found positive effects of a shorter study period and higher incentives on willingness to participate in passive app-based data collection. In a vignette-based study of factors affecting willingness to participate in an EMA study, Smyth et al. (2021) similarly found that shorter study duration, fewer prompts, and higher compensation increased willingness to participate and participation likelihood. To support these findings on hypothetical burden, in a systematic review of (non-experimental) electronic diaries Morren et al. (2009) found that shorter diaries were associated with higher compliance rates (see also Daniëls et al. 2021). We found no other research experimentally varying the burden of an intensive measurement task, but the broader literature on survey burden suggests that a request involving a shorter survey or task should have higher rates of compliance than that for a longer one. For instance, Crawford, Couper and Lamias (2001), Galesic and Bosnjak (2009) and Marcus et al. (2007) all found higher response rates to web surveys with a shorter stated length in the invitations (but see also Blumenberg et al. 2019; Lugtig and Luiten 2021;

Stone et al., 2022b). Read (2019) provides a detailed discussion of respondent burden in the context of app-based studies. These studies suggest that a smaller request will result in higher compliance rates than a larger request.

2.3 Position of request in the interview

There is some evidence from previous studies that participation in additional tasks for a survey depends on when the request is made. Jäckle et al. (2021) varied whether the request was made in-interview or in a follow-up mailing. Introducing the app as part of the interview significantly increased participation: 22.6% of respondents invited within the interview went on to use the app at least once to report a purchase, compared to only 12.4% of respondents invited by post.

The consent literature shows earlier placement of the consent to record-linkage question within an interview is associated with higher consent rates. Asking for consent after a module of questions related to the content of the data to be linked increased consent compared to asking at the end of the questionnaire (Sala, Knies and Burton 2014), and asking it at the beginning of the survey rather than the end had a positive effect (Eckman and Haas 2017; Sakshaug et al. 2019; Burton et al. 2022). However, Jäckle et al. (2022b) found an effect of placement only when the standard (more difficult) wording of the consent request was used; when an easy version of the request was presented, position in the survey did not matter. We are not aware of prior research on whether the position of a request within an interview affects recruitment for app-based studies.

2.4 Non-participation bias

Finally, we examine selection bias at difference stages of participation in the app study. As noted earlier, relatively few studies have examined selection biases associated with participation in intensive measurement studies. Jäckle et al. (2019) found extensive coverage bias in who has and does not have mobile devices, and some bias in who participates in a spending study, conditional on having a device. Subsequently, Jäckle et al. (2022a) found that adding a browser-based alternative to an app-based spending study reduced, but did not eliminate, participation bias. Lugtig, Roth and Schouten (2022) explored various stages of nonresponse (registration, activation and completion) of a 7-day travel diary app in the Netherlands. Using population registration data, they found substantial nonresponse biases, much of which arises during the recruitment phase. Age and education were key correlates of participation with older persons and those with lower education being under-represented.

Other studies focus on biases occurring at a single stage in the participation process. Keusch et al. (2020) examined coverage error in terms of ownership of a compatible Android smartphone for an app study in Germany (see also Haas et al., 2021). They found substantial biases in demographic variables (such as household size, income, and employment status) and in key substantive variables (such as general life satisfaction, satisfaction with living standards and social inclusion), even after adjustments for demographic differences. Elevelt, Lugtig and Toepoel (2019) compared those who participated in all parts of a smartphone-based time use study in the Netherlands with the full sample (members of a probability-based online panel). Participants had reported significantly more time working (+5 hours) and less time watching television (-4 hours) than people who did not participate. Stone et al. (2022a; under review) report substantial differences between those who participated in an EMA study and those who did not, among panellists in the Understanding America Study (UAS), a probability-based Internet panel in the U.S.. Compared with participants, those who did not participate were more likely to be older, male, to have less education and lower income, not working, have less confidence in their computer skills, and have worse self-reported health. They characterize many of the effects as of “substantial magnitude.” In sum there is little evidence to date on the biases introduced at different stages of participation in mobile app-based studies.

3 Research design and methods

3.1 Sample

We used the *Understanding Society* Innovation Panel (IP), a probability sample of households in Great Britain that interviews all household members aged 16+ annually (University of Essex, 2021). In earlier research (Jäckle et al. 2022a), we found that inviting panel members in the annual interview to download and use an app resulted in higher compliance rates than making a separate postal request following the interview. Given this, we embedded the invitation to the app study in the wave 13 annual interview (IP13).

The IP is part of *Understanding Society*: The UK Household Longitudinal Study, but fielded as a separate sample and used for methods testing and experimentation. The IP is a clustered and stratified sample selected from postal addresses, with an initial respondent sample of about 1,500 households and refreshment samples of about 500 respondent households added in waves 4, 7, 10 and 11. IP13 was conducted as a sequential mixed mode survey, with all sample households issued to web and non-respondent individuals followed up by CATI interviewers.[2] All IP13 respondents who had been interviewed at least once in previous years were eligible for the app study (n=2,152). About one-fifth (n=406, 18.9%) of cases eligible for the app study were interviewed by telephone. The household response rate for the IP13 survey was 63.6% with 78.6% of enumerated adults within those households completing the individual interviews. For further details on the survey design and fieldwork see the IP User Guide.[3].

3.2 The Understanding Wellbeing App

The “Understanding wellbeing” app was created and implemented by Connect Internet Solutions Ltd for the purposes of this study. The app was compatible with iOS and Android smartphones and tablets. Respondents were asked to download the app, either from the App Store (Apple) or from the Google Play Store (Android), and use it every evening for 14 days to answer questions about their experiences, relationships, emotions and wellbeing that day. Respondents were promised £1 for every day on which they completed the questionnaire. The invitation included a personal username and password for the respondent. They were then asked whether they had succeeded in downloading the app and if not, why not. All respondents, except those who said they did not want to download the app, also received an email reminder with their username and password (if their email address was known), alongside more detailed download information and an email address for any further queries. The wording of the invitation to the app and follow-up questions are presented in Appendix A.

---

[2] Since IP5 (2012), the Innovation Panel has used a mixed mode design with adults in two-thirds of households being issued web-first, with non-responders followed-up by CAPI interviewers, and one-third CAPI-first (with non-responders invited to complete online). However, IP13 was conducted during the COVID-19 pandemic, and face-to-face fieldwork was suspended. During this time, non-responders were followed up by telephone.
[3] https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide

The first time the respondent logged into the app they were shown a series of introductory pages summarising what the respondent was asked to do in the app (see Appendix B for the wording). The app dashboard included buttons linking to the daily survey, FAQs, and a summary of the rewards earned to date. The app included three questionnaires: a module of background questions that was only asked on the first day, the main questionnaire that was asked in the same way every day, and a debrief module that was only asked on day 14. The questionnaire was available between 5pm and 2am every evening. The app sent a ping notification at 5pm every day and another one at 9pm if the respondent had not yet completed the questionnaire for that day. See the wellbeing app study User Guide for more details on the app design and implementation (Jäckle et al. 2023).

The data from the annual IP interviews (University of Essex 2021) and the Wellbeing app study (University of Essex 2023) are available from the UK Data Service. The wellbeing app User Guide and questionnaire are also available from the UK Data Service (SN 9065).

3.3 Experimental design

The app study included three experiments with study protocols: bonus incentives, the length of the daily questionnaire, and the position of the invitation to the app study in the annual interview. The treatments for the first two experiments were flagged to the respondent in the invitation to the app and again in the introductory pages displayed the first time the app was opened. All randomised allocations were done at the household level, so that all members of a household received the same treatments. The allocations were split into equally sized groups and stratified to ensure that the allocations for each experiment are balanced across the treatments for the other experiments.

Bonus incentives: In addition to the £1 per completed daily questionnaire, a sub-set of respondents were promised a bonus if they completed all 14 days. The delivery of that bonus was however varied. Group 1 received no bonus incentive, group 2 were promised a £10 bonus if they completed all 14 days, and group 3 were promised £2.50 on four randomly selected days, if they completed the questionnaire on that day. The third treatment added an element of gamification, as the bonus incentive would come as a surprise, although the total achievable bonus of £10 was the same as in group 2.

Length of the daily questionnaire: The background and debrief questionnaires were the same for all sample members but the length of the daily questionnaire was varied. For group 1 the length of the daily questionnaire was about 2 minutes, for group 2 it was about 10 minutes. The short questionnaire was a subset of questions from the longer questionnaire.

Position of the invitation to the app study in the annual interview: The individual annual interview is on average about 45 minutes long. Group 1 were invited to the wellbeing app study early in the interview (after short modules on COVID-19, demographics, and mobile device use) and group 2 were invited to the app study at the end of the interview.

The treatment groups for the three experiments are balanced in terms of respondent age, sex, education, whether in paid work in the last week, marital status, whether has children younger than 16, and self-reported health with one exception: the late invitation group has a smaller proportion of respondents aged 51-70 and a higher proportion of respondents aged 71+ (difference of 4.2 percentage points) than the early invitation group (Pearson $\chi 2$ (5) = 10.2941, P = 0.067).

