Understanding Society

Working Paper Series

No. 2018-04

April 2018

Respondent burden in a Mobile App: evidence from a shopping receipt scanning study.

Brendan Read

Institute for Social and Economic Research, University of Essex

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Non-technical Summary

The use of mobile technologies such as mobile phones and tablets offers many new opportunities for data collection. New types of data might be able to be collected, or it may be possible to collect existing data types in new and innovative ways. As well as these new opportunities, there are new challenges. Again, these can both be unique to mobile data collection, or existing data collection challenges that are altered by using mobile devices to collect the data. This paper focuses on one challenge, respondent burden. Respondent burden describes the load that is placed upon people who participate in surveys by the demands the survey makes on them. This might be practical concerns, like how much time and effort the task takes to complete, called the objective burden. Otherwise, this might be participant's beliefs about how burdensome the task is, called subjective burden. Measures of both are considered along with several factors that might affect how burdened participants are.

The data used is from a study that makes use of an app for mobile devices to collect data about household spending, the *Understanding Society* Spending Study. Participants were asked to report their spending by either sending a photo of a receipt via the app developed by Kantar Worldpanel, entering information about a purchase manually, or reporting that they had not spent anything that day.

Measures of subjective and objective burden were found to be largely unrelated to one another, suggesting perceptions of burden seem to be separated from the objective burden of this task. This research also found that whilst there was a decrease in the amount of time it took participants to use the app as they continued participating, this was not matched by a decrease in how difficult they reported finding the task. The time it took to complete app uses was also found not to affect how likely participants were to drop out from the study.

Asking participants whether they were willing to complete survey tasks that were similar to the Spending Study turned out to be a good predictor of how burdensome the task would be. Those who were willing to use their camera for a survey task were more likely to report their time and effort was very well spent participating. Those who reported that they were willing to download an app to complete surveys took less time to use the app. Age had an interesting relationship with burden. Older participants took longer to complete app uses. However, at the same time, they tended to report being more interested in participating.

Respondent burden in a Mobile App: evidence from a shopping receipt scanning study.

Brendan Read – (University of Essex)

April 24, 2018

Abstract:

This study considers the burden placed on participants, subjectively and objectively, when asked to use a mobile app to scan shopping receipts. Using data from both the *Understanding Society* Spending Study, and the ninth wave of the *Understanding Society* Innovation Panel allow measures of burden and related characteristics to be identified. Subjective and objective burden were found to be seemingly unrelated to one another. There is evidence of older respondents facing greater objective burden, however there was some evidence that this did not correspond to an increase in the levels of subjective burden reported. Reported willingness to participate in a task of a similar nature proved to be indicative of both objective and subjective burden.

Keywords: Subjective, Objective, Cumulative, Fatigue, Expenditure, Measurement of Consumption, Household Panel Survey

JEL Classification: C18, C81, C83

Acknowledgements:

This study was supported by a +3 PhD studentship from the Economic and Social Research Council (ESRC) awarded to Brendan Read as part of an associated studentship funded from a grant awarded to Michaela Benzeval to conduct waves nine through eleven of *Understanding Society* [grant number ES/N00812X/1]. Data from the ninth wave of the *Understanding Society* Innovation Panel is used, which was funded by a grant awarded to Nick Buck and Michaela Benzeval to conduct waves six through eight of *Understanding Society* [grant number ES/K005146/1]. In addition, data from the *Understanding Society* Spending Study is used which was funded by a research grant from the ESRC Transformative Research scheme and the National Centre for Research Methods (NCRM) awarded to Annette Jäckle [grant number ES/N006534/1].

Contact information: Brendan Read, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex UK, CO4 3SQ email: bread@essex.ac.uk.

Introduction

Using mobile technologies to collect survey data offers many potential benefits (Link et al., 2014). Allowing respondents to complete surveys using their mobile devices allows them greater freedom to choose when and where they do this. In addition, a range of new data types can be collected through mobile technologies and apps including: 'voice, photography, video, text, email [and] GPS' (Link et al., 2014, p. 22). Along with the new data collection opportunities offered by using these new technologies, it is also important to consider the potential challenges they present. These could be challenges unique to data collection using a mobile device or app, or existing survey data collection challenges altered by the new context. This paper focuses on one such challenge, respondent burden, examining this in the context of mobile data collection using an app to measure spending.

Historically, there have long been concerns about the demands surveys place upon respondents and how this may affect the data collected (Ruch, 1941; Young and Schmid, 1956). One particular concern was that longer surveys may result in respondents being reluctant to participate. Another concern was that throughout a survey a respondent's reporting behaviour may change. More recently, such concerns have been conceptualised as respondent burden (Bradburn, 1978).

Burden is conceptualised as consisting of two dimensions: objective burden, the 'total time and financial resources expended by the survey respondent to generate, maintain, retain, and provide survey information' (Office of Management and Budget, 2006, p. 34); and subjective burden, 'the degree to which a survey respondent perceives participation in a survey research project as difficult, time consuming, or emotionally stressful' (Graf, 2008, p. 740). Both dimensions, and the relationship between them, are of interest in this paper.

These two dimensions can be thought of as existing on a continuum ranging from low to high levels of burden. In line with much of the existing research into response burden (e.g. Sharp and Frankel, 1983; Dale and Haraldsen, 2005; Yu et al., 2015) this paper focuses on participants, as measures of burden are not captured for non-participants.

The data collection task that is the focus of this paper is the *Understanding Society* Spending Study. Participants were asked to use an app every day for one month to scan shopping receipts, submit purchases made without obtaining a receipt, or report days without spending. Data from the app, accompanying debrief questionnaires, and wave nine of the *Understanding Society* Innovation Panel are used to examine the following research questions:

1. Are subjective and objective measures of burden related?

2. How do subjective and objective burden change over the course of the study?

3. Does objective burden predict breaks in participation?

4. What factors predict subjective and objective burden?

Background

The exact relationship between objective (also called actual) and subjective (also called perceived) burden has not always been clear. Bradburn, in his seminal discussion of respondent burden, suggested that ‘ “burdensomeness” is not to be an objective characteristic of the task, but is the product of an interaction between the nature of the task and the way in which it is perceived by the respondent’ (Bradburn, 1978, p. 49). This acknowledges the importance of the nature of the task, an objective set of features, but suggests that this is important only in terms of how it shapes the subjective perception.

More recent accounts have made the case for considering both the objective and subjective dimensions of burden (Ampt, 2003; Willeboordse, 1997). By considering both dimensions it is possible to acknowledge the role of objective burden in shaping subjective burden, whilst also considering objective burden in its own right, if for no other reason than the factors determining objective burden are likely to be more easily controllable by the survey practitioner.

Evidence for the relationship between subjective and objective measures of burden has been mixed. Dale and Haraldsen (2005) report a high correlation between subjective and objective measures of burden. However in this study the objective measure (how long it took to complete the survey) relies on self-reports and therefore it is less surprising that it correlates with other subjective measures.

Sharp and Frankel (1983) examined the relationship between a wider selection of measures of subjective and objective burden. They experimentally varied the objective length of the survey and the level of effort necessary to complete the survey. In addition, measures of objective burden including item refusal and nonresponse rates were collected. Subjective burden was captured through self-reports of: willingness to be re-interviewed, willingness to participate for longer, interest in the study, judgement as to important the study was, difficulty, whether time and effort was well spent, belief that the interview was the right length. The evidence suggested that a longer survey resulted in greater reports of subjective burden on the indicators related to length. However, there was little evidence of relationships between the other measures of burden examined.

Yu et al. (2015) attempted to disentangle the subjective from the objective by varying experimentally the actual length of a survey, and the presentation of that length, so as to examine whether separate effects of both increased objective burden and increased subjective burden could be observed. They found that not only did increasing the objective length of the survey increase the levels of reported burden but presenting the survey as longer and more burdensome also further increased the levels of reported burden.

Factors determining burden

Bradburn (1978) identified four survey characteristics that determine burden: survey length, the amount of effort required to complete the survey, the amount of emotional stress caused, and the frequency of interviewing . Haraldsen (2004) suggested three respondent chatacteristics as factors determining burden: the respondent's competence/ability, their interest/motivation, and their availability/opportunity to complete the task.

Such a dichotomy into survey and respondent characteristics is somewhat misleading. This is because it suggests that the seven factors identified are solely influenced by either design choices, or the nature of a respondent. Instead the case can be made that each of these seven factors ultimately ends up being determined by characteristics of both the survey and the respondent. For example, how long a survey takes to complete is both determined by the amount of content specified, and the variance in the length of time individuals take to respond.

Therefore, in this paper, the approach of combining the list of four factors suggested by Bradburn with those suggested by Haraldsen is taken, resulting in one list of seven factors that contribute to respondent burden. Most of these factors has been discussed in the existing survey methodological literature. The following review focuses on research findings that help inform the study of burden within the context of the Spending Study.

Survey length

Presenting information that suggests a longer survey to respondents has been found to have a negative impact on response rates in web surveys (Crawford et al., 2001; Galesic and Bosnjak, 2009), telephone surveys (Collins et al., 1988; Roberts et al., 2010), face-to-face surveys (Groves et al., 1999), and postal surveys (Yammarino et al., 1991; Dillman et al., 1993). However, when it comes to the actual time taken to complete a survey there is some evidence that those with the longest response times may be those individuals who have engaged the most with the topic of the survey, and for whom that topic is particularly relevant (Branden et al., 1995). Similarly, those respondents with the longest response times in a given wave of a panel study have been found to be more likely to respond in subsequent waves (Lynn, 2014). This suggests that expectations about survey length matter more than the actual experience of the time taken to participate.

Effort

Couper and Nicholls (1998) express concern that the shift from paper or interviewer-based modes to web modes of data collection may result in respondents having to expend more effort to participate. This is because some of the tasks traditionally performed by the data collector are instead coming to be performed by the respondent. This shift, whilst potentially beneficial in terms of reducing costs, or potentially reducing processing errors, comes at the cost of increasing the burden placed upon the respondent. It seems likely this holds true for much data collection using mobile devices or apps, which often require additional tasks to be performed, as is the case in the Spending Study.

Emotional stress

Typically research into the emotional stress caused by surveys has looked at the effect of sensitive questions on specific vulnerable populations. For example, emotional stress has been found to make participation harder in surveys on: sexual and physical violence among adults (Walker et al., 1997), bereavement (Dyregrov, 2004), and traumatic injuries (Ruzek and Zatzick, 2000).

