Understanding Society

Working Paper Series

No. 2013 – 05

August 2013

Propensity to Consent to Data Linkage: Experimental Evidence from the Innovation Panel on the Role of Three Survey Design Features

Emanuela Sala[1], Gundi Knies[2], Jonathan Burton[2]

[1] Dipartimento di Sociologia e Ricerca Sociale, Università degli Studi di Milano Bicocca

[2] 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

Data linkage is a procedure that links individual level administrative data to survey data. The procedure has a huge potential to inform further analyses and help us get a better picture of those who are using services provided by the government. One obstacle to realizing the full research potential is that not all survey respondents are happy for their data to be linked, and that some types of respondents are more likely to withhold consent than others. This means researchers often times cannot produce reliable results for specific groups of interest.

We know from previous research that the way in which we design the questionnaire and ask the question can influence the answer we get which means that there may be scope to achieve higher consent rates when we make changes to the questionnaire design. To investigate this empirically, we designed a number of experiments around the way in which we ask for consent to data linkage and implemented them on the UK Household Longitudinal Study: Innovation Panel.

We find that the survey design features which we hypothesised may lead to higher consent rates are indeed associated with higher consent propensities. In line with the hypothesis that you get better responses when the question is more salient, respondents were more likely to consent when consent to link to social security records held by DWP is asked in a block of questions relating to income and benefit receipt rather than at the end of the questionnaire. They were also more likely to consent if we asked them in a later wave of the survey when they had built up greater trust in the survey organization, and those who we have asked for consent before were more likely to consent if we reminded them of their previous decision. The most common reason for giving consent was wanting to be "helpful with the research", and the most common reason for withholding consent was concerns about sharing confidential data with third parties. Concerns over data sharing were also present among consenters.

Overall, the findings suggest that consent rates may be higher when the benefits of the data linkage are made clearer and if concerns over data sharing are alleviated. Whilst some of this may be achieved by carefully designing the consent question, a more effective interviewer training may also be considered.

Propensity to Consent to Data Linkage: Experimental Evidence from the Innovation Panel on the Role of Three Survey Design Features

Emanuela Sala[1], Gundi Knies[2], Jon Burton[2]

[1] Dipartimento di Sociologia e Ricerca Sociale, Università degli Studi di Milano Bicocca

[2] Institute for Social and Economic Research, University of Essex

Abstract: When performing data linkage, survey respondents need to provide their informed consent. Since not all respondents agree to this request, the linked dataset will have fewer observations than the survey dataset alone and bias may be introduced. By focusing on the role that survey design features play in gaining respondents' consent, this paper provides an innovative contribution to the studies in this field. Analysing experimental data collected in a nationally representative household panel survey of the British population, we find that interview features such as question format (dependent/independent questions) and placement of the consent question within the questionnaire have an impact on consent rates.

JEL classifications: C9 C18 C81

Key words: Data linkage, asking for consent, experiments, dependent interviewing, context effects, longitudinal studies, Innovation Panel.

Corresponding author's contact details: Emanuela Sala (emanuela.sala@unimib.it), Dipartimento di Sociologia e Ricerca Sociale, Università di Milano Bicocca, Italy.

Acknowledgments: We are grateful to Peter Lynn and participants of the International Panel Methods Workshop 2012 and the ESRA 2013 conference for their comments on a preliminary version of this manuscript.

1. Introduction

Linkage of person-based administrative data to survey data is becoming increasingly popular both in the UK and elsewhere[1]. Among the reasons why data linkage is so appealing to researchers is the potential to overcome some of the main challenges currently facing survey practitioners, for example, improving data quality, reducing survey costs in the longer term and easing respondent (and interviewer) burden. However, successful implementation of data linkage between survey data and person-based administrative data is a complex process.

One of the main hurdles in realising the full potential of linked survey and administrative data is the requirement, common to many countries, that survey respondents give their informed consent before the survey organisation can share the respondent's personal information with the data custodians of the administrative data for them (or a third party) to identify the survey respondent's record in their data base and to send the authorised information back to the survey team[2]. Since not all respondents agree to this request, the linked dataset will have fewer observations than the survey dataset alone and bias may be introduced.