3.4 Outcomes and covariates used in analyses

As with many multi-step processes like the task of downloading and using an app to answer daily questions, there are several outcomes of interest:

- Do they have a compatible mobile device?
- Did they agree to download and use the app?
- Did they use the app at least once?[4]
- How much (how often) did they use the app?

We do not examine the first of these in this paper, as all eligible respondents were invited, regardless of whether they had a compatible smartphone or not. In fact, 1,959 of the eligible 2,152 respondents (91.0%) reported having either a smartphone or a tablet or both. Our analyses therefore focus on the remaining three outcomes.

[4] We cannot identify those who downloaded and installed the app but did not complete any surveys, so use completion of at least one daily survey as evidence of successful installation of the app.

The analysis of participation bias uses data on respondent characteristics collected in the wave 13 annual interview. We use a combination of socio-demographic characteristics, health behaviours and outcomes that are directly related to some of the topics measured in the wellbeing app questionnaire, and mobile device use behaviours that are related to experience with using apps. The wording of the corresponding survey questions is documented in the wave 13 IP questionnaire.[5] The item missing rates for these variables range from 0 to 1%; missing observations are therefore set to modal values:

- Age coded in bands: 16-30, 31-40, 41-50, 51-60, 61-70, 71+.
- Sex: male, female.
- Highest educational qualification: degree, A-level or equivalent, GCSE or lower.
- In paid work in the last seven days: yes, no.
- Marital status: single never married, married or in civil partnership, separated or divorced or widowed.
- Has children aged below 16: yes, no.
- Self-rated health: excellent, very good, good, fair or poor.
- Longstanding illness or disability: yes, no.
- Mental health: low or high distress, based on median split of summed score of answers to 12 questions from the General Health Questionnaire (GHQ).
- Smokes cigarettes: yes, no.
- Number of doctor visits in the last 12 months: none, one or two, three or more.
- Mobile device use: smartphone and tablet, smartphone only, tablet only, neither.
- Frequency of using apps: every day, several times a week, several times a month, once a month or less, never.
- Uses health apps: yes, no.

4 Results

We first focus on the overall participation rates (RQ1) before examining the effect of the experimental manipulations (RQ2). All analyses are unweighted.

[5] https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires

RQ1: What are the overall rates of participation or compliance at each stage of the app study?

The first column of Table 1 shows that 815 of the 2,152 eligible respondents (37.9%) reported successfully downloading and logging in to the app during the IP13 interview. A similar percentage (37.6%) reported that they did not want to download the app. Those who were unwilling or unable to download the app at the time of the initial request were sent a follow-up email invitation, unless they said they did not want to participate in the app study or we did not have a valid email address. The original request plus the additional reminders resulted in a total of 960 eligible respondents (44.6%) using the app to record some data (bottom of column 2), which is higher than the 37.9% who originally reported downloading the data during the survey. In the last two columns of the table we present the percentages of respondents using the app by what they said they would do during the interview. Here we see that 93.3% of those who said they successfully downloaded the app and logged in at the time of the interview indeed had data recorded in the app. About a third of those who reported trying but failing to download the app (39.3%) and a similar proportion of those who reported not trying (38.4%) went on to download and use the app. This suggests that encouraging respondents to take action right away, but still following up those who were unable or reluctant to install the app at the time, is a useful strategy.

Table 1: Actual use of the app by action or intention reported during the IP survey

Reported action at the time of the IP survey	Overall (column %)	Actual use of app (row %)	
		Used app	Did not use app
Successfully downloaded & logged in	37.9% (815)	93.3% (760)	6.7% (55)
Tried but failed	6.3% (135)	39.3% (53)	60.7% (82)
Did not try	16.0% (344)	38.4% (132)	61.6% (212)
Did not want to	37.6% (809)	0.5% (4)	99.5% (805)
Don’t know, refused, no answer	2.3% (49)	22.4% (11)	77.6% (38)
Total (row %)	(2,152)	44.6% (960)	55.4% (1,192)

Table 2: Patterns of participation across the 14 days of the study

Pattern of participation across days	N	%	Cum. %	Example pattern
All 14 days completed	238	24.8	24.8	11111111111111
1 day missing	192	20.0	44.8	1111111.111111
Multiple 1 day gaps	225	23.4	68.2	111.11111.1111
Multi-day gap(s), last day=14	132	13.8	82.0	1.11111....111
Multi-day gap(s), last day=11,12,13	86	9.0	90.9	1.1111...1.1..
Multi-day gap(s), last day<=10	87	9.1	100.0	1..11.........
Total	960	100.0		

Notes: in the Example pattern, "1" denotes a study day the respondent completed, "." denotes a missing day.

Examining participation across the 14 study days (Table 2), we see that 24.8% of respondents who used the app at least once used it on all 14 days. A further 20.0% used the app on 13 of the 14 study days, and a further 23.4% had multiple gaps, but only ever missed one day at a time. That is, overall, 68.2% of participants provided complete or relatively complete data across the study period. There are a further 13.8% who had one or more gaps of multiple days but still used the app on day 14, as well as 9.0% who had longer gaps but still used the app on days 11, 12 or 13. Only 9.1% clearly dropped out of the study and did not use the app at all after day 10.

RQ2: Which features of the request increase participation in a daily app study?

Next we look at the effect of the experimental manipulations on participation. We test for effects on whether respondents installed the app, used it at least once, or used it on all 14 days, and then take a closer look at participation across the 14 days. Table 3 shows the outcomes by each of the experimental manipulations in turn. Overall, 960 eligible respondents (or 44.6%) completed at least one of the daily surveys, while 238 (or 11.1%) fully adhered to the protocol, completing all 14 days. The early invitation had a significant positive effect on the number of participants installing the app and completing at least one day. While the £10 bonus and the shorter daily questionnaire significantly increased the percentage of respondents who installed the app, these treatments had no effect on whether respondents actually used the app. None of the experimental manipulations significantly affected the percentage who completed all 14 study days.

Table 3: Effect of experiments on use of the app

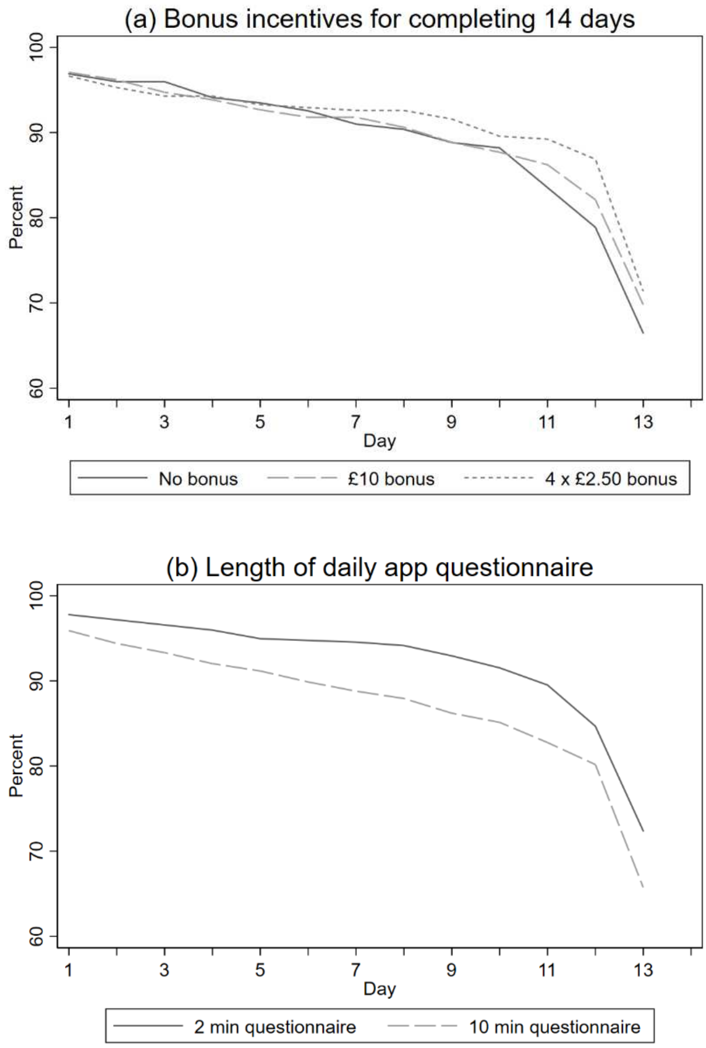
Experiments	(n)	Reported successfully downloading app in IP survey	Completed at least 1 daily survey	Completed all 14 days
Overall	(2,152)	37.9%	44.6%	11.1%
Incentive				
No bonus	(740)	36.1%	43.5%	9.6%
£10 bonus	(731)	41.6%	46.6%	11.1%
£2.50 x 4 bonus	(681)	35.8%	43.6%	12.6%
		$X^2(2)=6.503$, p=.039	$X^2(2)=1.86$, p=.394	$X^2(2)=3.32$, p=.190
Length of daily survey				
2 minutes	(1066)	40.9%	46.5%	11.2%
10 minutes	(1086)	34.9%	42.7%	11.0%
		$X^2(1)=8.24$, p=.0041	$X^2(1)=3.15$, p=.076	$X^2(1)=0.023$, p=.879
Position of invitation				
Early	(1098)	43.5%	48.6%	11.2%
Late	(1054)	32.0%	40.4%	10.9%
		$X^2(1)=30.54$, p=<.001	$X^2(1)=14.69$, p<.001	$X^2(1)=0.046$, p=.829