There has also been some evidence of question sensitivity as a barrier to participation amongst subgroups in general population surveys (Newman et al., 2001; Galea et al., 2005), though the characteristics of the affected subgroups identified have not always been clear. Kreuter et al. (2008) found that questions were more likely to be sensitive for respondents who belonged to groups with a sensitive status related to the concept being measured. This seems to support the idea that the amount of emotional stress caused by a survey instrument is not simply an innate characteristic of that given instrument, but it also shaped by the characteristics of the respondent receiving that instrument. As such, a given survey instrument may potentially be more stressful and thus produce higher burden for some individuals or subgroups of a sample as opposed to others.

Frequency

Less research has been conducted to examine the increased burden of being more frequently surveyed. In part, this may be due to difficulty in obtaining an accurate measure of how many other surveys respondents have taken part in prior to participation in any given survey.

Availability/Opportunity

To be able to participate sample members must have time available. As sample members only have a finite amount of time available to them they must make decisions about whether to spend that time participating. Framing this through the lens of traditional economic thought surrounding issues of resource scarcity (drawing upon Raiklin and Uyar, 1996), participation in the survey comes at the opportunity cost of not using their time for other activities. This cost is most sharply felt where time is a scarce resource. Previous research considering time constraints as a barrier to participation have found evidence to suggest that those who are more likely to have time constraints have a lower propensity to respond (Groves and Couper, 1998; Abraham et al., 2006).

As the data collected in the present study relied on the use of a mobile device, another important factor when considering the opportunity to participate is whether a sample member has access to a mobile device with which to take part in the study. Where a sample member does not have access to a mobile device, the objective burden of participating is clearly higher, as they must have the opportunity to either borrow or otherwise acquire access to a device to allow participation. Not possessing a mobile device may also contribute to the level of effort necessary to participate. Whilst a respondent may have the opportunity to gain access to a device, repeatedly acquiring that access may be considered too much effort, thus meaning the participant chooses either to participate less, or not at all.

Ability/Competence

Lower cognitive ability has been highlighted as a widely accepted cause of measurement error (Fricker and Tourangeau, 2010). Lower cognitive ability may result in greater difficulty completing a task, thus increasing the burden. Satisficing describes a response strategy where respondents attempt to reduce the burden of participation by producing sub-optimal (in the eyes of the survey practitioner) responses. Lower cognitive ability has been found to increase the likelihood of a respondent satisficing (Krosnick, 1991; Knäuper et al., 1997).

Lower device familiarity, or lower ability to complete survey tasks on a mobile device, has also been considered as something which may increase burden (Jäckle, Burton, Couper and Lessof, 2017). This may affect both the subjective burden, as sample members evaluate their ability to perform the task, and the objective burden, how well respondents are actually able to perform the task.

Motivation/Interest

One factor affecting a respondent's motivation is the topic or subject matter of the survey they are asked to complete. When being approached with a survey request, evidence suggests that if that request is related to a topic in which the respondent has been observed to have an interest, their propensity to respond will be increased (Groves et al., 2004). Conversely, a lack of interest has been found to result in a lower propensity to respond (Couper, 1997).

The consensus is that the use of incentives helps to motivate respondents, and improve the rate of participation (Armstrong, 1975; Singer et al., 1999). Typically, unconditional incentives have been found to be better motivators than conditional incentives (Church, 1993; Goyder, 1994; Young et al., 2015). However, there is evidence of a so-called "*ceiling effect*" when using incentives to promote response, with the impact of incentives being diminished when respondents are already motivated to take part in a survey (Groves et al., 2000; Zagorsky and Rhoton, 2008).

Dynamic burden

Much discussion of burden considers it as a static concept. Where the potential for dynamic change is considered this is implicit, rather than being explicitly stated. Insight into how burden may change dynamically throughout the course of data collection can be gained from explanations of respondents either breaking off or dropping out of web surveys or diary studies. Each of these contexts presents a different model for the decision-making process leading to participants ceasing to participate.

Accounts of break-off in web surveys have suggested participants go through of an ongoing decision-making process about whether to continue participating in a survey (Galesic, 2006; Haraldsen, 2004; Peytchev, 2009). Some of these analyses draw upon *decision field theory,* developed by

Busemeyer and Townsend (1993), which describes a dynamic decision-making process. One of the key aspects of decision field theory is the notion of an inhibitory threshold: 'the point which determines when the difference in the preference for one or the other action is large enough to provoke behaviour'(Galesic, 2006, p. 314). When respondents fall below this inhibitory threshold, they shift from making the decision to participate to making a decision to stop participating. In this account the preference to participate or not can rise or fall throughout the course of participation.

This seems to be consistent with existing understanding of subjective burden. Respondents possess an ongoing subjective perception of how burdensome the task is throughout the time that they are participating. This can rise or fall as a product of both the experience of participating or input from the survey practitioner. If the subjective burden is sufficiently large, the participant ceases to participate. As such, subjective burden can only be considered as a momentary concept, it does not make sense to talk of subjective burden in terms of a total value across the whole of the data collection process. Instead it is a latent value that fluctuates throughout the course of participation.

In comparison, it has been suggested that drop out in diary studies results from cumulative fatigue (Gillmore et al., 2001). Fatigue builds throughout participation and can therefore only increase as time goes on. Evidence of fatigue, as measured by a decrease in responding throughout the course of a diary study, has been mixed. There are examples of studies in which respondents show evidence of becoming fatigued (Gerstel et al., 1980; Leigh, 1993; Verbrugge, 1980) and some studies in which the effect does not seem to be present (Lemmens et al., 1988; Persky et al., 1981; Searles et al., 1995). Gillmore et al. (2001) suggest that both respondent and design characteristics may play a role in determining whether respondents become fatigued in a diary study. However, their attempts to identify examples of specific characteristics that contribute to fatigue were not able to provide much insight.

Objective burden seems to be consistent with this account of cumulative fatigue. We might consider that all participants start having expended zero time and energy participating and this increases throughout the course of data collection. As such, we can attribute a total value for the objective burden that each participant has experienced both up to any given moment and across the data collection process as a whole. In addition, if the data collection task is considered as a series of discrete tasks we can consider the objective burden of any particular task.

Overall, burden consists of the load placed upon respondents, this can be thought of in two dimensions the objective load placed upon them, and how they subjectively perceive this load. These two dimensions are both shaped by a range of seven factors that are influenced by characteristics of both the survey task and the respondents themselves. These dimensions then dynamically change throughout the process of data collection, in an iterative process, where experience shapes future burden.

Data

Study designs

To examine burden this research uses data from both wave nine of the *Understanding Society* Innovation Panel (IP9) and an inter-wave receipt scanning project known to participants as the *Understanding Society* Spending Study, which took place between waves nine and ten of the Innovation Panel. The main variables of interest are taken from the Spending Study, with variables from IP9 used as covariates for some of the analyses.

Innovation Panel

The Innovation Panel (University of Essex. Institute for Social and Economic Research, 2017) is one part of the UK Household Longitudinal Study, *Understanding Society*. The IP exists to allow the implementation of experiments and research into issues of data collection procedures within the context of longitudinal surveys. The sample design is a stratified, clustered sample of all households within Great Britain, south of the Caledonian Canal. The ninth wave contains the original sample along with refreshment samples from waves four and seven onwards. All household members aged sixteen and over at the time of interviewing are considered eligible for annual interviews. The data used in this paper come from the ninth wave which had a household response rate of 84.7% and an individual response rate of 85.4% within responding households (Jäckle, Gaia, Al Baghal, Burton and Lynn, 2017).

Spending Study

The Spending Study is part of a project to give a better account of household finances by developing innovative methods of collecting data on this topic. The study was conducted in partnership with Kantar Worldpanel, who developed the app. Respondents were tasked with downloading and using the app on their smartphone or tablet, to provide data about their spending across the span of a month.

Spending could be reported by scanning receipts, inputting a purchase without a receipt, or reporting a day in which nothing was spent.

The issued sample for the Spending Study consisted of all adult members (aged 16 or over) of households where at least one person in the household responded at IP9. Household members who are known to have refused to participate long-term in the Innovation Panel were not included in the sample.

Alongside the data collected via the app, the Spending Study also asked participants to complete several additional questionnaires, with questions regarding the experience of participating and some additional questions about their household expenditure. End of week surveys asked participants to reflect on the previous week's participation. An end of project questionnaire asked participants to reflect on the experience of participating as a whole. The end of project questionnaire was first implemented as an online survey, before a paper follow-up was sent out to those who had not initially responded to the online version.

Different incentive amounts for different forms of participation in the study were offered to participants, with the incentives being made available in the form of either Love2Shop gift vouchers or gift cards. These are redeemable in many high-street stores throughout the UK. There was an initial incentive for completing a registration survey and downloading an app with two conditions (£2 vs £6). All members of a given household received the same incentive treatment. Secondly, in an effort to further increase the rate of response, an additional £5 incentive was sent to members of a random half of all households where no-one had participated by the third week of the study. These first two incentives are included as covariates in the analyses presented here. In addition, participants received a 50p a day incentive for every day in which they used the app. Completion of each end of week survey earned a further 50p, and completing the end of project survey earned £3. Finally, a bonus of £10 was offered if a participant used the app every day for 31 days. Ultimately, this requirement was relaxed so that all participants who used the app on at least 27 days throughout the study received this bonus. Participants were sent an email at the end of each week updating them on how much they had earned in incentives so far.

Analytical sample

To allow covariates from IP9 to be used in the analyses in this paper only the 2,112 cases sample members who completed a full adult interview at IP9 were considered for the analytical sample. Of these IP9 respondents, 270 attempted to use the app, with 268 successfully completing an app use, a response rate of 12.7%. Table 1 documents the response rates at different stages of the study and how the analytical sample was derived.

Table 1: *Breakdown of response rates for different stages of the Spending Study.*

	n	% of issued sample	% of participants	% of analytical sample
Issued sample	2112	100.0		
Completed at least one app use	268	12.7	100.0	
Completed end of project survey	238	11.3	88.8	
Received subjective burden questions	224	10.6	83.6	
Analytical sample	223	10.5	83.2	100.0
Completed end of week surveys				
Week one	134	6.3	50.0	60.1
Week two	132	6.2	49.3	59.2
Week three	139	6.6	51.9	62.3
Week four	137	6.5	51.1	61.4

Of the 268 app users, 238 responded to the end of project survey (88.8%). As the subjective measures of burden were asked in the end of project survey the analytical sample for this paper is constrained to just those participants who completed this survey. Due to an error in the scripting of the web version of the end of project survey, fourteen participants who completed this survey did not receive the subjective burden questions. These fourteen cases were individuals who had not participated in the final week of the study and were allocated to receive questions about why they had dropped out. Instead these participants received a version of the questionnaire intended for non-participants, thus they were not asked any of the questions reflecting back on the experience of participating. This left 224 cases who received the subjective burden questions. Of the 224 cases, a single participant did not answer all of subjective burden questions, and was subsequently dropped from these analyses, leaving a final analytical sample of 223. This constitutes 10.5% of the issued sample and 83.2% of participants in the Spending Study. The average number of end of week surveys completed each week was around 136 out of the 223 analytical sample members. This was about 60% of the analytical sample.