A number of empirical studies have been published in recent years, examining consent rates and consent bias. Whilst the bulk of the literature is based on community health studies (for a comprehensive review, see Kho, Duffett, Willison, Cook, & Brouwers, 2009), a number of representative population studies have also reported findings on selectivity in consent (see, e.g., Knies, Burton, & Sala, 2012). Both strands focus in particular on bias with respect to respondent characteristics. By contrast, there is as yet very little methodological research on how the design of the questionnaire instrument may assist in achieving higher consent rates and help attenuate potential consent bias. Survey design decisions such as where to position the consent question in the questionnaire are often based on practical and operational considerations rather than on sound empirical evidence. The knowledge gap is particularly marked in the context of longitudinal studies where a number of additional design decisions arise. For example, in which wave of the survey should respondents be asked the consent question? Should the request be made early on in the life of the panel when the study has

[1] Major social surveys have linked their data ad hoc with a wide range of administrative data including data on benefit receipts, adolescents' school performance, health and morbidity (e.g., the US Current Population Survey, the Canadian Survey of Labour Income Dynamics, the UK Household Longitudinal Study, the German Labour Market and Social Security Study).

[2] In order to be valid, the decision to consent or not consent must be made by the respondent alone and there must not be any coercion. The respondent must be given full information about what their decision involves, including the benefits and risks, and they must have the capacity to understand the information provided to them.

suffered less from attrition which means that a greater number of people are asked for consent, or later on in the panel when fewer people participate in the study but their compliance, and potentially the consent rate, may be higher? Is there an impact on attrition of asking for consent to data linkage? How can respondents be re-asked or reminded of any consents they have given in the past in an ethical way but without jeopardizing the quality and quantity of linked data?

This study contributes to significantly enhance the current knowledge about how to ask for informed consent to data linkage by reporting empirical evidence from random treatment-control experiments on the performance of some of the aforementioned survey design options, many of which pertain specifically to longitudinal studies. More specifically, we analyse experimental data collected in a nationally representative household panel survey for Great Britain (the UK Household Longitudinal Study: Innovation Panel), which allow us to investigate empirically the trade-off between asking early on in the life of a panel and holding back the request until more rapport has been built, and to examine the effect of implementing different design options for confirming (or re-asking) consent. Moreover, we provide further empirical evidence on the effect of the question ordering, and provide a deeper understanding of the reasons why people give or withhold consent.

The findings suggest that interview features such as question format (dependent/independent questions) and placement of the consent question within the questionnaire have an impact on consent rates. We also find evidence that suggests that specific interviewer training and carefully drafted question wording help alleviate concerns about data linkage and therefore have a positive effect on consent rates. The study also provides practical guidance to survey methodologists and survey agencies on the implementation of eliciting consent to data linkage.

2. Linking administrative data to survey data. What are the drivers of consent?

Much of the survey research on data linkage has focused on identifying the correlates of respondents' propensity to consent[3]. In a nutshell, studies that have examined which respondent characteristics are associated with consent have typically found some association

[3] A small number of studies have taken the survey methodological work somewhat further. Jenkins et al. (2008) have examined the performance of different matching criteria; Sakshaug and Kreuter (2012) analyse selectivity in linked data; McKay (2013) has analysed selectivity in a linked dataset and provided guidelines for statistical remedies for dealing with bias.

with socio-demographic and socio-economic characteristics but there were no consistent drivers of consent across studies. The only exception is with commonly accepted markers of survey co-operation, altruism and trust. These are associated with a greater propensity to consent across studies (for a review see Korbmacher and Schroeder, 2013, and Sala, Burton & Knies, 2012).

Moreover, while a number of studies have found interviewer effects, when specific interviewer characteristics were considered in multivariate models, few of them were associated with consent. There was little agreement in findings across studies, albeit the evidence base in this field is very small. We are aware of only four studies. Korbmacher and Schroeder (2013) find that the interviewers' age, experience and performance matter whereas Sala, Burton, & Knies (2012), testing a similar range of characteristics, find that the interviewers' task-specific experience is associated with consent but not their experience more generally. Sakshaug et al. (2012) also find no evidence for interviewer demographic effects. A consistent finding across the studies investigating interviewer effects is that there is no empirical support for the hypothesis that interviewer attitudes and personality matter. This is true both when we look at interviewer personality traits and attitudes to persuading respondents more generally (see, e.g., Sala, Burton, & Knies, 2012) and also when considering their more specific attitude to consent to data linkage: Sakshaug et al. (2013) find no difference in obtaining consent between interviewers who would themselves consent to data linkage and those who would be reluctant to consent.