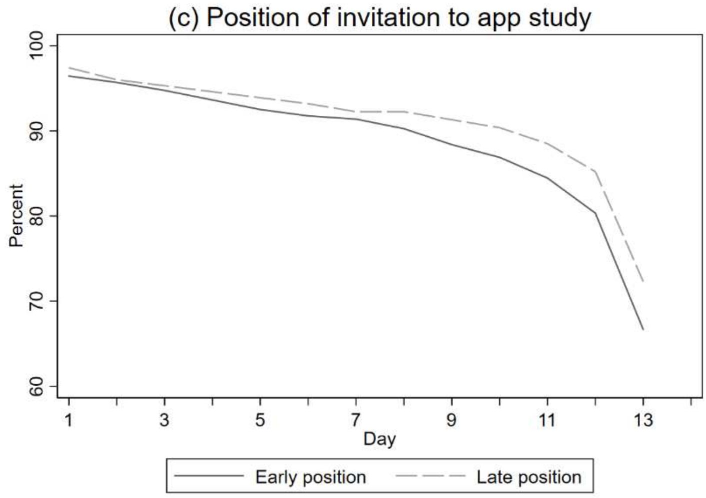
The results reported in Table 3 are not influenced by the mode in which respondents completed the IP13 interview (results not shown). Examining all three outcomes (whether downloaded the app, used it at least once, or used it on all 14 days), there are no significant interactions between any of the three treatment allocation variables and whether respondents completed the IP13 interview on the web or by telephone.

To examine the effects of the experimental treatments on participation over the 14 study days, we examine the survivor functions of all treatment groups. For each study day, the survivor function shows the percentage of participants who continued in the study and used the app on at least one later day. The incentive experiment did not have a significant effect on retention. Although there are some differences in Figure 1a, with 89% of participants in the 4 x £2.50 bonus group continuing in the study after day 11 compared to 86% of participants in the £10 bonus group and 84% in the no bonus group, the differences between treatment groups are not significant (log rank test for equality of survivor functions, Chi(2) = 2.40, P = 0.301). The short (2 minute) daily questionnaire however led to better retention than the longer (10 minute) daily questionnaire: 90% in the short group continued in the study past day 11, compared to 83% in the long group (log rank test, Chi(1) = 5.76, P = 0.016). In contrast, those invited to the app study early in the annual interview were less likely to remain in the study (84% continued past day 11) than those invited at the end of the survey (88% continued, log rank test Chi2(1) = 3.87, P = 0.049).

Appendix Figure 1 shows the daily participation rates for participants who used the app at least once, by experiment and treatment group. The differences between groups mirror those of the survivor functions.

Figure 1: Effect of experimental treatments on staying in the study (Survivor functions)





RQ3: What are the barriers to participation in an app-based study?

In this section we examine two types of barriers to participation: (1) reasons for not wanting to download the app reported by those who did not try, and 2) problems reported by respondents who attempted to download and log into the app.

Reasons for not wanting to download the app

After asking respondents in the interview whether they had managed to download the app, those who said *"I do not want to download the app"* were asked *"To help us plan future studies, could you explain why you did not try to download and log in to the app? Please select all that apply."* The most frequent reasons mentioned (Table 4) were lack of interest or time: 58.8% of respondents selected at least one of the options *"I don't want to participate in additional survey tasks"* (28.7%), *"Do not have time to take part"* (17.2%), or "*Not interested in answering additional questions on this topic"* (12.7%). Technical or ability related reasons were less prevalent: 32.1% of respondents selected at least one of the options *"No smartphone or tablet which can download apps"* (12.4*%), "Not able or confident to download apps onto my smartphone or tablet"* (9.4%), *"Do not want to take up storage space on my smartphone or tablet"* (7.4%), or *"No Internet access"* (5.0%). Privacy and/or data security concerns were reported by 11.0% of respondents: 8.5% said they were *"Not willing to share this kind of information"* and 2.3% that they were *"Not confident that information would be held securely"*.

Table 4: Reasons for not trying to install and log in to the app

Reasons for not installing app	N mentions	% of respondents
I don't want to participate in additional survey tasks	261	28.7
Do not have time to take part	156	17.2
Not interested in answering additional questions on this topic	115	12.7
No smartphone or tablet which can download apps	113	12.4
Not able or confident to download apps onto my smartphone or tablet	85	9.4
Not willing to share this kind of information	77	8.5
Do not want to take up storage space on my smartphone or tablet	67	7.4
No Internet access	45	5.0
Not confident that information would be held securely	21	2.3
Other reason	76	8.4

Notes: Based on 809 respondents. Respondents would select more than one response option, therefore the column percentages add up to more than 100%.

There are few differences between the experimental treatment groups in the reasons respondents gave for not installing the app. For each of the nine potential reasons respondents could give, we computed a $\chi 2$ test for each of the three randomised experiments, of whether the percentage of respondents who selected that reason was different between the treatment groups. Of the 27 $\chi 2$ tests only two are statistically significant. In the 2-minute questionnaire length group 4.3% said they were not confident that the information would be held securely, compared to 1.1% in the 10-minute questionnaire group (Pearson $\chi 2(1) = 7.9887$, $P = 0.005$). In the treatment group that were offered no bonus for completing the app study each day 14% said they were not willing to share this kind of information, compared to 7.3% in the £10 bonus group and 6.9% in the randomised bonus group (Pearson $\chi 2(2) = 10.4475$, $P = 0.005$).

Examining the reasons for not installing the app by mode and device with which the respondent completed the IP13 interview shows that respondents who completed the survey with a telephone interviewer were most likely to say that they did not have a smartphone or tablet which can download apps (35.1%), followed by not wanting to participate in additional survey tasks (25.4%). Respondents who completed the survey online were most likely to say that they did not want to participate in additional survey tasks (35.0% of respondents who completed the survey on a PC and 38.9% of those who completed on a tablet) or that they did not have time (32.0% of those who completed the survey on a smartphone).

Problems with installing and logging into the app

Respondents who reported they had *"Successfully downloaded and logged into the app"* were asked an open-ended question "*Did you have any difficulties installing and logging into the app?"* Similarly, those who reported they had *"Tried to download the app but did not succeed"* were asked *"What difficulties did you have installing and logging into the app?"* Table 5 shows the distributions of the coded responses to these open questions, by whether the respondent had successfully installed the app or tried but failed.

Table 5: Problems installing the app, by whether respondent managed or failed to install it

Problems installing app	Successfully downloaded and logged into app		Tried to download but did not succeed	
	N	%	N	%
Don't know/refused	22	2.7	5	3.7
No problem	694	85.2	2	1.5
Couldn't find it	35	4.3	56	41.5
Log-in errors	14	1.7	20	14.8
Device not compatible	0	0.0	14	10.4
Don't know how to do it	2	0.2	9	6.7
Couldn't type username/password	9	1.1	9	6.7
App did not download	0	0.0	6	4.4
Take up too much space	7	0.9	5	3.7
App froze when trying to log in	1	0.1	2	1.5
App very slow	13	1.6	2	1.5
Other	18	2.2	5	3.7
Total	815	100.0	135	100.0

Respondents who managed to install and login to the app were most likely to say they had not had any problems (85.2%), although some said they could not find the app (4.3%).

Respondents who tried but failed to install the app were most likely to say they could not find the app (41.5%), followed by log-in errors (14.8%), their device not being compatible (10.4%), not knowing how to do it (6.7%), or problems with typing the username or password (6.7%). Note that in this study we told respondents the name of the app and asked them to search for it in their app store. Including a direct link or a QR code leading directly to the app in the relevant store would presumably reduce the problems respondents have with finding the app.

There are no significant differences between experimental treatment groups in the problems reported with installing the app (early versus late position of the invitation to the app study in the questionnaire, incentive treatment, and the length of the daily app questionnaire).

Examining the problems reported with installing the app by mode and device with which the respondent completed the IP13 interview shows that respondents who used a smartphone to complete the interview online were most likely to report that they did not have any problems installing the app (85.6%), followed by respondents who used a PC (including laptops and netbooks, 73.1%), and respondents who used a tablet (69.4%). Of respondents who completed the survey with a telephone interviewer fewer (60.2%) reported no problems with installing the app.

RQ4: What selection biases (if any) are introduced at different stages of the app study? Do study protocols affect biases?