The analyses presented here are constrained to the analytical sample, though those analyses which only examined objective measures of burden, were repeated with all 268 app users. The differences between the two specifications were for the most part minimal, with notable differences highlighted throughout the results section of this paper.

Measures of burden

Objective measures of burden

Four measures of objective burden were derived from paradata collected by the app: the number of app uses each partcipant completed, the total time they spent completing these app uses, their average time per app use, and the durations of the individual app uses. The first two of these measures capture the total cumulative burden of individuals across the course of the whole study. The latter two instead attempt to measures the amount of objective burden per app use.

Descriptive statistics for these four measures, both broken down by type of app use, and pooled across all types of app use are presented in Table 2. The assumption here is that a longer period of time or more app uses equals a greater objective burden placed upon the participant.

Table 2: *Descriptive statistics for the four measures of objective burden.*

	$\bar{x}$	s	Q_1	Q_2	Q_3
Number of app uses completed by each participant					
All app uses	46	21	33	43	55
Receipts scanned	24	18	12	20	33
Purchases without receipts	15	12	5	11	21
Nothing bought	11	8	5	9	15
Average duration of app uses for participants (seconds)					
All app uses	34	15	24	32	41
Receipts scanned	52	28	35	44	61
Purchases without receipts	36	25	23	31	42
Nothing bought	11	7	7	9	13
Total duration of app uses for participants (seconds)					
All app uses	1575	978	854	1327	2175
Receipts scanned	1131	816	532	940	1596
Purchases without receipts	479	404	195	377	637
Nothing bought	103	85	43	85	139
Duration of each app uses (seconds)					
All app uses	35	41	14	24	40
Receipts scanned	47	47	23	33	53
Purchases without receipts	32	34	17	25	36
Nothing bought	10	12	5	7	10

The total number of app uses for the analytical sample of 223 participants was 10,381. Of these 202 were identified as potential outliers, taking longer than 10 minutes to complete. It is possible that some of these app uses did represent active participation for a longer period. However, it is also likely that some of these longer durations were the result of respondents either choosing or being forced to pause part way through an app use, and subsequently resuming later. As it is impossible to disentangle long active response times from long response times produced by a period of inactivity, together with the potential bias these large results may have produced, they were dropped from analysis. This left an analytical sample of 10,179 app uses, which is the base used for all of the analyses presented here.

The mean number of app uses completed by an individual was 46, which is about one or two app uses per day throughout the course of the study. The mean time to complete an individual app use was 35 seconds. The grand mean of the mean time taken by each respondent to complete their app uses was 34 seconds. The mean total time taken by an individual to complete all their app uses was 1575 seconds, this equates to a little over 26 minutes throughout the course of the study.

Subjective measures of burden

Four measures of subjective burden were taken from the end of project survey. All four measures were adapted from measures used by Sharp and Frankel (1983). The distributions for these four subjective measures were skewed towards lower levels of burden. This, combined with the relatively small analytical sample size, means that the number of responses in the categories representing highest burden was typically quite small. The decision was made to recode these variables into four dichotomous measures. Specifications for models using both the original form of these variables and the dichotomised form were considered, however the original form resulted in a number of empty cells at certain levels of the four measures of subjective burden in the multivariate analysis or resulted in estimations being made from a very small number of cases. In most cases this violated the proportional odds assumption of the ordered logistic regression models. Therefore, the dichotomised specifications of models are presented here. The original and recoded responses to these variables can be found in Table 3.

One of these four measures, self-rated ease or difficulty participating in the study, was also asked each week in the end of week surveys, reflecting on the previous week. A week by week breakdown of the response distributions for this variable can be found in Table 4.

Table 3: *Response distributions for four subjective measures of respondent burden (original and recoded).*

	Original n	Original %		Recoded n	Recoded %
Likelihood - *'Imagine you were being asked to do this Spending Study for the first time. Based on your experience, how likely would you be to participate?'*					
Very likely	150	67.3	Higher likelihood	150	67.3
Somewhat likely	57	25.6	Lower likelihood	73	32.7
Somewhat unlikely	11	4.9			
Very unlikely	5	2.2			
Time/effort - *'Overall do you feel that the time and effort you put into participating in the Spending Study was...'*					
Very well spent	112	50.2	More well spent	112	50.2
Somewhat well spent	106	47.5	Less well spent	111	49.8
Not very well spent	5	2.2			
Interest - *'Overall how interesting was participating in the Spending Study?'*					
Very interesting	88	39.5	Higher interest	88	39.5
Somewhat interesting	111	49.8	Lower interest	135	60.5
Not interesting	24	10.8			
Difficulty- *'Overall, how easy or difficult did you find completing the Spending Study?'*					
Very easy	88	39.5	Lower difficulty	88	39.5
Somewhat easy	95	42.6	Higher difficulty	135	60.5
Somewhat difficult	36	16.1			
Very difficult	4	1.8			

Table 4: *Response distributions for end of week measure of Spending Study difficulty listed for each week and pooled across all weeks.*

Week	Very easy		Somewhat easy		Somewhat difficult		Very difficult		Missing	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
1	56	25.1	55	24.7	20	9.0	3	1.4	89	39.9
2	53	23.8	51	22.9	25	11.2	3	1.4	91	40.8
3	58	26.0	53	23.8	23	10.3	5	2.2	84	37.7
4	57	25.6	63	28.3	15	6.7	2	0.9	86	38.6
Pooled	224	25.1	222	24.9	83	9.3	13	1.5	350	39.2

Predictors of burden

A number of potential predictors of burden were identified including: existing mobile device usage behaviours, willingness to complete survey tasks with a mobile device, existing financial behaviours, demographic characteristics and the two incentive treatments outlined earlier. The seven factors affecting burden identified in the background section were reflected upon when deciding which variables to examine. Which of the seven factors it is believed that the predictor may effect is noted throughout. Descriptive statistics for each of these predictor variables can be found in Table 5.

Table 5: *Descriptive statistics for predictors of burden.*

		n	%
Initial incentive	*£2.00*	97	43.5
	£6.00	126	56.6
Received unconditional 5 incentive	*Yes*	39	17.5
	No	184	82.5
Uses device for taking photos	*Yes*	201	90.1
	No	22	9.9
Uses device for online banking	*Yes*	158	70.9
	No	65	29.1
Uses device to install apps	*Yes*	180	80.7
	No	43	19.3
Willingness to download a survey app	*Not willing*	44	19.7
	Willing	179	80.3
Willingness to use the camera on device to take photos or scan barcodes	*Not willing*	38	17.0
	Willing	185	83.0
Frequency of checking bank balance	*Less than once a week*	43	19.2
	Once a week or more	181	80.8
Keeps a budget	*Yes*	116	52.0
	No	107	48.0
Poverty threshold	*Below the threshold*	28	12.6
	Above the threshold	195	87.4
Time constrained	*Yes*	65	29.1
	No	158	70.9
Disabled/ long term illness	*Yes*	56	25.1
	No	167	74.9
Gender	*Male*	87	39.0
	Female	136	61.0
Age	$\bar{x}$	44	
	s	15	
	Q_1	31	
	Q_2	43	
	Q_3	53	
Level of education	*Less than a degree*	124	55.6
	Degree or higher	99	44.4

Activities performed on mobile device - *Ability/Motivation/Emotional stress*

Questions about whether respondents performed a range of activities on their mobile device were asked to respondents who reported access to either a smartphone or tablet. Respondents were presented with a list of possible activities and asked, '*Do you use your smartphone for the following activities?*' Of those activities three were identified as being related to the Spending Study. The first two of these, '*Taking photos*', and '*Installing new apps (e.g., from iTunes, Google Play Store)*', were both necessary skills to participate in the study. Being familiar with performing either of these tasks likely increased the ability of participants to take part in the study, thus decreasing the burden they faced.

The third activity, '*Online banking (e.g., checking account balance, transferring money)*', was a related skill which was included with the idea that those respondents who did this would likely be more comfortable accessing and transmitting their financial information through an app. It was felt that this greater comfort performing the task of transmitting financial information digitally might result in less emotional stress when participating in the study, meaning the burden for those participants used to doing this would be decreased. It was also considered possible that those who checked their finances online may have more interest in the topic of the study, increasing their motivation, thus reducing the subjective burden of participation.

As respondents were asked this set of questions for both mobiles and tablets, each of these activities was coded 1 if the respondent reported performing the activity on either device, or 0 if they did not report performing it on either. As those without access to either device did not receive these questions, these respondents were also coded to 0, with the assumption that without access to a device they could not perform these actions.

Willingness to perform survey activities on mobile device - *Motivation*

A series of hypothetical questions about willingness to perform different survey activities on mobile devices were asked. Of these, two were felt to be directly related to the tasks performed in the Spending Study, and likely therefore to be indicative of greater motivation to participate.The assumption here is that reporting being willing to perform this task would likely mean that the participant would be more likely to surpass the initial inhibitory threshold for deciding to participate, and as such their subjective perception of burden would be lower from the onset.

Respondents were asked '*How willing would you be to carry out the following tasks on your smartphone/tablet for a survey?*' The two items included are willingness to '*Download a survey app to complete an online questionnaire*' and '*Use the camera of your smartphone to take photos or scan barcodes*'. Both items were measured on a four-point scale of '*not at all willing/a little willing/somewhat willing/very willing*'. Where the respondent was asked both for tablet and smartphone the higher value of their two answers was taken. This was on the assumption that respondents would choose to use the device they had reported being the most willing to perform the task on. Two alternative specifications were considered, one keeping the original four answer categories, another collapsing these variables into not at all willing vs any of the other levels of willingness. On examination of the alternative specifications, the important distinction seems to be whether the participant was willing or not, as opposed to the degree of willingness; therefore, the dichotomous specification is presented here. Again, these questions were filtered on device access, and subsequently sample members who did not receive their questions were coded to 0.