The main focus of the present study lies in examining whether design and implementation of the consent instrument affect consent rates. Although much understudied, a number of studies have considered the association between interview features and consent. For example, Jenkins et al. (2008) find that the length of the interview (interpreted as a proxy for interviewer-respondent rapport) and the level of understanding of survey questions are predictors of consent. Sala, Burton & Knies (2012) find that survey "fidelity", the interview sequence and the number of consents that have already been given by other household members are related to consent. To our knowledge there is only one study which considers specific design features of the consent instrument on the consent rate. Sakshaug et al. (2013) find that the consent rate was 9.6 percentage points higher among those who were asked for consent at the start of a telephone interview compared to those who were asked at the end. The study also finds no effect on consent rates of mentioning data linkage as a route to reduce the burden on the respondent ("To keep the interview as brief as possible…").

3. Research hypotheses

When talking about interview features we will refer, more specifically, to three aspects which are relevant to the case of data linkage, especially when it is performed in a longitudinal context: the location of the consent question in the questionnaire (context effects), the time in the life of the panel when the consent question is asked and the type of question that is used to phrase the consent question.

3.1 The placement of the consent question

Research has shown that responses to survey questions may be influenced by prior questions (McClendon and O'Brien, 1988; Schuman and Presser, 1981; Tourangeau, Rips, & Rasinski, 2000). A question which causes a respondent to consider a particular subject may affect the way that they respond to a subsequent question. This has been found to occur with general well-being questions (McClendon and O'Brien, 1988) and fear of crime questions (Yang and Hinkle, 2012). The phenomenon has been referred to as the "context effect" or the "question ordering effect".

We aim to investigate whether consent rates vary according to the placement of the consent question. As we mentioned in the introduction, survey design decisions regarding the placement of the consent questions are often based on practical and operational decisions. Where the consent question is accompanied by an information leaflet for the sample member to read, and a form which needs to be signed to record consent, the consent question is typically asked at the end of the interview so as not to break up the flow of the interview. However, we argue that consent rates may be higher when the consent question is asked after a series of questions on a similar topic ("in context") than when the consent question is asked at the end of the questionnaire. The underlying mechanism is that if the request for consent appears in a context, this makes the request more salient, hence, the respondent will be more likely to agree to data linkage. Therefore, we test whether asking for consent to link to administrative data about the receipt of state benefits after a section in the interview that asks about these benefits leads to a higher level of consent than asking at the end of the interview. We hypothesise that having just been asked, and answering, a series of questions about the receipt of a large number of state benefits and payments, the respondent will be more likely to consent to linkage to data about those benefits. This may be because the respondent will want to appear more consistent ("assimilation effect") or want to reduce future redundancy

and short-cut the questionnaire (e.g., “If I consent to this, I might not be asked these questions again next year”).

3.2 The time in the life of the panel when the consent question is asked

In the case of longitudinal studies, responses to survey questions may be influenced by a number of factors, including answers to questions administered in previous interviews (i.e., panel conditioning) and the time in the life of the panel when a specific question is asked. We intend to explore whether consent rates vary according to the wave of the panel in which the consent question is asked. Similar to the previous research question, studies on this topic are lacking. Our hypothesis is that consent rates are higher for respondents who are interviewed later in the life of a panel. Research has shown that asking for consent to link survey data to administrative data may be a sensitive topic (Sala, Burton, & Knies, 2012), therefore higher consent rates may be gained when respondents have developed trust in the survey organization and are engaged in the survey, i.e., later in the life of the panel.

3.3 The type of survey questions

It is a well-known fact in the survey literature that responses to survey questions also vary according to the type of question respondents are administered as well as to the question wording (see, for example, Belli et al. 1999; Prohaska, Brown, & Belli (1998); Schuldt, Konrath, & Schwarz, 2011; Singer at al. 2010; Tourangeau, Rips, & Rasinski, 2000). We aim to investigate whether consent rates vary by the type of survey question respondents are administered. This research question is also driven by practical motivations, as ethical guidance in many countries requires survey organizations to periodically give sample members a chance to change their minds about consent. There are a number of ways this can be done and different data holders may have different preferences.