Table 6 documents the selection biases that result from restricting the sample to different subsets: those who downloaded and logged into the app during the IP13 interview, those who used the app at least once, and those who used the app on all 14 days. Following the approach by Couper et al. (2018), the first column shows the composition of the full sample. Note that this is conditional on completing the IP13 interview – we do not account for nonresponse or attrition in the Innovation Panel itself. The bias columns show the difference in an estimate *y* between the selected sub-sample (s) and the full sample (f):

$$bias(y) = y_s - y_f$$

We calculate the standard error of the estimated bias following Lee (2006) as:

$$se(y_s - y_f) = \frac{n_f - n_s}{n_f}\sqrt{var(y_s) + var(y_{ns})}$$

Where y_{ns} is the estimate for the sub-sample that is not selected. We then test the significance of the bias estimates using large sample z-tests, dividing each estimated bias by its standard error. Statistically significant biases ($p<0.05$) are highlighted in bold in Table 6. As an example, 13.7% of the full sample are in the 16-30 age group. Among the sub-sample who used the app at least once 17.0% are in this age group, resulting in a bias of 3.3 percentage points. For variables with two categories the bias estimates are symmetrical. For example, in the sub-sample who used the app at least once, women are over-represented by 3.2 percentage points and men under-represented by the same amount. We therefore only include one of the two categories in the table. For variables with more than two categories all categories are included. To compare the extent of biases across the sub-samples, the final row in Table 6 presents the mean absolute bias across all estimates: for each sub-sample we sum the absolute values of all biases and average them.

Overall, the mean absolute bias is lowest for the sub-group that used the app at least once (4.6), followed by the sub-group who used the app on all 14 days (4.9), with the largest overall bias in the sub-group who installed the app during the IP13 interview (5.2). These suggest that the biases are highest at the initial compliance, but do not compound with each successive stage of the protocol (although the variances increase as the sample sizes decline).

For most characteristics the patterns of biases are consistent across the three sub-groups, with younger age groups, those with a degree, those in work, and those who use health apps over-represented, and those who are divorced, separated or widowed, those with a long-standing illness or disability, those who use only a tablet or neither a tablet nor a smartphone, and those who never install and use mobile apps under-represented in all sub-groups. There are however some characteristics where biases are only significant for some of the sub-groups: women, those who have children below the age of 16, those with excellent or very good health, and those with above median levels of mental distress are only significantly over-represented in the first two sub-samples but not in the sub-sample who used the app on all 14 days. Smokers are only significantly under-represented in the third sub-sample. The only characteristic for which there is no significant bias in any of the categories in any of the sub-groups, is the frequency of visiting a doctor in the last 12 months.

The largest biases are in the characteristics related to mobile device use. Using the sub-group who used the app at least once as an example, those who use health apps are over-represented by 15.9 percentage points, those who use apps every day are over-represented by 14.3 percentage points, and those who use both a smartphone and a tablet are over-represented by 10.7 percentage points. The next largest biases are related to socio-demographic characteristics: those in the 71+ age group are under-represented by 11.0 percentage points, those not in work by 10.9 percentage points, and those with GCSEs, lower or no educational qualification by 7.6 percentage points. The biases related to health outcomes and behaviours are smaller in size, with those with fair or poor health under-represented by 3.4 percentage points, and those with a disability by 3.2 percentage points.

Table 6: Bias at different stages of participation

Variable	Level	Full sample	Downloaded app		Used app		Used app 14 days	
		%	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)
Age	16-30	13.7	**4.9**	**(0.778)**	**3.3**	**(0.678)**	**6.0**	**(2.263)**
	31-40	12.1	**5.0**	**(0.712)**	**4.5**	**(0.560)**	**5.9**	**(2.170)**
	41-50	15.9	**5.3**	**(0.841)**	**5.1**	**(0.664)**	0.9	(2.264)
	51-60	20.1	2.0	(1.052)	**1.8**	**(0.898)**	-2.8	(2.367)
	61-70	19.8	**-5.3**	**(1.220)**	**-3.8**	**(1.084)**	0.0	(2.435)
	71+	18.4	**-12.0**	**(1.276)**	**-11.0**	**(1.264)**	**-10.0**	**(1.912)**
Sex	Female	55.7	**2.6**	**(1.328)**	**3.2**	**(1.135)**	4.8	(2.962)
Education	GCSEs, other, none	35.3	**-9.0**	**(1.468)**	**-7.6**	**(1.319)**	**-8.8**	**(2.810)**
	A-level	19.8	0.3	(1.092)	-0.1	(0.961)	2.9	(2.499)
	Degree	44.9	**8.7**	**(1.206)**	**7.7**	**(1.023)**	**5.9**	**(3.004)**
In work	Yes	53.2	**12.7**	**(1.140)**	**10.9**	**(0.979)**	**9.4**	**(2.903)**
Marital status	Single, never married	25.6	**4.1**	**(1.106)**	1.4	(1.006)	1.7	(2.694)
	Married, civil partnership	56.6	0.3	(1.364)	**2.6**	**(1.143)**	2.7	(2.989)
	Separated, divorced, widowed	17.8	**-4.3**	**(1.162)**	**-4.0**	**(1.062)**	**-4.4**	**(2.183)**
Has kids <16	Yes	13.1	**2.6**	**(0.844)**	**3.4**	**(0.654)**	2.9	(2.158)
Health	Excellent	10.4	**1.9**	**(0.773)**	**2.0**	**(0.630)**	2.2	(1.957)
	Very good	38.2	**3.1**	**(1.281)**	**2.5**	**(1.108)**	3.8	(2.975)
	Good	34.3	-1.5	(1.339)	-1.1	(1.168)	-2.0	(2.886)
	Fair/poor	17.0	**-3.5**	**(1.126)**	**-3.4**	**(1.026)**	-4.0	(2.149)
Disability	Yes	29.7	**-3.2**	**(1.325)**	**-3.2**	**(1.184)**	**-5.3**	**(2.705)**
Mental health	Above median distress	54.9	**3.3**	**(0.013)**	**2.3**	**(0.012)**	-0.3	(0.030)
Smoker	Yes	11.1	-1.4	(0.914)	-1.2	(0.813)	**-4.4**	**(1.662)**

Doctor visits in last 12 months	None	31.4	-0.6	(1.294)	-0.2	(1.119)	-1.6	(2.819)
	One or two	42.8	1.6	(1.337)	0.6	(1.175)	5.5	(3.002)
	Three or more	25.8	-1.0	(1.230)	-0.5	(1.065)	-3.9	(2.593)
Mobile device use	Smartphone and tablet	57.2	**9.6**	**(1.195)**	**10.7**	**(0.984)**	**10.9**	**(2.803)**
	Smartphone only	26.9	**2.5**	**(1.167)**	1.0	(1.036)	0.0	(2.710)
	Tablet only	7.3	**-4.6**	**(0.860)**	**-4.5**	**(0.868)**	**-3.1**	**(1.344)**
	Neither	8.7	**-7.5**	**(0.950)**	**-7.1**	**(1.006)**	**-7.8**	**(0.888)**
Frequency of using apps	Every day	38.8	**15.8**	**(0.999)**	**14.3**	**(0.804)**	**11.7**	**(2.932)**
	Several times a week	15.3	**2.6**	**(0.915)**	**2.5**	**(0.761)**	2.3	(2.271)
	Several times a month	9.4	1.3	(0.755)	0.5	(0.673)	0.7	(1.817)
	Once a month or less	13.0	-1.4	(0.970)	-0.5	(0.831)	0.9	(2.089)
	Never	23.5	**-18.4**	**(1.389)**	**-16.7**	**(1.409)**	**-15.5**	**(1.943)**
Uses health apps	Yes	39.5	**18.5**	**(0.942)**	**15.9**	**(0.769)**	**18.1**	**(2.842)**
	Mean absolute bias		5.2		4.6		4.9	
	Total (N)	2,152	815		960		238	

Note: Statistically significant bias estimates (p < .05) are in bold.

To examine whether the protocols for implementing the app study affect selection biases, we replicate the bias estimates for the sub-sample who used the app at least once, by experimental treatment group (Appendix Table 1). While there are some differences between treatment groups in the extent of participation bias, there are no clear patterns. Overall, the mean absolute bias is lower in the 2 minute questionnaire group than the 10 minute questionnaire group (4.1 compared to 5.2), somewhat lower in the £2.50 x 4 incentive group than the no bonus or £10 bonus groups (4.4 compared to 4.8 and 5.2) and there is little difference between the early and late invitation groups (4.8 and 4.6).

We test for differences in bias between experimental treatment groups by testing for interactions. We estimate a separate logit model for each respondent characteristic and each treatment allocation variable: we regress the probability that a respondent used the app at least once on the treatment allocation for the experiment, the respondent characteristic, and their interaction. The cells shadowed in grey in Appendix Table 1 indicate statistically significant interactions ($p<0.05$). In the experiment with the position of the invitation to the app, the biases regarding mobile device use and use of health apps are significantly different between the 'early' and 'late' invitation groups. There is however no clear pattern that biases are lower in one treatment group than the other. In the questionnaire length experiment there are also two interactions that are significant. In the 2 minute group women are not over-represented among participants, while they are over-represented by 5.8 percentage points in the 10 minute group. Those who use health apps are somewhat less over-represented in the 2 minute questionnaire group (13.0 percentage points) than in the 10 minute questionnaire group (18.8 percentage points). These differences contribute to somewhat lower mean absolute bias in the 2 minute group than the 10 minute group. In the incentive experiment, the biases regarding marital status and whether the respondent has a long-standing illness or disability are significantly different, however there are no systematic differences between treatment groups in the extent of bias.