Existing financial behaviors - *Ability/Motivation*

As with the existing mobile device behaviors, reported participation in certain existing financial behaviors are suggested as potential indicators of increased evidence of interest in the topic of the Spending Study. In line with existing evidence that interest results in a greater motivation to respond (Groves et al., 2004) it is expected that participants who engage in these financial behaviours will typically report being less burdened.

One measure used was an indicator measuring if respondents kept a budget. Respondents were asked ‘*Now, thinking about different ways that people have of managing their finances, how, if at all, do you record your budget?*’ which was coded 0 if they did not report keeping any form of budget and 1 if they did. Respondents were asked ‘*How often do you check your bank balance?*’ with ‘*most days/ at least once a week/ a couple of times a month/ at least once a month/ less than once a month/ never’* as response options. The original variable was highly skewed and therefore recoded into a binary indicator of high or low frequency for analysis with’*most days/at least once a week’* being coded as 1, and ‘*a couple of times a month/at least once a month/less than once a month/never*’, coded 0.

As these measures are tied to skills related to tracking your finances (keeping receipts, being aware of how much you have spent, etc.) it also seems likely that those participants who already take part in these activities may have more ability to complete the task as they already possess a number of useful associated skills.

Poverty indicator - *Emotional stress*

Given the subject of the Spending Study, it was considered that the topic of the survey may be sensitive for those with the lowest household incomes, and thus cause more emotional stress, meaning the task was more burdensome. As such, an indicator was derived marking the threshold under which individuals were considered to be living in poverty. This was defined as those individuals whose equivalised net household income fell below 60% of the median equivalised net household income. As the Innovation Panel only derives gross income, not net, this figure was first calculated for the seventh wave of the main *Understanding Society* (US7) sample (this wave having occurred for the most part in the same year as IP9). The resulting figure was £922.67. Equivalised gross household income for US7 respondents was then regressed on their equivalised net household income. The resulting regression coefficient was then used to calculate a corresponding gross poverty threshold from the earlier net threshold. The resulting threshold was £1025.38, which was applied to the analytical sample, to derive the final poverty indicator. All individuals whose household equivalised gross income fell below this threshold were considered to be living in poverty.

Time constraints - *Opportunity*

Participants with greater time constraints seem likely to have less opportunities to participate. An indicator of this was derived taking into account a number of factors. Participants were considered time constrained if they reported working more than forty hours a week, either in employment or self-employment. Those with a commute of greater than an hour to get to work each day were also coded as time constrained. In addition to this, participants were considered time constrained if they had any children under the age of five living in the household. The final derived variable took the value of 1 if a respondent met any of the criteria for being considered time constrained, or otherwise took a value of 0.

Disability or illness - *Ability*

An indicator for whether an individual had reported to be suffering from any long-standing physical or mental impairment, illness or disability was included as an indicator of participants' ability to participate in the Spending Study. Reporting such a longstanding illness or disability is considered here to reduce ability to participate. This was coded 1 if they reported that they did have a longstanding illness or disability, and 0 if they did not.

Level of education - *Ability*

Level of education was included as a proxy for cognitive ability. Participants' level of education was coded as 1 for a degree or above and 0 if a respondent's highest level of qualification was lower than this. Participants with higher education are expected to find the task easier. This may result in the task taking them less time to complete. It may also result in them reporting finding the task easier, and this may translate to other measures of subjective burden also being lower.

Demographics

Two demographic control variables were included in the analyses. Sex was coded as 0 for male respondents, and 1 for female. Age was included as a continuous variable, and the possibility of a curvilinear relationship was explored, however the introduction of a squared age term did not show evidence of such a relationship, and this squared term was subsequently removed from the analyses presented here.

Results

To address the four research questions in this paper, two different units of analysis are used throughout, either: participants, or the individual app uses, with app uses being clustered within participants. All standard errors are calculated adjusting for the complex clustered design of the Innovation Panel.

RQ1: *Are subjective and objective measures of burden related?*

For this first research question the unit of analysis is participants. As the four subjective measures of burden are measured at a participant level, the three objective measures chosen to be introduced in this analysis are those that are calculated at the participant level. To examine the relationship between objective and subjective indicators the matrix of correlations between the seven indicators was initially examined. An exploratory factor analysis was then carried out, examining the underlying structure of the seven indicators.

Polychoric correlations were used due to the potential drawbacks of using other correlation measures: neither Pearson's r or Spearman's ρ are appropriate as the subjective measures of burden used here are binary, Kendall's τ is suitable for binary measures, but the resulting correlation matrix cannot be used for factor analysis. The approach of using polychoric correlations to allow both ordinal correlations, and a subsequent factor analysis has previously been advocated by Maydeu-Olivares and D'zurilla (1995), Flora and Curran (2004) and Holgado-Tello et al. (2010) and is thus adopted here. These correlations were calculated using the user-written "*polychoric*" package written for Stata by Stas Kolenikov (2008) and are presented in Table 6.

Table 6: *Correlation matrix of the bivariate relationships between different measures of burden.*

	Likelihood	Time/ effort	Interest	Difficulty	Average time	Total time	No. of app uses
Likelihood	1.00						
Time/effort	0.51	1.00					
Interest	0.42	0.44	1.00				
Difficulty	0.66	0.62	0.67	1.00			
Average time	0.10	0.20	−0.15	−0.05	1.00		
Total time	−0.15	0.08	−0.23	−0.13	0.67	1.00	
No. of app uses	−0.27	−0.07	−0.19	−0.13	0.08	0.76	1.00

Notes: $n = 223$; Correlations between subjective measures are polychoric, correlations between objective measures and subjective measures are polyserial, correlations between objective measures are Pearson's r correlations.

Likelihood Time/effort Interest Difficulty Average time Total time Number of app uses

Using established thresholds for interpreting correlations (Hinkle et al., 2003) most of the relationships between each pairing of the four subjective measures fell within the range of moderate positive correlations (0.50 to 0.70). The only exceptions to this were the relationship between interest in the study and value placed on the study; and between interest and likelihood of participation. Here the correlations were lower, though both were above 0.40, indicating a low positive correlation.

The correlations between each of the subjective measures and the objective measures of burden produced coefficients that fell below the threshold for a remarkable relationship, falling within the range of -0.30 to 0.30. This seems to suggest that the subjective measures captured are not associated with any of the three measures of objective burden considered here.

Total time showed a moderate to strong relationship to both the number of app uses, and the average time taken to complete app uses. This is not a surprise as increases in either of these two variables would have been expected to increase the total time taken to complete app uses. The number of app uses did not show a strong association with the average time taken to complete an app use.

Before performing the exploratory factor analysis, a common test for the appropriateness of applying a factor structure to a set of variables was conducted. Bartlett (1951) suggests the test of sphericity to offer validation for one of the assumptions of factor analysis, namely that the variables are not orthogonal from one another. A result of $\chi^2 = 1087.53$, $df = 21$, $p < 0.001$ is indicative that the variables are not orthogonal from one another, and therefore suitable for factor analysis.

Having established the appropriateness of using factor analysis on the seven variables, a principal factors factor analysis was conducted, with an orthogonal varimax rotation. This was calculated using the earlier matrix of polychoric correlations. Only those factors that were above the threshold of the Kaiser criterion (Kaiser, 1960), an eigenvalue of 1.0, are presented. This produced a structure with three factors, and the factor loadings for each variable with relation to these factors are presented in Table 7.

Table 7: *Factor analysis of the structure of seven indicators of respondent burden.*

	Factor One	Factor Two	Factor Three	Uniqueness	KMO
Likelihood	0.69	−0.21	0.11	0.45	0.77
Time/effort	0.69	−0.01	0.19	0.49	0.84
Interest	0.68	−0.13	−0.16	0.48	0.77
Difficulty	0.88	−0.06	−0.05	0.22	0.68
Average time	0.02	0.10	0.94	0.10	0.24
Total duration	−0.06	0.78	0.61	0.02	0.39
App uses	−0.09	0.96	0.00	0.07	0.30
Eigenvalue	2.19	1.59	1.32		
Overall					0.50

Notes: $n = 223$; Factor structure after orthogonal varimax rotation applied; Factors with Eigenvalues higher than 1 presented.

For the first factor each of the four subjective measures of burden produced a factor loading greater than the suggested threshold of 0.60 (Guadagnoli and Velicer, 1988) suggesting strong associations between each of these variables the underlying latent variable. There is very little evidence of an association between the objective measures of burden and this underlying factor, further reinforcing the idea that the subjective measures and the objective measure are capturing different aspects of burden. The other two factors are largely related to a single variable, either the number of app uses, in the case of factor two, or average time taken to complete app uses for factor three. That total duration strongly loads onto each of these factors is again not surprising as this measure is a product of the other two variables. This factor structure seems to be indicative of a lack of underlying latent factors that can be inferred from the objective measures captured.

A test for the Kaiser-Mayer-Olkin measure of sampling adequacy (Kaiser, 1970; Kaiser and Rice, 1974) was also conducted with an overall result of 0.50; applying the criteria set out by Kaiser and Rice (1974) this value comes at the very lowest end of values considered appropriate for factor analysis. However, examining this for individual variables indicates that the subjective measures of burden are better suited for factor analysis than the objective measures. The four subjective measures ranged from 0.68 to 0.84, values that can be considered suitable for factor analysis. This compares to values ranging from 0.24 to 0.39 for the objective measures. This seems to further reinforce the notion that there is a latent structure underlying the four subjective burden measures, whereas the three objective measures do not seem to be related in this way.