In a longitudinal context we may distinguish between dependent and standard independent questions. Dependent interviewing (DI) is a standardized questioning method particular to longitudinal surveys that is widely used on major surveys internationally. It uses data gathered in previous interviews with the respondent to formulate question text. This practice can be distinguished from the standard independent interviewing (INDI), which makes no reference to data previously collected to phrase questions or route respondents through questionnaires (Mathiowetz and McGonagle 2000; Lynn et al. 2006).

We hypothesize that overall consent rates to data linkage may vary according to the type of question respondents are asked. We also argue that responses given in the past may also play a key role in the mechanisms that lead respondents to consent in a later wave. In particular, we believe that respondents tend to agree with information that is fed-forward from previous interviews, regardless of their specific content. This hypothesis is driven by two considerations. First, there is evidence that survey respondents like to be consistent when they are responding to survey questions (Groves, Cialdini, & Couper, 1992). If they answer in a contradictory way, they may appear to the interviewer to be indecisive or give the impression that they had answered 'wrongly' in the past. Thus, by reminding respondents of their response, they are likely to give the same answer (in our case, a yes or a no to a consent question). When respondents are asked the question independently, they have no reminder of their previous answer and so they are able to make the decision at the time without worrying about consistency. Second, there is evidence that shows that DI questions may facilitate the response process and ease respondent burden (Sala, Burton, & Knies, 2012). As argued by Tourangeau (1984), the response process is structured in four main steps: (1) understanding the question; (2) retrieving the relevant information; (3) making a judgment; and (4) selecting a response. In the case of consent to data linkage question, where respondents need to process difficult information and make a decision in relatively little time, DI may affect the second and the third step. By reminding the respondent of their previous response they are given an opportunity to short-cut this cognitive process by giving them an easy response; to agree with their previous answer. This short-cut is justified by the respondent 'trusting' their earlier thought processes, rather than thinking through the response from the beginning.

4. Data

We use data collected in the fourth round of interviews of the Innovation Panel (IP). The analyses also draw on the longitudinal nature of the study by including information from previous waves, mainly wave 1.

The IP is a longitudinal household panel study, representative of the population living in Great Britain in 2008. Interviews take place annually. The IP is part of *Understanding Society*: The UK Household Longitudinal Study (UKHLS), one of the major investments in the social sciences research infrastructure in the UK[4]. It is managed on behalf of the ESRC by

[4] More details on Understanding Society can be found at https://www.understandingsociety.ac.uk/.

the Institute for Social and Economic Research (ISER) at the University of Essex. It is a resource for carrying out innovative longitudinal experimental and methodological research. It is used, for example, to test different fieldwork designs, new questions and new ways of asking existing questions (for a review, see Budd et al. 2012, Burton et al. 2011). Findings from Innovation Panel inform the design of *Understanding Society* as well as other longitudinal surveys worldwide.

The IP sample is a clustered, stratified and equal probability design of almost 1,500 households (at wave 1, in 2008). At IP4, a refreshment sample of 960 issued households was added to the original sample. The achieved sample at IP4 consisted of 910 households and 1,456 adults in the original sample, and 464 households and 723 adults in the refreshment sample.

The standard IP design, in terms of questionnaire content and sample following rules, is modelled on the main-stage of *Understanding Society*. The survey collects a wide range of information including job and education, fertility histories, health conditions, personal finances, social and political participation, and social attitudes[5].

At IP4, eligible adults were interviewed using Computer Assisted Personal Interviewing (CAPI) and either a paper self-completion questionnaire or a self-completion questionnaire carried out using the lap-top (Computer Assisted Self Interviewing, CASI). Young respondents (aged 10-15) were interviewed using a paper self-completion questionnaire[6].

At IP4, thirteen experiments were implemented. They include the investigation of the impact of incentives on response rates, efficiency of fieldwork and costs, the evaluation of a new contact approach in order to improve fieldwork efficiency, the comparison of potentially ambiguous questions with improved versions, the assessment of the impact of question placing and phrasing, and an investigation of panel conditioning in addition to the experiments covered in this paper (for a review of the main findings, see Budd et al., 2012).