5 Discussion

Intensive measurement studies – particularly those exploiting technologies like smartphone apps – are transitioning from small-scale studies among cooperative volunteers to inclusion in large-scale probability samples of the general population. This transition raises questions about maximizing participation and minimizing participation bias in such studies, and points to the need to develop protocols for increasing participation and reducing differential participation, potentially affecting inference from these samples of willing participants.

Our study demonstrated the substantial sample loss suffered by studies such as this: fewer than half of those invited (44.6%) downloaded and used the app at least once, and only 11.1% fully complied with the protocol, answering all 14 days of the daily app survey. Examining the patterns of participation across the 14 study days, however, shows that although most participants skipped some days, only a minority actually dropped out. This suggests that data from intensive measurement studies are ideally analysed using statistical methods that can deal with differing numbers of observations across respondent.

We tested three interventions to increase participation in the app-based daily diary survey. We found that offering bonus incentives conditional on completing all 14 study days had little effect on the outcomes we examined. The lack of incentive effect mirrors the mixed results in the literature, suggesting that the effects of incentives are not consistent, and may depend on the features of the incentive offered.

The length of the daily questionnaire did not affect whether respondents used the app, but did affect retention: Participants in the short questionnaire group were less likely to drop out than those in the long questionnaire group.

The timing of the request had a significant effect on uptake, with 48.6% of those asked earlier in the IP13 interview downloading and using the app at least once, compared with 40.4% of those asked at the end of the IP13 interview. This finding is consistent with the literature cited earlier that asking for consent at the beginning of the survey is associated with higher rates of consent than asking it later. Our study is the first to find that this effect may extend to other types of requests. However, this did not translate into higher rates of full adherence to the protocol (completing all 14 days of the app study). In fact, participants in the early invitation group were more likely to drop out of the study than those in the late invitation group.

Our follow-up questions asked of those who were unwilling to participate in the app study or who reported unsuccessfully trying to download the app help identify important barriers to participation. Finding ways to overcome these hurdles is an important next step. While some of the technical barriers (like failing to find the app or problems logging in) could be overcome more easily, allaying concerns about the time commitment or overcoming a lack of interest could take more work. For example, the offer of feedback may be motivating for some, but of little interest to others (see Wenz et al. 2020).

Sample loss is only part of the story. If those lost do not differ from those who fully participated, then smaller sample sizes only lessen statistical power. However, if participants differ from non-participants on key variables of interest, the resulting bias may affect inferences drawn from studies such as this. Our finding that there is substantial bias related to access and use of technology – that those who fully participated in the app study are much more tech-savvy than those who dropped out at various points – raises concerns about selection bias. While the biases are smaller for variables related to health, the sample of app participants is generally more healthy than the full IP13 population, suggesting the potential for bias on outcomes related to health and wellbeing. While some biases (e.g., socio-demographic differences) can be accounted for with weighting, overcoming the digital divide is more challenging, and may undermine some of the benefits of app-based data collection. That is, offering an alternative mode (such as paper diaries) or providing smartphones to those without access may bring more people in, but come at the cost of efficiency of data collection and the other benefits that technology brings (such as in-app reminders or prompts and the speed of completion).

Our study has several limitations. First, we test only one type of intensive measurement, a daily app study over 14 days. As a case study, our results may not generalize to other types of additional requests made of survey participants. Second, the experimental features were limited to those that would be practical and affordable in studies such as this. That is, we did not maximize the differences in incentive amounts to produce a significant effect; rather, we chose incentives and survey length manipulations that were reasonable. This may have diminished the possibility of detecting significant differences, but increases the practical utility of our findings. Third, our design deliberately included all IP13 panellists, rather than restricting the sample to those known to have a compatible mobile device. While imposing such a restriction this may have resulted in higher participation rates, prior reports of ownership of or access to suitable device might not capture the temporal change in technology use. This also allows us to explore lack of access as a reason for non-participation, rather than taking it as given. Finally, our study is embedded in an ongoing longitudinal study. On the one hand, this means that we are dealing with a set of relatively engaged panellists who are familiar with the study and may be more positively disposed to such requests. That is, we consider participation in the app study conditional on completing the IP13 interview. Wave nonresponse or attrition in the Innovation Panel may introduce additional biases relative to the broader population. It also means we have a set of covariates to explore the conditional selection bias, and potentially correct for those we find. On the other hand, these same panellists may feel that that are giving enough to the study, and are disinclined to give more. That is, our results may not generalize to cross-sectional surveys, but are relevant to panel studies where the richness of longitudinal data already collected on respondents can be further enhanced with these ancillary studies.

In summary, these findings point to the importance of developing methods to increase participation in intensive measurement studies and, in particular, to paying attention to differences in participation across key subgroups of the population. Finding ways to increase participation of under-represented groups in these ancillary studies is important in maximizing the value of the additional data collected.

References

Axén, I., & Bodin, L. (2016). Searching for the optimal measuring frequency in longitudinal studies -- an example utilizing short message service (SMS) to collect repeated measures among patients with low back pain. *BMC Medical Research Methodology*, 16: 119, https://doi.org/10.1186/s12874-016-0221-4.

Bandarian-Balooch, S., Martin, P.R., McNally, B., Brunelli, A., & Mackenzie, S. (2017). Electronic-diary for recording headaches, triggers, and medication use: Development and evaluation. *Headache: The Journal of Head and Face Pain*, 57: 1551-1569, https://doi.org/10.1111/head.13184.

Bissonnette, J.M. (2008). Adherence: A concept analysis. *Journal of Advanced Nursing*, 63: 634-643, https://doi.org/10.1111/j.1365-2648.2008.04745.x.

Bolger, N., Davis, A., & Rafaeli, E. (2003), Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54 (1): 579-616, https://doi.org/10.1146/annurev.psych.54.101601.145030.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. The Guilford Press, New York.

Bond, D.M., Hammond, J., Shand, A.W., & Nassar, N. (2020). Comparing a mobile phone automated system with a paper and email data collection system: Substudy within a randomized controlled trial. *JMIR Mhealth and Uhealth,* 8(8):e15284, https://doi.org/10.2196/15284.

Broderick, J.E., & Stone, A.A. (2006). Paper and electronic diaries: Too early for conclusions on compliance rates and their effects--Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006). *Psychological Methods*, 11(1): 106-111, https://doi.org/10.1037/1082-989X.11.1.106.

Burton, J., Jäckle, A., & Couper, M.P. (2022). How to maximise informed consent to data linkage? The effects of placement, context, and order of consent questions. Presented at the Panel Survey Methods Workshop, Lausanne (July 2022).

Couper, M.P., Gremel, G., Axinn, W., Guyer, H., Wagner, J.N., & West, B.T. (2018). New options for national population surveys: The implications of internet and smartphone coverage. *Social Science Research*, 73: 221-235, https://doi.org/10.1016/j.ssresearch.2018.03.008.

Courvoisier, D.S., Eid, M., & Lischetzke, T. (2012). Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. *Psychological Assessment*, 24(3), 713–720, https://doi.org/10.1037/a0026733.

Courvoisier, D.S., Eid, M., Lischetzke, T., & Schreiber, W.H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, 10(1), 115–124, https://doi.org/10.1037/a0017813.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method*. Journal of Nervous and Mental Disease*, 175 (9): 526–536, https://doi.org/10.1007/978-94-017-9088-8_3.

Daniëls, N.E.M., Hochstenbach, L.M.J., van Zelst, C., van Bokhoven, M.A., Delespaul, P.A.E.G., & Beurskens, A.J.H.M. (2021). Factors that influence the use of electronic diaries in health care: Scoping review. *JMIR Mhealth and Uhealth*, 9 (6): e19536, https://mhealth.jmir.org/2021/6/e19536.

De las Cuevas, C. (2011). Towards a clarification of terminology in medicine taking behavior: Compliance, adherence and concordance are related although different terms with different uses. *Current Clinical Pharmacology*, 6(2): 74-77, https://doi.org/10.2174/157488411796151110.

De Vries, L.P., Baselmans, B.M.L., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies*. Journal of Happiness Studies,* 22: 2361-2408, https://doi.org/10.1007/s10902-020-00324-7.

Diener, E., & Tay, L. (2014). Review of the day reconstruction method (DRM). *Social Indicators Research*, 116 (1): 255-267, https://www.jstor.org/stable/24720529.

Duif, M., Thewissen, V., Wouters, S., Lechner. L., & Jacobs, N. (2020). Associations between affect and alcohol consumption in adults: An ecological momentary assessment study. *The*

American Journal of Drug and Alcohol Abuse, 46 (1): 88-97, https://doi.org/10.1080/00952990.2019.1635606.