RQ2: ***How do subjective and objective burden change over the course of the study?***

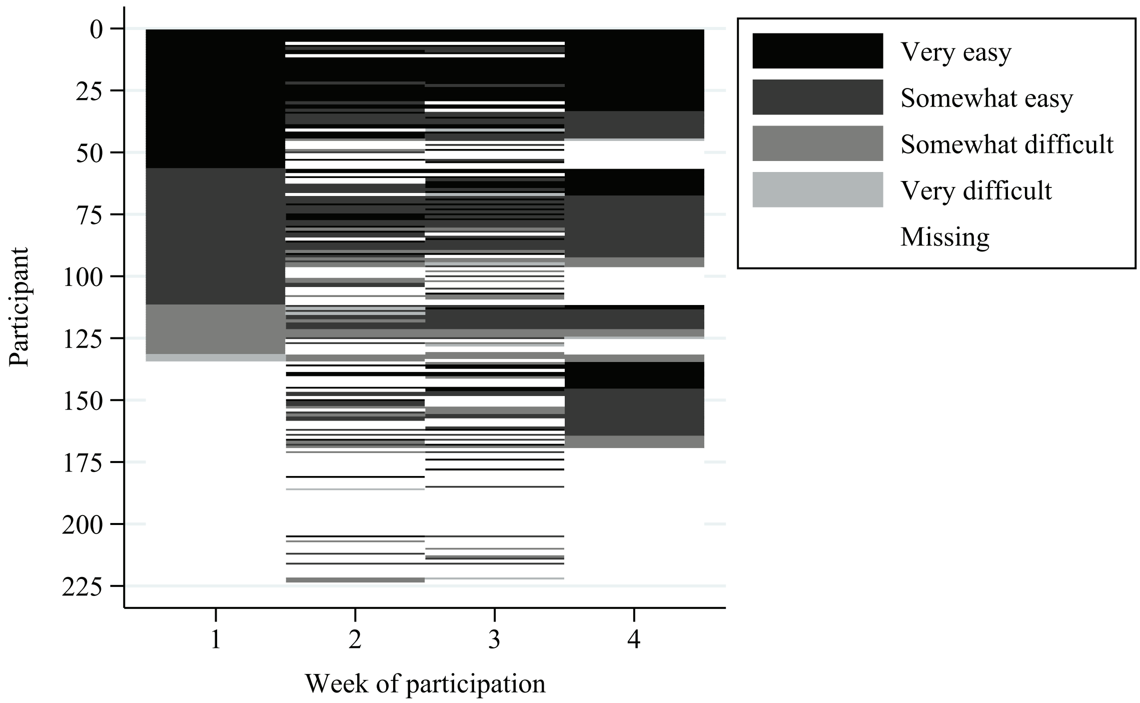
Subjective burden

To investigate the change in subjective burden across the four weeks of participation the sequence of responses to the weekly difficulty question are examined. These sequences are plotted in Figure 1. Each line in the graph represents the sequence for a single participant. The "*sq*" set of sequence analysis packages written for Stata by Kohler et al. (2006) were used to produce this plot.

The resulting array of sequences seems to indicate no systematic change in reported burden across the four weeks of participation. One pattern that might have been expected would be that respondents who were not initially burdened accumulate burden, echoing the fatigue observed to occur in some diary studies (Gerstel et al., 1980; Leigh, 1993; Verbrugge, 1980). Conversely, it might be expected that respondents who are initially burdened find themselves adapting to the task, and subsequently their reported levels of burden would decrease. Neither of these patterns is observed in the sequences presented in the graph in Figure 1.

To formally test whether there were any within individual trends in self-reported difficulty a fixed-effects regression model was estimated. One challenge that arises in fitting this model is how best to treat the large volume of missing reports that are present in the data. One approach is to treat these as a substantive category, indicative of high levels of burden, with the assumption that a high level of burden would cause a participant to be less likely to complete an end of week survey. A fixed effects regression including missing reports as a substantive category, representing the highest level of burden, produces a coefficient of $\beta = -0.03$, $p > 0.05$, 95% CI $[-0.11,\ 0.04]$. Excluding these missing

Figure 1: *Sequence analysis graph documenting the sequence of weekly reported difficulty participating in the Spending Study.*



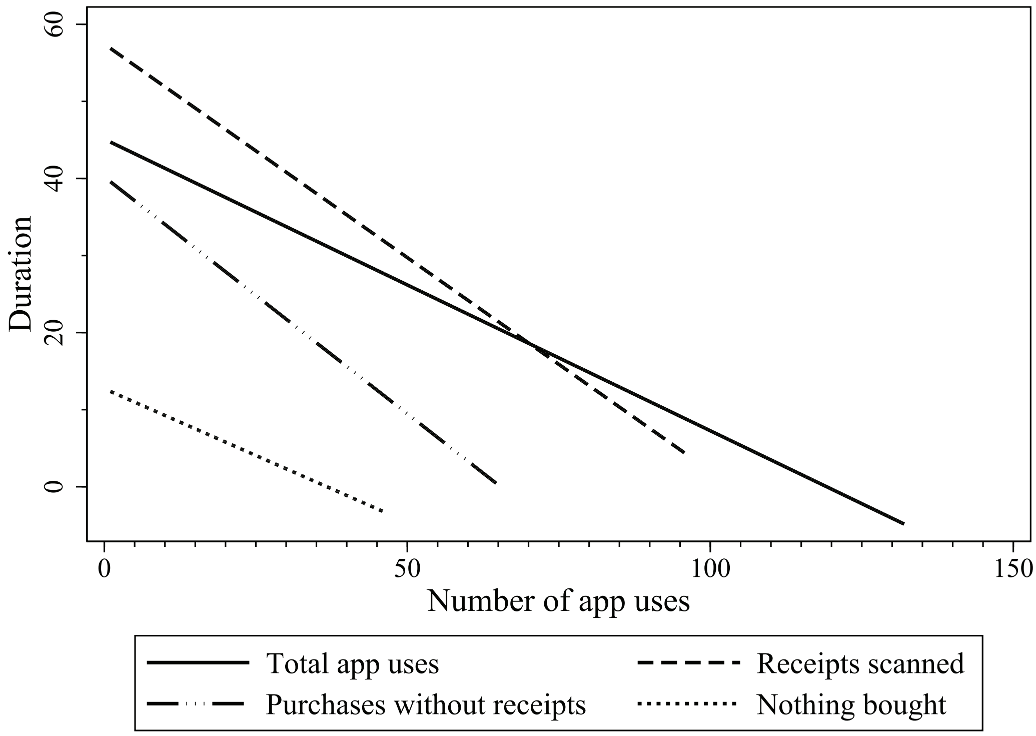
reports avoids the assumption that these are a substantive category of burden but results in an unbalanced panel. The resulting coefficient for a model excluding missing reports is $\beta = -0.01$, $p > 0.05$, 95% *CI* $[-0.06,\ 0.04]$. Neither of these specifications of the model produces a result that is indicative of an underlying pattern across time. This is consistent with the lack of a pattern present in the sequence analysis graph.

Objective burden

To examine the change in objective burden across the course of the study trends in the duration of app uses as a participant completes more app uses were modelled. The unit of analysis is app uses clustered within individuals. Fixed-effects models are fitted to look at the within individual changes. Four separate models were specified, one measuring the change across all app uses and three models measuring the changes within each of the three types of app use. Lines fitted for each of these four models are plotted in Figure 2. The overall trend was a decrease in the time it took to complete app uses with participants typically taking 0.4 seconds less to complete each subsequent app use

($\beta = -0.38$, $p < 0.001$, 95% CI $[-0.43, \ -0.33]$).

Figure 2: *Fixed-effects regression models of changes in app use duration as participation continues split by type of app use.*



The model was then repeated for each type of app use, with the predictor variable becoming the number of that type of app use that had been completed. The decision was made to run the models separately to help understand whether the overall trend was truly the product of decreases in time, or whether there was a compositional effect born out of shifting from the more time-consuming scanning of receipts to the other two less time-consuming methods. The results suggest that participants became quicker, around half a second with each subsequent app use for all three types of app use: $\beta = -0.55$, $\beta = -0.61$ and $\beta = -0.35$ for receipts scanned, purchases submitted without receipts, and submissions of nothing bought that day, respectively (95% CIs $[-0.68, \ -0.43], [-0.80 - 0.43]$ and $[-0.50, \ -0.19]$ respectively, all $p - values < 0.001$).

RQ3: *Does objective burden predict breaks in participation?*

To determine whether there was evidence that a higher objective burden resulted in temporary or permanent break-off Cox proportional-hazard regression models were fitted. Three models were specified, measuring breaks in participation in different ways. In the first model, the outcome variable is dropout from the Spending Study. Participants were considered to have dropped out (and thus exited from the analysis) after the last day on which they used the app within the 28 days from when they first used the app. There were therefore 223 spells, with one for each participant, running from when they began the study, until the last day on which the app was used.

The second model examined is the time until the first day on which the participant did not use the app. Again, there are 223 spells, this time running from when participants began the study until the first day on which the app was not used. Once the participant missed a day of app use they exit from the analysis.

The third model included repeated spells of participation: when a participant missed a day of app use a new spell began from the day they resumes using the app. Participants remained in the study throughout repeated spells of participation, with the exit condition for this model being dropout, as defined in the first model. This final model consists of 1374 spells. All three models use the Breslow method for handling tied failures (Breslow, 1974). The results of all three models are documented in Table 8.

The main predictor of interest is the average duration of app uses, up to that point in the study. This is a time varying measure, that is recalculated for each day. The proportions of app uses to date that are purchases without receipts and submissions of nothing bought are included as control variables. These are included because the three different types of app use differed in the amount of time taken to complete them. This could lead to a confounding compositional effect if participants have completed different proportions of different types of app uses.

Table 8: *Cox regression models examining whether objective burden is predictive of dropout or gaps in participation.*

	Dropout		First day missed		All days missed	
	HR	*SE*	*HR*	*SE*	*HR*	*SE*
Average duration	0.98*	0.01	1.00	0.00	1.00	0.00
Prop. of purchases without receipts	1.28	0.76	0.97	0.32	1.19	0.26
Prop. of submissions of nothing bought	1.29	0.97	1.89*	0.58	1.24	0.34
Wald χ^2	7.12		6.77		1.04	
n	223		223		223	
Spells	223		223		1374	

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

In terms of dropout, higher average time taken to complete app uses is associated with a lower risk of dropping out ($HR = 0.98$, $p < 0.05$). There is a 2% decrease in the expected hazard associated with a one second increase in average time taken to complete app uses. This is the opposite of the expected relationship. To better understand this result, it has been noted that it can be informative to convert hazard into a corresponding measure of effect size (Azuero, 2016). As the thresholds set out by Azuero consider only ratios that represent increases in hazard to compare the observed ratio to it is first necessary to calculate the inverse ratio, that is, the increase in hazard predicted by shorter average duration. This is given by one over the original hazard ratio, resulting in a value of ($HR = 1.02$); this falls below the suggested threshold for a small effect of 1.14, suggesting the observed effect may be inconsequential. Further doubt is cast on whether there is an effect of average duration on dropout when considering the full sample of 268 app users, where this result was not statistically significant ($HR = 1.00$, $p > 0.05$).

For both time until the first day missed, and time until each day missed the hazard ratio was not statistically significantly different dependent upon the average duration of app uses ($HR = 1.00$ and $HR = 1.00$ respectively, both $p - values > 0.05$). There was a higher risk of those participants with a higher proportion of reports of nothing bought initially missing a day of using the app ($HR = 1.89$, $p < 0.05$. It is possible that this was due to the task being less salient for these participants, as they were not making purchases as frequently. However, that there is not an equivalent increase in the risk of dropping out, or for the repeated spells model, suggests that this may have only been an issue as participants initially became used to participating.