4.1 Collecting informed consent to data linkage in IP

The Innovation Panel offers a valuable opportunity to address some of the outstanding research questions around best practices for collecting informed consent, in particular those paramount in longitudinal study designs. The decision to seek respondents' consent to data

[5] The IP4 questionnaire is available at https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires.
[6] Further information on the IP can be found at https://www.understandingsociety.ac.uk/about/innovation-panel.

linkage in the IP was driven by scientific motivations as well as practical considerations such as the relatively low consent rates to data linkage in the British Household Panel Survey (BHPS) (Knies, Burton, & Sala, 2012; Sala, Burton, & Knies, 2012) and the plans to implement data linkage on the associated Understanding Society study. To address some of the concerns and the research questions of the IP design team, a detailed plan to ask for consent to link respondents' survey data to a wide range of administrative records was developed and a number of experiments were designed and implemented over time.

The process of asking for consent in the IP is similar to the one implemented in the BHPS and described in detail in Knies, Burton, & Sala (2012). In a nutshell, there is a brief preamble to the consent question which informs the respondent that the study would like to add information from specific (named) administrative records to the responses given in the interview; there is an information leaflet which provides further information on what the data linkage involves, and, in order to give permission, the respondents need to sign a consent form. Consent is asked at the end of the interview so that signing the forms and reading the information leaflet does not interrupt the flow of the interview. The outcome of the consent question is recorded in CAPI and a copy of the signed consent forms is kept by the respondent and the original is collected by the survey organisation, reconciled against the data and then sent to ISER for secure storage for future reference. The IP4 protocol differs from the standard process of collecting informed consent for a number of aspects relating to the experimental manipulation.

Consent to data linkage was collected at a number of different stages. At IP1, adults in a random two-thirds of households were asked for their written permission to link to benefit data held by the Department for Work and Pension (DWP) and HM Revenue and Customs (HMRC). This was to evaluate whether there is an effect of asking for consent to data linkage on attrition, which is a major concern for longitudinal studies. IP1 also asked adults aged 16-24 and adults responsible for children aged 4-15 for their permission to link to education records. At IP2, adult respondents and adults responsible for children aged 0-15 years were asked for their written permission to link to administrative health data. Some respondents (i.e., new entrants, IP1 non-consenters, etc.) were also asked for their consent to the education data linkage. Note that there was no experimentation with consent protocols in IP2 or IP3. At IP4, all adult respondents were asked for consent to the benefit data linkage.

Table 1 reports the consent rates achieved on IP as recorded in CAPI. It is important to note that IP respondents may revoke their consent anytime and there are currently no plans to perform data linkages for consenters.

Table 1
Percentage of the sample agreeing to administrative data linkage on Innovation Panel, Waves 1-4.

Administrative data type	Wave 1	Wave 2	Wave 3	Wave 4
National Insurance contributions, benefits and tax records, savings and pensions[A]	56.9 [a]			62.2
Education: 4-15 year olds[B]	65.1 [b]	55.6 [c]		
Education: 16-24 year olds[B]	69.0 [b]	78.9 [c]		
Health: 0-15 year olds[C]		72.6 [d]		
Health: 16+ year olds[C]		79.6 [d]		

Notes: Unweighted data
[A] From the Department of Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC).
[B] From the Department for Children, Schools and Families (DCSF) at IP1, and the English Department for Children, Schools and Families, the Welsh Department for Children, Education, Lifelong Learning, and Skills, the Scottish Government Education Directorate, or the Department of Education / Education and Skills Authority in Northern Ireland at IP2.
[C] From the National Health Service (NHS), Department of Health, General Registration Office and the Office for National Statistics.
[a] Experimental allocation of two-thirds of the sample.
[b] Asked of 16-24 year olds, plus the responsible adult for children aged 4-15.
[c] For new entrants or those who had not consented at IP1 and were aged 4-24.
[d] All responding adults, plus the responsible adult was asked for consent for children aged 0-15.

4.2 Experimenting with different ways of asking for consent at IP4

To address our research questions we developed three experiments:

Experiment 1. Context effects

IP4 adult respondents (aged 16+) were randomly allocated to two treatment groups[7]: one group were asked for their consent at the end of the questionnaire (control group, "at the end"), the other group were asked for consent after a module of questions which asked about the receipt of state benefits and other payments (treatment group, "in context").

[7] In practice, all random allocations were at the household level, and all adults within a household were in the same experimental treatment group.

Experiment 2. Time in the life panel effects

IP1 adult respondents were randomly allocated to two treatment groups: two-thirds of the sample (control group) were asked for consent at IP1; the remaining one-third of the sample were to be asked for consent later in the life of the panel, i.e., at IP4 (treatment group).