Elevelt, A., Lugtig, P., & Toepoel, V. (2019). Doing a time use survey on smartphones only: What factors predict nonresponse at different stages of the survey process? *Survey Research Methods*, 13 (2): 195-213, *Survey Research Methods*, 13 (2): 195-213, https://doi.org/10.18148/srm/2019.v13i2.7385.

Gellad, W.F., Thorpe, C.T., Steiner, J.F., & Voils, C.I. (2017). The myths of medication adherence. *Pharmacoepidemiology and Drug Safety*, 26 (12): 1437-1441, https://doi.org/10.1002/pds.4334.

Green, A.S., Rafaeli, E., Bolger, N., Shrout, P.E., & Reis, H.T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11 (1): 87–105, https://doi.org/10.1037/1082-989X.11.1.87.

Griffin, J., & Patrick, M.E. (2014). Nonresponse bias in a longitudinal measurement design examining substance use across the transition out of high school. *Drug and Alcohol Dependence*, 143: 232-238, https://doi.org/10.1016/j.drugalcdep.2014.07.039.

Haas, G.-C., Kreuter, F., Keusch, F., Trappmann, M. & Bähr, S. (2021). Effects of incentives in smartphone data collection. In: Hill, C.A., Biemer, P.P., Buskirk, T.D., Japec, L., Kirchner, A., Kolenikov, S. & Lyberg, L.E. (Eds.) *Big data meets survey science. a collection of innovative methods*. Hoboken: Wiley, pp. 387–414.

Hu, M., Melipillán, E.R., West, B.T., Kirlin, J.A., & Paniagua, I. (2020). Response patterns in a multi-day diary survey: Implications for adaptive survey design. *Survey Research Methods*, 14(3), 289-300, https://doi.org/10.18148/srm/2020.v14i3.7465.

Jäckle, A., Burton, J., Couper, M.P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Response rates and response biases. *Survey Research Methods*, 13 (1): 23-44, https://doi.org/10.18148/srm/2019.v1i1.7297.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., & Walzenbach, S. (2021). Understanding and improving data linkage consent in surveys. *Understanding Society Working Paper Series No. 2021-01*. Colchester: University of Essex.

Jäckle, A., Wenz, A., Burton, J., & Couper, M.P. (2022a). Increasing participation in a mobile app study: The effects of a sequential mixed-mode design and in-interview invitation. *Journal of Survey Statistics and Methodology*, 10 (4): 898–922, https://doi.org/10.1093/jssam/smac006.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., & Walzenbach, S. (2022b). Consent to data linkage: Question wording and format experiments. *Understanding Society Working Paper 2022-05,* Colchester: University of Essex.

Jäckle, A., Burton, J., Couper, M.P., Perelli-Harris, B., & Vine. J. (2023). The UK Household Longitudinal Study Wellbeing App User Guide. Colchester: Institute for Social and Economic Research, University of Essex. Available at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=9065.

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I.H.A., Wen, C.K.F., & Field, M. (2019), Compliance with ecological momentary assessment protocols in substance users: a meta-analysis, *Addiction*, 114: 609-619, https://doi.org/10.1111/add.14503.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., & Stone, A.A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306 (5702): 1776-1780, https://www.jstor.org/stable/3839780.

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, first published April 7, https://doi.org/10.1177/0049124120914924.

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., Trappmann, M., & Eckman, S. (2022). Non-participation in smartphone data collection using research apps. Journal of the Royal Statistical Society Series A, 1-21, https://doi.org/10.1111/rssa.12827.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M.P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, 83 (S1): 210-235, https://doi.org/10.1093/poq/nfz007.

Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38 (5): 533-549, https://doi.org/10.1177/0894439318816389.

Laughland, A., & Kvavilashvili, L. (2018). Should participants be left to their own devices? Comparing paper and smartphone diaries in psychological research. *Journal of Applied Research in Memory and Cognition,* 7 (4): 552-563, https://doi.org/10.1016/j.jarmac.2018.09.002.

Lee, C.M., Fairlie, A.M., Ramirez, J.J., Patrick, M.E., Luk, J.W., & Lewis, M.A. (2020). Self-fulfilling prophecies: Documentation of real-world daily alcohol expectancy effects on the experience of specific positive and negative alcohol-related consequences. *Psychology of Addictive Behaviors*, 34 (2): 327-334, https://doi.org/10.1037/adb0000537.

Levinson, C.A., Sala, M., Fewell, L., Brosof, L.C., Fournier, L., & Lenze, E.J. (2018), Meal and snack-time eating disorder cognitions predict eating disorder behaviors and vice versa in a treatment seeking sample: A mobile technology based ecological momentary assessment study. *Behaviour Research and Therapy*, 105: 36-42, https://doi.org/10.1016/j.brat.2018.03.008.

Liddle, J., Wishink, A., Springfield, L., Gustafsson, L., Ireland, D., & Silburn, P. (2017). Can smartphones measure momentary quality of life and participation? A proof of concept using experience sampling surveys with university students. *Australian Occupational Therapy*, 64: 294-304, https://doi.org/10.1111/1440-1630.12360.

Lugtig, P., Roth, K., & Schouten, B. (2022). Nonresponse analysis in a longitudinal smartphone-based travel study. *Survey Research Methods*, 16 (1): 13-27, https://doi.org/10.18148/srm/2022.v16i1.7835.

Lutfey, K.E., & Wishner, W.J. (1999). Beyond compliance is adherence. Improving the prospect of diabetes care. *Diabetes Care*, 22, 635-639, https://doi.org/10.2337/diacare.22.4.635.

May, M., Junghaenel, D.U., Ono, M., Stone, A.A., & Schneider, S. (2018). Ecological momentary assessment methodology in chronic pain research: A systematic review. *The Journal of Pain*, 19 (7): 699-716, https://doi.org/10.1016/j.jpain.2018.01.006.

McCool, D., Schouten, B., Mussmann, O., & Lugtig, P. (2021). An app-assisted travel survey in Official Statistics. Possibilities and challenges. *Journal of Official Statistics*, 37 (1): 149–170, https://doi.org/10.2478/jos-2021-0007.

Miller, S., Ainsworth, B., Yardley, L., Milton, A., Weal, M., Smith, P.W.F., & Morrison, L. (2018). Supporting systematic assessment of digital interventions: a framework for analysing and measuring usage and engagement data (AMUsED). *Journal of Medical Internet Research*, https://doi.org/10.2196/10966.

Patrick, M.E., Maggs, J.L., & Lefkowitz, E.S. (2015). Daily associations between drinking and sex among college students: A longitudinal measurement burst design. *Journal of Research on Adolescence*, 25, 377-386, https://doi.org/10.1111/jora.12135.

Phillips, M.M., Phillips, K.T., Lalonde, T.L., & Dykema, K.R. (2014). Feasibility of text messaging for ecological momentary assessment of marijuana use in college students. *Psychological Assessment*, 26(3), 947–957, https://doi.org/10.1037/a0036612.

Raffaelli, B., Mecklenburg, J., Scholler, S. et al. (2021). Primary headaches during the COVID-19 lockdown in Germany: Analysis of data from 2325 patients using an electronic headache diary. *Journal of Headache and Pain*, 22: 59, https://doi.org/10.1186/s10194-021-01273-z.

Revilla, M., Couper, M.P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, 13 (2): 223-252, https://doi.org/10.12758/mda.2018.01.

Ságvári, B., Gulyás, A., & Koltai, J. (2021). Attitudes towards participation in a passive data collection experiment. *Sensors*, 21 (18): 6085, https://doi.org/10.3390/s21186085.

Shiffman, S., Stone, A.A., & Hufford, M.R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4: 1-32, https://doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Settineri S., Frisone F., Merlo E.M., Geraci D., & Martino G. (2019). Compliance, adherence, concordance, empowerment, and self-management: Five words to manifest a relational maladjustment in diabetes. *Journal of Multidisciplinary Healthcare*, 12: 299-314, https://doi.org/10.2147/JMDH.S193752.

Stone, A.A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24 (3): 236–243, https://doi.org/10.1207/S15324796ABM2403_09.

Stone, A.A., Shiffman, S., Schwartz, J.E., Broderick, J.E., & Hufford, M.R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, 24: 182-199, https://doi.org/10.1016/S0197-2456(02)00320-3.

Stone, A.A., Schneider, S., Smyth, J.M., Junghaenel, D.U., Wen, C., Couper, M.P., & Goldstein, S. (2022a). Shedding light on participant selection bias in ecological momentary assessment (EMA) studies: Findings from an internet panel study. Under review.

Stone, A.A., Schneider, S., Smyth, J.M., Junghaenel, D.U., Couper, M.P., Wen, C., Mendez, M., Velasco, S., and Goldstein, S. (2022b). A population-based investigation of participation rate and self-selection bias in momentary data capture and survey studies. Under review.

Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2020). Understanding willingness to share smartphone-sensor data. *Public Opinion Quarterly*, 84 (3): 725-759, https://doi.org/10.1093/poq/nfaa044.

Struminskaya, B., Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., & Dolmans, R. (2021). Sharing data collected with smartphone sensors: Willingness, participation, and nonparticipation bias. *Public Opinion Quarterly*, 85 (S1): 423-462, https://doi.org/10.1093/poq/nfab025.