RQ4: *What factors predict subjectiveand objective burden?*

Subjective burden

Table A1 in the Appendix shows the bivariate relationship between the predictors of burden and each of the four subjective measures of burden. Multivariate analyses were completed using four logistic regression models, with each of the four measures of subjective burden captured in the end of project survey as the dependent variable in one of the models. Each of the four dependent variables was coded such that 0 meant lower burden, and 1 meant an increased burden. The unit of analysis is the 223 participants. The results of the four models are documented in Table 9.

Throughout, where a statistically significant predictor is observed, this is compared to a series of thresholds for odds ratio values that correspond to recognised thresholds for effect size as measured by

Table 9: *Logistic regression models examining the multivariate relationship between predictors of burden and four measures of subjective burden.*

	Likelihood		Time/effort		Interest		Difficulty	
	OR	*SE*	*OR*	*SE*	*OR*	*SE*	*OR*	*SE*
£6 incentive treatment	0.96	0.36	0.99	0.32	1.22	0.38	1.61	0.53
Received additional incentive	1.18	0.54	1.56	0.71	0.95	0.45	0.77	0.3
Uses device for taking photos	5.34*	3.34	1.87	1.04	0.65	0.43	2.04	1.32
Uses device for online banking	0.53	0.19	0.60	0.21	0.80	0.32	0.52	0.28
Uses device to install apps	1.22	0.56	1.08	0.54	2.34	1.26	0.55	0.34
Willing to download app	0.78	0.43	2.45	1.32	1.68	0.75	1.37	0.71
Willing to use camera	0.46	0.28	0.30*	0.16	0.32	0.19	1.09	0.62
Checks balance once a week or more	0.80	0.29	1.03	0.38	0.48	0.21	1.90	0.78
Keeps a budget	0.87	0.31	0.86	0.24	0.84	0.23	1.88	0.55
Below the poverty threshold	2.51	1.36	0.65	0.34	0.59	0.31	2.43	1.55
Time constrained	0.73	0.26	0.91	0.29	0.81	0.30	0.77	0.26
Degree or higher	1.38	0.44	1.87*	0.54	1.86	0.62	1.39	0.39
Disabled/ long term illness	0.58	0.23	0.58	0.21	0.64	0.25	0.56	0.21
Female	1.05	0.35	0.76	0.22	1.18	0.35	0.89	0.26
Age	1.00	0.01	1.00	0.01	0.97**	0.01	1.01	0.01
Constant	0.35	0.28	1.01	0.86	14.75*	16.37	0.30	1.08
F	0.96		0.88		1.36		1.68	

Notes: $n = 223$; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Cohen's d. These thresholds are those set out by Cohen (1969) who suggests that $d = 0.20$, $d = 0.50$ and $d = 0.80$ represent a small, medium and large effect size respectively. The formula below, as set out by (Borenstein et al., 2009), allows the conversion of the threshold values of Cohen's d to log odds ratios, which can then be converted to odds ratios.

$$LogOddsRatio = d\frac{\pi}{\sqrt{3}} \tag{1}$$

This results in values of $OR = 1.44$, $OR = 2.48$ and $OR = 4.27$ corresponding to small, medium and large effect sizes respectively. To establish thresholds for odds ratios below one the inverse values for these effect size thresholds can be calculated by one over each respective value, resulting in $OR = 0.69$, $OR = 0.43$ and $OR = 0.23$, corresponding to small, medium and large effect sizes respectively.

Across all four models the two incentive treatments were not significant predictors of the respective measures of subjective burden. It is possible that this may be a result of so called 'ceiling effects' (Groves et al., 2000) as to the effectiveness of incentives in the presence of other motivating factors. This seems plausible given the seemingly high initial inhibitory threshold to participate (as suggested by the low response rate) together with relatively little variability in the level of self-reported burden.

Both perhaps suggest that participants had to be quite highly motivated to participate, thus meaning the additional effect of a larger incentive was negligible.

For all four models, downloading apps and online banking did not prove to be statistically significantly predictors of any of the four measures of subjective burden. However, using a mobile device to take photos did significantly increase the odds of reporting a lower likelihood of participating in the Spending Study if asked for the first time ($OR = 5.34$, $p < 0.05$), corresponding to a large effect size.

Gender, disability/long term illness, poverty and time constraints were not significant predictors across any of the four models. Participants who reported their highest level of education as a degree or higher had significantly higher odds of reporting that their time and effort was less well spent as compared to those with lower levels of education ($OR = 1.87$, $p < 0.05$) though this effect is seemingly small. This perhaps reflects a greater value placed upon their time by these participants.

Age was a significant predictor of interest, with older respondents reporting finding the study more interesting than younger respondents ($OR = 0.97$, $p < 0.01$). Though this was a seemingly negligible effect when comparing year to year, the effect was more substantial when comparing across a larger difference in age. For example, when comparing the first and third quartile of age ($Q_1 = 31$, $Q_3 = 53$) the odds ratio is $OR = 0.49$, a medium sized effect.

Willingness to download an app to complete survey tasks was not a significant predictor of any of the four measures of subjective burden. Willingness to use a camera to take photos or scan barcodes was a significant predictor of how well participants reported finding their time and effort spent participating. Those who reported being willing to use their camera to take photos for a data collection task had significantly lower odds of reporting lower levels of satisfaction with how well spent their time and effort was ($OR = 0.30$, $p < 0.05$) when compared to those who were not willing, again a medium sized effect.

Objective burden

The bivariate relationship between the predictors of burden and the time taken to complete app uses are documented in Table A2 in the Appendix. To understand which factors are predictive of the objective burden experienced by respondents the same covariates that were explored as predictors of subjective burden were included in a model with the duration of individual app uses as the dependent variable.

This shifted the unit of analysis from participants down to the level of individual app uses. A mixed effects regression model was used to account for the clustering of app uses within individual participants. The results from the model are presented in Table 10. Type of app use was included to control for the differences in typical durations of each of the three types of app use.

Table 10: *Mixed effects regression model examining the multivariate relationship between predictors of burden and the time taken to complete app uses.*

	β	*SE*
Six pounds incentive treatment	1.16	1.60
Received additional incentive	−0.12	1.76
Uses device for taking photos	0.18	3.94
Uses device for online banking	−2.60	2.23
Uses device to install apps	1.79	2.78
Willing to download app	-10.00^{*}	4.48
Willing to use camera	1.00	3.93
Checks balance once a week or more	6.49^{**}	2.27
Keeps a budget	−0.22	1.56
Below poverty threshold	4.96	4.14
Time constrained	−0.29	1.3
Degree or higher	−0.43	1.7
Disabled/ long term illness	2.03	1.68
Female	2.68	1.59
Age	0.44^{***}	0.05
Type of purchase		
Reference: (Scanned receipts)		
Purchase without receipt	-14.36^{***}	1.57
Nothing bought	-39.50^{***}	1.97
Constant	27.88^{***}	5.44
Wald χ^2	796.47^{***}	

Notes: $n = 10,179$ app uses, across 223 participants;$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Neither receipt of the higher initial incentive or receipt of the additional incentive proved to be a significant predictor of response times. This is not entirely surprising, it seems more plausible that if an effect of incentives were to be observed it would come when examining subjective burden, with the assumption that an increased incentive would lead to greater motivation, thus reducing the subjective burden of the task. However, it was considered possible that a larger incentive may have given the impression of greater importance of the task to respondents, thus potentially leading to greater care taken completing the task. These two covariates were retained for this reason, though it turns out there is no evidence of such a relationship.

Surprisingly, those participants who reported using their mobile devices for taking photos, online banking, or installing apps at IP9 were not significantly faster at completing app uses. It was expected that having these existing skills would reflect a greater competency in usage of mobile devices and that this would result in shorter app use durations.

Those respondents reporting a long-term illness or disability did not take longer to complete app uses, this perhaps can be explained by the fact that this variable encompasses a wide array of medical conditions, many of which may not be expected to have a direct impact upon participation. Cognitive ability, as measured by level of education, did not have a significant association, though it is unclear whether a better indicator of this characteristic would have revealed an association. Participants whose income fell below the poverty threshold were also not statistically significantly different in how long it took them to complete app uses.

In terms of reported willingness to perform survey tasks on mobile devices, willingness to download an app to complete survey tasks was found to be predictive of app use duration. Respondents who reported being willing were around 10 seconds faster ($\beta = -10.00$, $p < 0.05$) than those who reported not being willing to download a survey app. Surprisingly, willingness to use a camera for survey tasks, which is more directly tied to completing app uses, was not found to be a significant predictor of duration.

Age was found to be a significant predictor of the time taken to complete app uses, with each additional year older a participant was resulting in their app uses typically being just under half a second longer in duration ($\beta = 0.44$, $p < 0.001$). By again comparing the first and third quartiles of age ($Q_1 = 31,\ Q_3 = 53$) it is possible to get a better understanding of the effect of age on duration within the sample. The predicted duration for an individual at Q_3 compared to one at Q_1 is 9.70 seconds longer. One explanation for this is that it is consistent with evidence of a second-level digital divide in skills, with technical capability being less amongst older individuals (Loges and Jung, 2001).

When it comes to existing financial behavior keeping a budget did not have a significant effect upon the length of time it took respondents to complete app uses. However, checking one's bank balance more frequently did have an effect. Participants who checked their bank account at least once a week took just under 6 seconds longer to complete app uses than those who checked less frequently ($\beta = 6.49$, $p < 0.01$). It is possible this may reflect those respondents tying participation in the Spending Study to their existing tracking of their finances, thus taking greater care. Another possible explanation is that those who check their finances more regularly may take more of an active interest in their financial situation, and thus the data they are providing us with may also be of greater interest to them, meaning they spend more time reflecting on this. When considering all 268 participants in the study this relationship appears diminished ($\beta = 4.53$, $p > 0.05$); though in this instance this coefficient is just outside of the realm of statistical significance ($p = 0.052$).

Discussion

This paper examines respondent burden within the context of data collection using a mobile app situated as an additional study alongside an existing household panel survey. Whilst this paper is limited by the small number of participants who make up the final analytical sample, several contributions to the understanding of burden are made.

The results of RQ1 seem to support the notion that subjective and objective burden arise separately from one another. The four measures of subjective burden were strongly correlated with one another, and also showed strong evidence of mapping onto a latent variable that is seemingly consistent with an underlying concept of subjective burden. This highlights the potential for future use of multi-item scales to capture subjective perceptions of burden.