Note that there was an error in the implementation of this experiment. In the first two months of fieldwork *all* IP1 respondents were asked for consent. Implications for evaluating Experiment 2 are discussed in the analysis methods section.

Experiment 3. Question wording effects

IP4 respondents who had been asked for consent to link to benefit data at IP1 were randomly allocated to two treatment groups: one group were asked the consent question independently, i.e., they were not reminded of their previous answer (control group, "INDI"), the other group were reminded of their IP1 response and were asked if they were (still) (un)willing to allow the data linkage (treatment group, "DI"). The allocation to this experiment was done independently to the context effect experiment.

The wording of the verbal consent question for all experimental groups is provided in Appendix 1. Respondents who gave verbal consent were then asked for their written consent in a follow-up question, which was administered at the end of the interview including for those asked for consent 'in context' (i.e., the treatment group of Experiment 1)[8].

In addition, all respondents were asked a follow-up question on reasons for consenting or not consenting to data linkage at the end of the interview. Upon the administration of the follow up question, the interviewer coded whether or not the respondent had changed their mind. Last but not least, some sections of the IP4 interviews were audio-recorded, including the consent to data linkage question (consent to audio-recoding: 68.4 per cent)[9].

[8] Overall, 3.9 per cent of all IP4 respondents who provided verbal consent did not provide written consent. Respondents who had given consent at IP1, were dependently asked at IP4 and confirmed their consent were not asked to sign a consent form (again).

[9] Data from the audio recordings is currently prepared for analysis under the project "Understanding non-response on Understanding Society", funded by the NCRM.

5. Methods of analysis

To address the research questions, we use both bivariate and multivariate logistic regression analysis. The dependent variable is a dummy variable that indicates whether respondent has provided verbal consent to perform data linkage: a value of 1 indicates that consent has been provided and a value of 0 indicates consent was withheld. The key independent variable is an indicator of the experimental treatment groups.

In the bivariate analysis we compare consent rates at the relevant waves, for the relevant respondents and samples. The analysis plan is described in details in the last column of Table 2. A standard t-test is used to test for differences in consent rates. In the regression analysis, we include additional control variables such as age, gender, employment status, net earnings, number of times the respondent was interviewed prior to IP4, a proxy for cognitive skills (the level of understanding of the questionnaire) and the type of considerations while deciding about data linkage.

Note that due to an implementation error of Experiment 2 in IP1, for a robustness check, we will restrict the IP4 analysis sample to respondents who were interviewed in the last sample months. Respondents in the IP1 experimental treatment groups may be viewed as those most difficult to get hold of so we will need to compare them with the most difficult to get hold of respondents in IP4.

Results from the bivariate and multivariate analysis consider the complex sampling design of Innovation Panel: results are weighted for unequal selection probabilities as well as non-response. For results drawing on just the IP4 sample (assessment of Experiment 1, analysis of reasons for consent/non-consent) we use cross-sectional weights; for results drawing on the continuing IP1 sample (assessment of Experiments 2 and 3), we use longitudinal population weights. The analysis is carried out using Stata Version 12.1 (StataCorp, 2012). To account for the complex survey design, we use the svy suite of commands.

Table 2
Overview of the design of IP4 consent experiments

Experiment	Treatment group		Assessment
	Control (C)	Treatment (T)	
Experiment 1. Context effects	Consent question asked at the end of the questionnaire	Consent question asked after the benefit module	Comparisons of consent rates at IP4 between the C and T groups. Eligible sample for C: IP4 respondents allocated to C responding at IP4 (N=1,114) Eligible sample for T: IP4 respondents allocated to T, responding at IP4 (N=1,065) The refreshment sample is included in the analysis.
Experiment 2. Time of the life panel effects	Benefit consent question asked at IP1	Benefit consent asked at IP4	Comparisons of consent rate between the C and T groups respectively at IP1 and IP4. Eligible sample for C: continuing IP1 respondents in IP4 allocated to C group responding at IP4 (N=1,096). Eligible sample for T: continuing IP1 respondents in IP4 allocated to T, responding at IP4 (N=174)
Experiment 3. Question wording effects	Independent interviewing question asked	Dependent interviewing question asked	Comparisons of consent rates at IP4 between the T and C groups, also broken down by the respondents' previous answers to the IP1 consent question. Eligible sample for C: IP4 respondents allocated to C responding at IP4 (N=510) Eligible sample for T: IP4 respondents allocated to T responding at IP4 (N=578) Refreshment sample is excluded from the analysis

6. Results

Table 3 shows the results of the first experiment, the placement in the interview of the consent request.