Tilson, H.H. (2004). Adherence or compliance? Changes in terminology. *Annals of Pharmacotherapy*, 38(1): 161-162, https://doi.org/10.1345/aph.1D207.

Thurman, S.M., Wasylyshyn, N., Roy, H., Lieberman, G., Garcia, J.O., Asturias, A., et al. (2018). Individual differences in compliance and agreement for sleep logs and wrist actigraphy: A longitudinal study of naturalistic sleep in healthy adults. *PLoS ONE* 13(1): e0191883, https://doi.org/10.1371/journal.pone.0191883.

University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: Innovation Panel, Waves 1-13, 2008-2020. [data collection]. 11th Edition. UK Data Service. SN: 6849, http://doi.org/10.5255/UKDA-SN-6849-14.

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Innovation Panel Wellbeing App Study, 2020. [data collection]. 1st Edition. UK Data Service. SN: 9065, http://doi.org/10.5255/UKDA-SN-9065-1.

van Berkel, N., Goncalves, J., Hosio, S., Sarsenbayeva, S., Velloso, E., & Kostakos, V. (2020). Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies*, 134: 1-12, https://doi.org/10.1016/j.ijhcs.2019.10.003.

Vhaduri, S., & Poellabauer, C. (2016). Human factors in the design of longitudinal smartphone-based wellness surveys. IEEE International Conference on Healthcare Informatics (ICHI), pp. 156-167, https://doi.org/10.1109/ICHI.2016.24.

Vhaduri, S., & Poellabauer, C. (2017). Design factors of longitudinal smartphone-based health surveys. *Journal of Healthcare Informatics Research*, 1: 52–91, https://doi.org/10.1007/s41666-017-0003-8.

Vine, J., Burton, J., Couper, M.P. & Jäckle, A. (2021). Measuring smartphone operating system versions in surveys: How to identify who has devices compatible with survey apps. Paper presented at the virtual General Online Research (GOR) Conference, September.

Wen, C.K.F., Schneider, S., Stone, A.A., & Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research*, 19: e132, https://doi.org/10.2196/jmir.6641.

Wenz, A., Jäckle, A., Burton, J., & Couper, M.P. (2020). The effects of personalized feedback on participation and reporting in mobile app data collection. *Social Science Computer Review*, 40 (1): 165-178, https://doi.org/10.1177/0894439320914261.

Wenz, A., Jäckle, A., & Couper, M.P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods*, 13 (1): 1-22, https://doi.org/10.18148/srm/2019.v1i1.7298.

Appendix A: Well-being App Question Wording

1) As you know Understanding Society is all about understanding the lives of people in the UK. To broaden out the research that can be done with Understanding Society data, we would like you to participate in a study that collects data about relationships and emotional wellbeing. We would like you to download the Understanding Well-being app and use it every evening for 14 days, to record how your day has been.

Answering daily questions will take about [2 minutes / 10 minutes] of your time each evening.

As a thank you for participating in this study, you will receive £1 for each day on which you complete the [questions. /, plus an additional £10 if you complete the questions on each of the 14 days /, plus you will have a chance to receive additional rewards totalling £10 if you complete the questions on each of the 14 days.]

1. Continue

2) We would like you to try this now. Please go to the App Store for Apple devices or the Google Play Store for Android devices and download the Understanding Well-being app.

Here is the Understanding Well-being logo and your login details for the app.

Please make a note of these:

User Name: [xxxx]

Password: [yyyy]

1. Continue

3) Please tell us whether you have managed to download the app?

1. Successfully downloaded and logged into the app
2. Tried to download the app but did not succeed
3. Have not yet tried to download and log into the app
4. I do not want to download the app

4) If successfully installed the app OR did not manage to install the app:

Did you have any difficulties installing and logging into the app? / What difficulties did you have installing and logging into the app?

[open text]

5) If did not want to download and log into app:

To help us plan future studies, could you explain why you did not try to download and log in to the app?

1. No Internet access
2. No smartphone or tablet which can download apps
3. Not able or confident to download apps onto my smartphone or tablet
4. Not confident that information would be held securely
5. Do not want to take up storage space on my smartphone or tablet
6. Not willing to share this kind of information
7. Not interested in answering additional questions on this topic
8. Do not have time to take part
9. I don’t want to participate in additional survey tasks
97. Other (please specify)

Appendix B: Introductory pages in the Understanding Well-Being app

{Display Understanding Society logo}

UNIVERSE: display if day = 1

Screen 1:

Thanks for downloading the Understanding Well-Being app!

We would like you to answer a set of questions about how your day has been, each evening for the next 14 days. The survey will be open each day from 5pm until 2am.

Screen 2:

Answering these daily questions will take about {if ff_applengthw13=1: 2 minutes / if ff_applengthw13=2: 10 minutes.}

Screen 3:

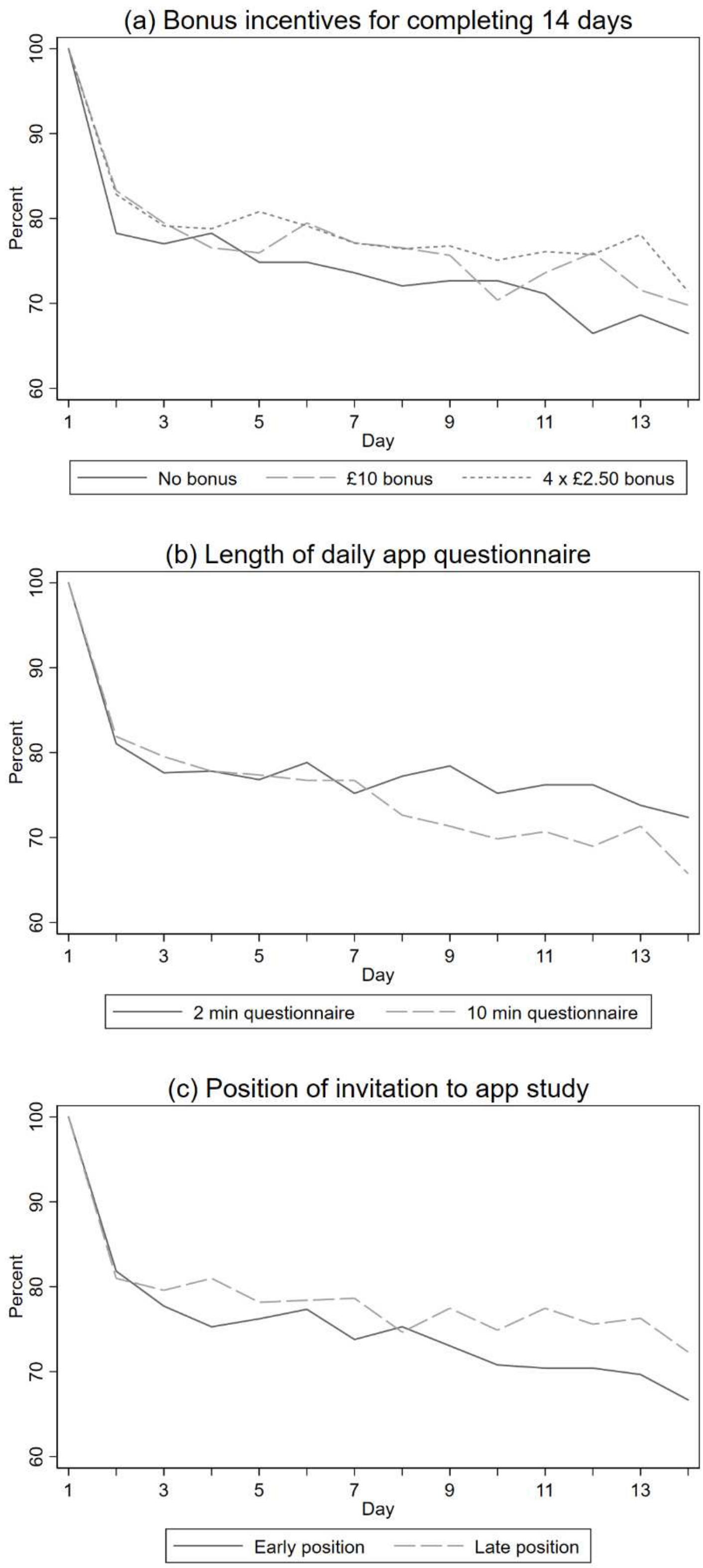
As a thank you for participating in this study, you will receive £1 for each day on which you complete the questions {if ff_appincentw13=1: . / if ff_appincentw13 = 2: , plus an additional £10 if you complete the questions on each of the 14 days. / if ff_appincentw13 = 3: plus you will have a chance to receive additional rewards totalling £10 if you complete the questions on each of the 14 days.}

Screen 4:

If possible, please try to answer these questions on your own, without discussing them with anyone else. For data security reasons, you will not be able to go back to previous questions.

By using this app you consent to *Understanding Society* using the data you submit via the app.