This was not the case for objective burden, where measures were less clearly related to one another. This is probably to be expected as these different measures are capturing objective burden in different ways. This highlights the importance of careful consideration when attempting to measure objective burden, as this can be considered either on an event level, or cumulatively across data collection.

That the three subjective measures not related to time spent participating are not strongly correlated with measures of time as proxies of objective burden is not surprising. This is consistent with previous research which has found a lack of correlation between measures of objective burden and subjective measures not explicitly asking about length (Sharp and Frankel, 1983; Oomens and Timmermans, 2008).

Perhaps more surprising is that the subjective measure asking about whether time and effort spent participating was well spent is also not strongly correlated with objective measures of time taken to complete app uses. Subjective measures asking about survey length have typically been found to have a strong association with objective length (Dale and Haraldsen, 2005; Sharp and Frankel, 1983). It is possible that the lack of correlation here may be a result of asking about effort as well as time (though this is the same as in the case of Sharp and Frankel); or it could reflect the disconnect between subjective and objective indicators of burden that has at times been observed (Oomens and Timmermans, 2008). More research is necessary to better understand the relationship between subjective and objective burden. Qualitative accounts of how objective burden feeds into subjective perceptions of a task may help to shed light on the relationship between experienced burden and subjective perceptions of burden.

In terms of how burden changes over time (RQ2) the results of the analysis of reported difficulty throughout the course of the study suggest that there is no evidence of systematic changes in subjective burden. It seems likely that in the case of the Spending Study this was because there was a high initial inhibitory threshold that was necessary to surpass to begin participating and that this may have resulted in subjective burden being typically quite low among participants, and indeed, this can be seen in the original distribution of the four subjective measures.

The time taken to participate showed consistent signs of decreasing as participation continued. This is reassuring, as it suggests that the objective burden of each task performed decreased as the number of tasks performed increased. What is less clear is whether this reduction in burden is the result of a learning effect with increases in participant ability, or whether participants were expending less effort to participate in the task, impacting on the quality of the data collected. Examination of indicators of data quality looking for evidence of satisficing behaviour would help to better understand the mechanism driving the reduction in time taken to participate.

The effect of cumulative burden on continued participation was small. Respondents who on average took longer to participate had a lower risk of dropping out (RQ3). However, this effect was minimal, and disappeared when considering all app users.

When it comes to uncovering which factors predict subjective and objective burden (RQ4) it seems clear that more work is necessary to help better identify these factors. This echoes the difficulties found in uncovering the characteristics which determine whether respondents experience fatigue in a diary study (Gillmore et al., 2001). That said, this paper does begin to find some evidence of the importance of certain factors.

Those who reported being willing to download an app to complete survey tasks using a mobile device turned out to be significantly faster at completing app uses. Likewise, those who reported being willing to use a camera to complete survey tasks were more likely to report their time and effort were well spent. This perhaps suggests respondents may be quite good at self-assessing what their experience of participating will be like, both in terms of the objective burden the task presents, and how they will subjectively perceive the burden of the task. This echoes the previous finding that willingness is predictive of propensity to respond (Jäckle, Burton, Couper and Lessof, 2017), with respondents who reported themselves as being very or somewhat willing to download an app to complete survey tasks being eight percentage points more likely to participate. That willingness should prove to be predictive of both participation, together with subjective and objective burden, is a positive argument for making use of hypothetical willingness questions to inform decisions about the use of alternative methods of survey data collection.

When it comes to looking at other factors that predict respondent burden, age proved to have an interesting relationship with different measures of burden. In the first instance, older participants tended to take longer to complete app uses than their younger counterparts. However, this did not translate to greater subjective burden amongst older participants, with age in fact being predictive of increased interest in the study. It is possible longer durations are indicative of reduced mobile technology skills amongst older participants (this is consistent with findings in the general population (Loges and Jung, 2001)), in which case it is reassuring that older participants still report finding the task interesting despite the greater objective burden.

One important caveat throughout is that the distribution of burden captured in the end of project survey does not fully reflect the full continuum of burden. For those respondents for whom the subjective burden was greatest it seems likely that they never surpassed the initial inhibitory threshold necessary to begin participating (for more information on who participated in the Spending Study see Jäckle, Burton, Couper and Lessof, 2017). In addition to this, the end of project survey does not fully capture burden even amongst participants. It seems plausible that those participants in the Spending Study who chose not to complete the additional end of project survey may have been amongst those most burdened by the task. In addition to this, the omission of the small portion of end of project respondents who did not receive the correct questionnaire version further contributes to an inability to account for the full spectrum of burden. Future research into respondent burden may benefit from finding ways of considering burden for both respondents and non-respondents.

In summary, this paper contributes to the understanding of respondent burden, and specifically to the context of burden in data collection using mobile devices. Generally, subjective burden was unrelated to the objective burden of the task. Further research is needed to provide a better account of the factors leading to burden, and how those factors affect subjective burden both for sample members when they decide whether to participate or not, and, conditional on participation, the decisions made about subsequent reporting behaviour.

References

Abraham, K., Maitland, A. and Bianchi, S. M. (2006), 'Non-response in the american time use survey: Who is missing from the data and how much does it matter?', *NBER Technical Working Paper Series* (328).

Ampt, E. (2003), Respondent burden, *in* P. Stopher and P. Jones, eds, 'Transport Survey Quality and Innovation', Emerald Group Publishing, Bingley, West Yorkshire, UK., pp. 507–521.

Armstrong, J. S. (1975), 'Monetary incentives in mail surveys', *Public Opinion Quarterly* **39**(1), 111–116.

Azuero, A. (2016), 'A note on the magnitude of hazard ratios', *Cancer* **122**(8), 1298–1299.

Bartlett, M. S. (1951), 'The effect of standardization on a χ^2 approximation in factor analysis', *Biometrika* **38**(3-4), 337–344.

Borenstein, M., Hedges, L. V., Higgins, J. P. and Rothstein, H. R. (2009), *Introduction to Meta-Analysis*, John Wiley & Sons, Chicago, IL.

Bradburn, N. (1978), Respondent burden, *in* 'Proceedings of the Section on Survey Research Methods', American Statistical Association, Alexandria, VA, pp. 35–40.

Branden, L., Gritz, R. and Pergamit, M. (1995), The effect of interview length on nonresponse in the national longitudinal survey of youth, *in* 'Proceedings of the 1995 Census Bureau Annual Research Conference', Arlington, VA, pp. 129–154.

Breslow, N. (1974), 'Covariance analysis of censored survival data', *Biometrics* **30**(1), 89–99.

Busemeyer, J. R. and Townsend, J. T. (1993), 'Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment', *Psychological review* **100**(3), 432–459.

Church, A. H. (1993), 'Estimating the effect of incentives on mail survey response rates: A meta-analysis', *Public Opinion Quarterly* **57**(1), 62–79.

Cohen, J. (1969), *Statistical power analysis for the behavioral sciences*, Academic Press, New York, NY.

Collins, M., Sykes, W., Wilson, P. and Blackshaw, N. (1988), Diffusion of technological innovation: Computer assisted data collection in the u.k., *in* R. Groves, P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II and J. Waksberg, eds, 'Computer assisted survey information collection', John Wiley & Sons, New York, NY, book section Nonresponse: The UK experience.

Couper, M. P. (1997), 'Survey introductions and data quality', *Public Opinion Quarterly* **61**(2), 317–338.

Couper, M. P. and Nicholls, II, W. L. (1998), The history and development of computer assisted survey information collection methods, *in* M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II and J. M. O'Reilly, eds, 'Computer assisted survey information collection', John Wiley & Sons, New York, NY.

Crawford, S. D., Couper, M. P. and Lamias, M. J. (2001), 'Web surveys', *Social Science Computer Review* **19**(2), 146–162.

Dale, T. and Haraldsen, G. (2005), Embedded evaluation of perceived and actual response burden in business surveys, *in* D. Hedlin, T. Dale, G. Haraldsen and J. Jones, eds, 'Developing methods for assessing perceived response burden', Eurostat, Luxembourg, pp. 112–125.

Dillman, D. A., Sinclair, M. D. and Clark, J. R. (1993), 'Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys', *Public Opinion Quarterly* **57**(3), 289–304.

Dyregrov, K. (2004), 'Bereaved parents experience of research participation', *Social science & medicine* **58**(2), 391–400.

Flora, D. B. and Curran, P. J. (2004), 'An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data', *Psychological methods* **9**(4), 466–491.

Fricker, S. and Tourangeau, R. (2010), 'Examining the relationship between nonresponse propensity and data quality in two national household surveys', *Public Opinion Quarterly* **74**(5), 934–955.

Galea, S., Nandi, A., Stuber, J., Gold, J., Acierno, R., Best, C. L., Bucuvalas, M., Rudenstine, S., Boscarino, J. A. and Resnick, H. (2005), 'Participant reactions to survey research in the general population after terrorist attacks', *Journal of Traumatic Stress* **18**(5), 461–465.

Galesic, M. (2006), 'Dropouts on the web: Effects of interest and burden experienced during an online survey', *Journal of Official Statistics* **22**(2), 313–328.

Galesic, M. and Bosnjak, M. (2009), 'Effects of questionnaire length on participation and indicators of response quality in a web survey', *Public Opinion Quarterly* **73**(2), 349–360.

Gerstel, E. K., Harford, T. C. and Pautler, C. (1980), 'The reliability of drinking estimates obtained with two data collection methods', *Journal of Studies on Alcohol* **41**(1), 89–94.

Gillmore, M. R., Gaylord, J., Hartway, J., Hoppe, M. J., Morrison, D. M., Leigh, B. C. and Rainey, D. T. (2001), 'Daily data collection of sexual and other healthrelated behaviors', *Journal of Sex Research* **38**(1), 35–42.

Goyder, J. (1994), 'An experiment with cash incentives on a personal interview survey', *Journal of the Market Research Society* **34**(4), 1–7.

Graf, I. (2008), Respondent burden, *in* P. Lavrakas, ed., 'Encyclopedia of survey research methods', Sage Publications, Thousand Oaks, CA, p. 740.

Groves, R. M. and Couper, M. P. (1998), *Nonresponse in household interview surveys*, John Wiley & Sons, New York, NY.

Groves, R. M., Presser, S. and Dipko, S. (2004), 'The role of topic interest in survey participation decisions', *Public Opinion Quarterly* **68**(1), 2–31.