Table 3

Consent rates for respondents asked in context (treatment group) and asked at the end of the interview (control group), Experiment 1.

	Consent rate	S.E.	95per cent Confidence Interval		N
			Lower bound	Upper bound	
Asked at the end	0.58	0.03	0.52	0.64	1,114
Asked in context	0.65*	0.02	0.60	0.69	1,065

Notes: Standard errors adjusted for the complex survey design. Results weighted for unequal selection probabilities and non-response. Sample includes all IP4 adult respondents. Differences in experimental group means * significant at the 0.1 level, ** significant at the 0.05 level, *** significant at the 0.01 level.

Source: Innovation Panel, Waves 1-4, IP4 release.

We find evidence that consent rates vary by the position of the consent question: respondents who were asked the consent question "in context" are, on average, 7 percentage points more likely to consent than respondents who are asked for consent at the end of the questionnaire (65 per cent compared to 58 per cent, two-sample $t(60)=2$, $p=0.05$). The result is robust also when we absorb further population heterogeneity (i.e., when we include in a logistic regression model controls for age, gender, employment status, net earnings, as well as how often the respondent has given an interview (all not statistically significant), respondent suspicion (negative association with consenting) and good understanding of the questionnaire (positive association with consent), but becomes not statistically significant if we include markers for whether the respondent mentioned any concerns or considerations when deciding on whether or not to consent to data linkage (results reported in Appendix 2).

Table 4 reports the result of Experiment 2 which was started in IP1 and concluded in IP4. There is some indication that consent varies by the stage in the life of a panel in which the data linkage question is asked (at least over the four-year-period that we are looking at). A greater share of continuing IP1 respondents who were first asked for consent at IP4 consented to economic record linkage (71 per cent) than was true for IP1 respondents who were asked at

IP1 (60 per cent). The difference is statistically significant at the 90 per cent level, t(60)=-1.66, p=0.10.

Table 4
Consent rates for respondents asked in first interview (treatment group) and asked at the fourth interview (control group), Experiment 2.

	Consent rate	S.E.	95 per cent Confidence Interval		N
			Lower bound	Upper bound	
Experimental groups					
Asked at first interview (IP1)	0.60	0.02	0.55	0.65	1,096
Asked at the fourth interview (IP4)	0.71*	0.05	0.60	0.82	174
Excluding first sample months					
Asked at first interview (IP1)	0.53	0.05	0.44	0.63	358
Asked at the fourth interview (IP4)	0.64	0.08	0.48	0.81	54
Base consent rates					
IP1 sample in IP1	0.57	0.02	0.53	0.61	2,073
IP4 refreshment sample in IP4	0.61	0.03	0.55	0.66	723

Notes: Standard errors adjusted for the complex survey design. Results weighted for unequal selection probabilities and non-response (and attrition). Differences in experimental group means * significant at the 0.1 level, ** significant at the 0.05 level, *** significant at the 0.01 level.

Source: Innovation Panel, Waves 1-4, IP4 release.

However, allocation to experimental groups did not take place during the first two months of fieldwork in IP1, due to a survey implementation error. This means that treatment (i.e., being asked later in the life of the panel, e.g., at wave 4) is disproportionate among respondents who were interviewed at a later stage of fieldwork. Such respondents tend to be more difficult to interview because of their busy schedules or greater reluctance to participate. It may well be that this confounds the expected positive effect of rapport because we try to identify this effect among the most difficult to get respondents who may be the least responsive to such a treatment effect. In support of the argument, when we exclude from the IP1 sample respondents who were interviewed in the first two months, and from the continuing IP1 sample those interviewed in the first month[10] we find that the consent rate in these groups is overall lower. However, the difference in the consent rate between the two groups is virtually

[10] Very few interviews on IP4 took place outside the first month of fieldwork; excluding the first two months of IP4 would leave less than 10 respondents to evaluate the experiment.

unchanged; whilst 64 per cent of continuing IP1 respondents who were first asked for consent at IP4 give consent, the figure amounts to 53 per cent in the group who were asked for consent at IP1. The difference in means is not statistically significant, t(60)=-1.17, p=0.25.