Appendix Figure 1: Daily participation rates by experimental treatment group



Appendix Table 1: Participation bias estimates for the sub-sample who used the app at least once, by experimental treatment group

		Position of invitation to app				Length of app questionnaire				Bonus incentive					
		Early		Late		2mins		10mins		None		£10		£2.50 x 4	
Variable	Level	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)	Bias	(s.e.)
Age	16-30	**4.2**	**(0.765)**	2.0	(1.169)	**2.0**	**(0.976)**	**4.7**	**(0.946)**	**2.6**	**(1.184)**	**4.7**	**(0.986)**	2.3	(1.366)
	31-40	**4.4**	**(0.675)**	**4.7**	**(0.929)**	**4.9**	**(0.789)**	**4.1**	**(0.799)**	**5.1**	**(0.897)**	**4.3**	**(0.983)**	**4.0**	**(1.038)**
	41-50	**4.4**	**(0.814)**	**6.2**	**(1.079)**	**5.1**	**(0.799)**	**5.3**	**(1.062)**	**6.1**	**(1.158)**	**5.7**	**(0.993)**	**3.4**	**(1.312)**
	51-60	1.2	(1.221)	2.2	(1.339)	2.1	(1.204)	1.4	(1.335)	2.3	(1.589)	2.3	(1.352)	0.9	(1.722)
	61-70	**-4.0**	**(1.466)**	**-3.5**	**(1.618)**	**-4.6**	**(1.537)**	-2.8	(1.534)	**-6.2**	**(2.021)**	-3.0	(1.747)	-1.9	(1.834)
	71+	**-10.1**	**(1.728)**	**-11.6**	**(1.869)**	**-9.5**	**(1.752)**	**-12.6**	**(1.799)**	**-9.9**	**(2.079)**	**-14.0**	**(2.314)**	**-8.7**	**(2.161)**
Sex	Female	2.9	(1.463)	**3.6**	**(1.766)**	0.6	(1.609)	**5.8**	**(1.598)**	**4.7**	**(1.939)**	1.9	(1.895)	2.9	(2.069)
Education	GCSEs, other, none	**-6.1**	**(1.709)**	**-9.4**	**(2.034)**	**-7.2**	**(1.808)**	**-7.9**	**(1.920)**	**-7.3**	**(2.281)**	**-8.1**	**(2.217)**	**-7.0**	**(2.362)**
	A-level	-0.9	(1.298)	0.8	(1.442)	-0.5	(1.335)	0.4	(1.385)	-0.4	(1.735)	-0.4	(1.575)	0.6	(1.679)
	Degree	**7.0**	**(1.297)**	**8.6**	**(1.614)**	**7.7**	**(1.389)**	**7.5**	**(1.511)**	**7.7**	**(1.772)**	**8.5**	**(1.644)**	**6.4**	**(1.920)**
In work	Yes	**11.6**	**(1.192)**	**10.1**	**(1.592)**	**10.3**	**(1.336)**	**11.3**	**(1.435)**	**10.8**	**(1.726)**	**12.5**	**(1.522)**	**9.0**	**(1.855)**
Marital	Single, never married	0.7	(1.307)	2.4	(1.559)	0.6	(1.423)	2.2	(1.427)	-0.4	(1.818)	**5.1**	**(1.427)**	-0.9	(1.983)
status	Married, civil partner	**4.3**	**(1.416)**	0.2	(1.855)	2.7	(1.560)	2.6	(1.673)	3.8	(1.961)	-2.4	(2.017)	**6.9**	**(1.935)**
	Separated, divorced, widowed	**-5.0**	**(1.473)**	-2.5	(1.557)	**-3.2**	**(1.434)**	**-4.8**	**(1.564)**	-3.4	(1.837)	-2.7	(1.640)	**-6.0**	**(2.036)**
Has kids <16	Yes	**3.7**	**(0.757)**	**2.9**	**(1.109)**	**2.8**	**(0.904)**	**4.0**	**(0.949)**	**4.4**	**(1.122)**	**4.1**	**(1.009)**	1.4	(1.289)
Health	Excellent	0.6	(0.843)	**3.9**	**(0.959)**	1.7	(0.852)	**2.4**	**(0.930)**	0.3	(1.349)	**3.5**	**(0.833)**	**2.2**	**(1.035)**
	Very good	**4.3**	**(1.347)**	0.2	(1.809)	0.7	(1.567)	**4.4**	**(1.568)**	3.2	(1.918)	**4.7**	**(1.728)**	-1.0	(2.131)
	Good	-1.3	(1.528)	-1.1	(1.800)	0.7	(1.551)	-3.2	(1.752)	0.2	(1.956)	**-4.7**	**(2.089)**	1.5	(2.021)
	Fair/poor	**-3.6**	**(1.394)**	**-3.1**	**(1.528)**	**-3.0**	**(1.387)**	**-3.6**	**(1.513)**	**-3.7**	**(1.762)**	**-3.6**	**(1.695)**	-2.7	(1.885)
Disability	Yes	-2.9	(1.532)	-3.4	(1.833)	-2.8	(1.619)	**-3.7**	**(1.730)**	-1.6	(1.949)	**-7.0**	**(2.133)**	-0.7	(2.067)
Mental health	Above median distress	**3.7**	**(1.440)**	0.5	(1.851)	2.9	(1.554)	1.6	(1.709)	1.7	(2.037)	2.2	(1.899)	3.1	(2.052)
Smoker	Yes	**-2.3**	**(1.153)**	0.2	(1.172)	-0.8	(1.129)	-1.7	(1.170)	-2.5	(1.552)	-0.3	(1.216)	-0.7	(1.431)
Doctor visits	None	-2.0	(1.540)	1.7	(1.646)	-0.5	(1.537)	0.3	(1.630)	-2.0	(2.045)	1.3	(1.781)	0.2	(1.986)
in last 12	One or two	1.3	(1.484)	0.2	(1.853)	0.0	(1.627)	1.3	(1.698)	1.8	(2.003)	1.0	(1.927)	-1.2	(2.191)
months	Three or more	0.7	(1.313)	-1.9	(1.712)	0.5	(1.414)	-1.6	(1.595)	0.2	(1.806)	-2.2	(1.816)	1.0	(1.921)

Mobile device use	Smartphone + tablet	**9.7**	**(1.264)**	**11.9**	**(1.537)**	**9.8**	**(1.352)**	**11.5**	**(1.431)**	**10.6**	**(1.725)**	**10.3**	**(1.611)**	**11.3**	**(1.771)**
	Smartphone only	2.2	(1.277)	-0.6	(1.673)	0.8	(1.430)	1.0	(1.504)	1.7	(1.751)	1.5	(1.712)	-0.6	(1.938)
	Tablet only	**-4.2**	**(1.203)**	**-4.7**	**(1.263)**	**-5.1**	**(1.302)**	**-3.9**	**(1.159)**	**-4.3**	**(1.423)**	**-5.5**	**(1.599)**	**-3.6**	**(1.498)**
	Neither	**-7.6**	**(1.506)**	**-6.5**	**(1.372)**	**-5.6**	**(1.325)**	**-8.6**	**(1.493)**	**-8.0**	**(1.747)**	**-6.3**	**(1.697)**	**-7.2**	**(1.765)**
Frequency of using apps	Every day	**14.2**	**(0.935)**	**14.6**	**(1.354)**	**13.1**	**(1.110)**	**15.5**	**(1.167)**	**13.3**	**(1.446)**	**14.3**	**(1.281)**	**15.2**	**(1.460)**
	Several times/week	**2.3**	**(0.995)**	**2.4**	**(1.179)**	2.1	(1.056)	**2.9**	**(1.099)**	2.5	(1.313)	1.1	(1.436)	**3.8**	**(1.167)**
	Several times/month	-0.2	(0.930)	1.4	(0.992)	0.5	(0.924)	0.5	(0.981)	1.2	(1.155)	-0.2	(1.195)	0.5	(1.145)
	Once a month or less	0.2	(1.016)	-1.3	(1.340)	0.3	(1.081)	-1.3	(1.262)	1.3	(1.347)	-0.4	(1.282)	-2.5	(1.694)
	Never	**-16.5**	**(1.994)**	**-17.1**	**(1.995)**	**-15.9**	**(1.991)**	**-17.5**	**(1.993)**	**-18.3**	**(2.428)**	**-14.9**	**(2.363)**	**-17.1**	**(2.525)**
Health apps	Yes	**16.9**	**(0.874)**	**14.3**	**(1.340)**	**13.0**	**(1.132)**	**18.8**	**(1.044)**	**16.2**	**(1.320)**	**16.7**	**(1.192)**	**14.5**	**(1.505)**
	Mean absolute bias	4.8		4.6		4.1		5.2		4.8		5.2		4.4	
	Total (N)	1,098		1,054		1,066		1,086		740		731		681	

Note: Statistically significant bias estimates (p < .05) are in bold. For characteristics highlighted in grey there is a significant difference between the treatment groups. Interactions tested by estimating logit regressions of the probability that the respondent used the app at least once, regressed on the characteristic, the incentive treatment groups and their interactions.