Groves, R. M., Singer, E. and Bowers, A. (1999), 'A laboratory approach to measuring the effects on survey participation of interview length, incentives, differential incentives, and refusal conversion', *Journal of Official Statistics* **15**(2), 251–268.

Groves, R., Singer, E. and Corning, A. (2000), 'Leverage-saliency theory of survey participation: description and an illustration', *Public Opinion Quarterly* **64**(3), 299–308.

Guadagnoli, E. and Velicer, W. (1988), 'Relation of sample size to the stability of component patterns', *Psychological Bulletin* **103**, 265–75.

Haraldsen, G. (2004), 'Identifying and reducing response burdens in internet business surveys', *Journal of Official Statistics* **20**(2), 393410.

Hinkle, D. E., Wiersma, W. and Jurs, S. G. (2003), *Applied statistics for the behavioral sciences*, Houghton Mifflin, Boston, MA.

Holgado-Tello, F. P., ChacnMoscoso, S., BarberoGarca, I. and VilaAbad, E. (2010), ‘Polychoric versus pearson correlations in exploratory and confirmatory factor analysis of ordinal variables’, *Quality & Quantity* **44**(1), 153–166.

Jäckle, A., Burton, J., Couper, M. P. and Lessof, C. (2017), ‘Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: response rates and response biases’, *Understanding Society Working Paper Series* **2017**(9).

Jäckle, A., Gaia, A., Al Baghal, T., Burton, J. and Lynn, P. (2017), Understanding society: The uk household longitudinal study innovation panel, waves 1-9, user manual, Report, ISER, University of Essex.

Kaiser, H. F. (1960), ‘The application of electronic computers to factor analysis’, *Educational and Psychological Measurement* **20**(1), 141–151.

Kaiser, H. F. (1970), ‘A second generation little jiffy’, *Psychometrika* **35**(4), 401–415.

Kaiser, H. F. and Rice, J. (1974), ‘Little jiffy, mark iv’, *Educational and Psychological Measurement* **34**(1), 111–117.

Knäuper, B., Belli, R. F., Hill, D. H. and Herzog, A. R. (1997), ‘Question difficulty and respondents’ cognitive ability: The effect on data quality’, *Journal of Official Statistics* **13**(2), 181–199.

Kohler, U., Luniak, M. and Brzinsky-Fay, C. (2006), ‘sq. [stata package]’.
URL: *http://www.stata-journal.com/software/sj6-4*

Kolenikov, S. (2008), ‘polychoric. [stata package]’.
URL: *http://www.komkon.org/ tacik/stata/*

Kreuter, F., Presser, S. and Tourangeau, R. (2008), ‘Social desirability bias in cati, ivr, and web surveys the effects of mode and question sensitivity’, *Public Opinion Quarterly* **72**(5), 847–865.

Krosnick, J. A. (1991), ‘Response strategies for coping with the cognitive demands of attitude measures in surveys’, *Applied Cognitive Psychology* **5**(3), 213–236.

Leigh, B. C. (1993), ‘Alcohol consumption and sexual activity as reported with a diary technique’, *Journal of Abnormal Psychology* **102**(3), 490–493.

Lemmens, P. H. H. M., Knibbe, R. A. and Tan, F. (1988), ‘Weekly recall and diary estimates of alcohol consumption in a general population survey’, *Journal of studies on alcohol* **49**(2), 131–135.

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J. and Langer Tesfaye, C. (2014), ‘Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the aapor task force on emerging technologies in public opinion research’, *Public Opinion Quarterly* **78**(4), 779–787.

Loges, W. E. and Jung, J. (2001), ‘Exploring the digital divide: Internet connectedness and age’, *Communication research* **28**(4), 536–562.

Lynn, P. (2014), ‘Longer interviews may not affect subsequent survey participation propensity’, *Public Opinion Quarterly* **78**(2), 500–509.

Maydeu-Olivares, A. and D’zurilla, T. J. (1995), ‘A factor analysis of the social problem-solving inventory using polychoric correlations’, *European Journal of Psychological Assessment* **11**(2), 98–107.

Newman, E., Willard, T., Sinclair, R. and Kaloupek, D. (2001), 'Empirically supported ethical research practice: The costs and benefits of research from the participants view', *Accountability in Research* **8**(4), 309–329.

Office of Management and Budget (2006), Standards and guidelines for statistical surveys, Report, Office of Management and Budget.

Oomens, P. and Timmermans, G. (2008), The dutch approach to reducing the real and perceived administrative burden on enterprises caused by statistics, Paper present to the 94th DGINS Conference, Vilnius, Lithuania.

Persky, H., Strauss, D., Lief, H. I., Miller, W. R. and O'Brien, C. P. (1981), 'Effect of the research process on human sexual behavior', *Journal of Psychiatric Research* **16**(1), 41–52.

Peytchev, A. (2009), 'Survey breakoff', *Public Opinion Quarterly* **73**(1), 74–97.

Raiklin, E. and Uyar, B. (1996), 'On the relativity of the concepts of needs, wants, scarcity and opportunity cost', *International Journal of Social Economics* **23**(7), 49–56.

Roberts, C., Eva, G., Allum, N. and Lynn, P. (2010), 'Diffusion of technological innovation: Computer assisted data collection in the U.K.', *ISER Working Paper Series* **2010**(36).

Ruch, F. L. (1941), 'Effects of repeated interviewing on the respondent's answers', *Journal of Consulting Psychology* **5**(4), 179–182.

Ruzek, J. I. and Zatzick, D. F. (2000), 'Ethical considerations in research participation among acutely injured trauma survivors: An empirical investigation', *General Hospital Psychiatry* **22**(1), 27–36.

Searles, J. S., Perrine, M., Mundt, J. C. and Helzer, J. E. (1995), 'Self-report of drinking using touch-tone telephone: extending the limits of reliable daily contact', *Journal of Studies on Alcohol* **56**(4), 375–382.

Sharp, L. M. and Frankel, J. (1983), 'Respondent burden: A test of some common assumptions', *Public Opinion Quarterly* **47**(1), 36–53.

Singer, E., van Hoewyk, J., Gebler, N. and McGonagle, K. (1999), 'The effect of incentives on response rates in interviewer-mediated surveys', *Journal of Official Statistics* **15**(2), 217–230.

University of Essex. Institute for Social and Economic Research (2017), 'Understanding society: Innovation panel, waves 1-9, 2008-2016. [data collection]. 8th edition. uk data service. sn: 6849, http://doi.org/10.5255/ukda-sn-6849-9'.

Verbrugge, L. M. (1980), 'Health diaries', *Medical Care* **18**(1), 73–95.

Walker, E. A., Newman, E., Koss, M. and Bernstein, D. (1997), 'Does the study of victimization revictimize the victims?', *General Hospital Psychiatry* **19**(6), 403–410.

Willeboordse, A. (1997), Minimizing response burden, *in* A. Willeboordse, ed., 'Handbook on design and implementation of business surveys', Eurostat, Luxembourg, pp. 111–118.

Yammarino, F. J., Skinner, S. J. and Childers, T. L. (1991), 'Understanding mail survey response behavior a meta-analysis', *Public Opinion Quarterly* **55**(4), 613–639.

Young, J. M., O'Halloran, A., McAulay, C., Pirotta, M., Forsdike, K., Stacey, I. and Currow, D. (2015), 'Unconditional and conditional incentives differentially improved general practitioners' participation in an online survey: randomized controlled trial', *Journal of Clinical Epidemiology* **68**(6), 693–697.

Young, P. V. and Schmid, C. F. (1956), *Scientific social surveys and research : an introduction to the background, content, methods, principles, and analysis of social studies*, Prentice-Hall, Englewood Cliffs, NJ.

Yu, E., Fricker, S. and Kopp, B. (2015), Can survey instructions relieve respondent burden?, Paper presented to the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Zagorsky, J. L. and Rhoton, P. (2008), 'The effects of promised monetary incentives on attrition in a long-term panel survey', *Public Opinion Quarterly* **72**(3), 502–513.

Appendix

Table A1: *Pearson χ^2 tests examining the bivariate relationship between predictors of burden and four measures of subjective burden.*

	Likelihood		Time/effort		Interest		Difficulty	
	χ^2	F	χ^2	F	χ^2	F	χ^2	F
£6 incentive treatment	0.36	0.1	1.16	0.50	1.16	0.50	5.11	1.65
Received additional incentive	1.99	0.61	2.25	0.95	0.46	0.20	2.10	0.70
Uses device for taking photos	1.97	0.64	0.66	0.35	0.29	0.17	1.23	0.43
Uses device for online banking	4.11	1.44	0.79	0.42	0.58	0.29	3.72	1.20
Uses device to install apps	1.23	0.41	0.04	0.02	1.96	1.08	3.75	1.23
Willing to download app	11.55	1.36	3.3	0.54	2.76	0.49	12.17	1.38
Willing to use camera	14.72	1.71	6.21	0.99	3.08	0.52	15.16	1.69
Checks balance once a week or more	2.94	1.00	1.51	0.79	1.3	0.65	3.52	1.26
Keeps a budget	3.22	1.00	0.20	0.10	1.44	0.69	5.17	1.84
Below the poverty threshold	11.20	3.03*	1.88	0.86	0.7	0.29	5.6	1.47
Time constrained	8.76	3.32*	0.28	0.13	0.91	0.38	1.10	0.36
Degree or higher	2.87	1.03	4.49	2.52	6.94	3.20*	1.50	0.55
Disabled/ long term illness	3.78	1.19	3.30	1.48	2.59	1.08	4.02	1.41
Female	1.13	0.36	1.04	0.51	3.51	1.78	2.72	0.94

Table A2: *Two-tailed t-tests examining the bivariate relationship between predictors of burden and a measure of objective burden, the time taken to complete app uses.*

	$\bar{x_1} - \bar{x_2}$	SE	t
£6 incentive treatment	−0.65	2.18	−0.30
Received additional incentive	−0.13	2.16	−0.06
Uses device for taking photos	7.52	4.57	0.10
Uses device for online banking	7.07**	2.23	3.17
Uses device to install apps	5.67	2.89	1.96
Willing to download app	9.35**	3.23	2.89
Willing to use camera	5.60	3.00	1.86
Checks balance once a week or more	0.63	1.98	0.32
Keeps a budget	1.58	2.06	0.77
Below the poverty threshold	−3.64	4.33	−0.84
Time constrained	4.47*	2.06	2.16
Degree or higher	−0.11	2.04	−0.06
Disabled/ long term illness	−2.92	2.10	−1.39
Female	−2.53	1.85	−1.37