Table 5 reports the results of the DI experiment. Non-consenters at IP1 had a 22 percentage point higher probability to consent in IP4 if they were not reminded of their decision in IP1 (i.e., 46 compared to 24 per cent), t(60)= 3.14, p=0.00. Consenters at IP1 had a 26 percentage point higher probability to consent when reminded that they had consented to the linkage in IP1, t(60)= -5.86, p=0.00. In other words, respondents tend to be consistent with their previous decision when reminded of that decision. These results are robust to including further controls in multivariate regression models, see Appendix 3.

Table 5
Consent rates for respondents asked in first interview and who were reminded of their previous decision (treatment group) or asked consent without reminding them of their previous decision (control group), Experiment 3.

	Consent rate	S.E.	95 per cent Confidence Interval		N
			Lower bound	Upper bound	
Non-consenters at IP1					
Asked IND	0.46	0.06	0.35	0.57	193
Asked DI	0.24***	0.05	0.13	0.34	219
Consenters at IP1					
Asked IND	0.68	0.04	0.59	0.76	317
Asked DI	0.94***	0.02	0.91	0.97	359

Notes: Standard errors adjusted for the complex survey design. Results weighted for unequal selection probabilities and non-response (and attrition). Differences in experimental group means * significant at the 0.1 level, ** significant at the 0.05 level, *** significant at the 0.01 level.

Source: Innovation Panel, Waves 1-4, IP4 release.

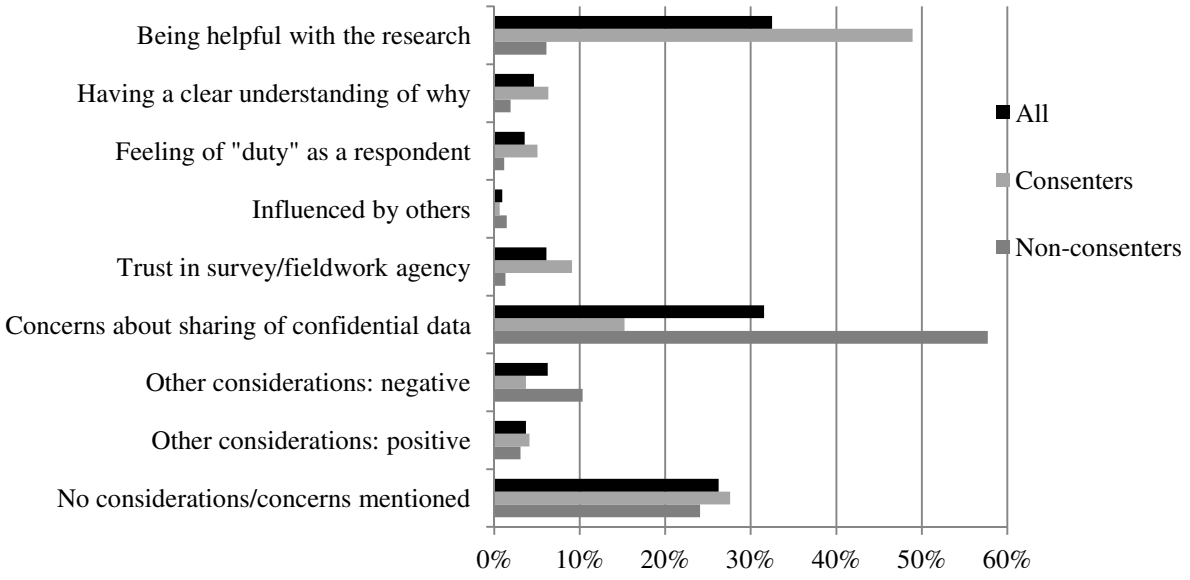
6.1 Why do people (withhold) consent to data linkage?

To further explore the mechanisms that influence respondents' consent, we also looked at the reasons they gave while agreeing or disagreeing to data linkage. Respondents were asked about what they considered when they gave their response. The exact question wording was:

"Different things can be important when deciding to give consent to add information from DWP administrative records to survey data. What were you considering when answering?". Respondents could name more than one reason, and the interviewer coded their response to a pre-set list of categories with two "other" categories where the reason was recorded by the interviewer verbatim.

Graph 1 shows the reasons for agreeing or disagreeing to the data linkage request by consent status.

Graph 1
Reasons for agreeing or disagreeing to data linkage by consent status



Source: Innovation Panel, Wave 4. All results consider the complex survey design and are weighted.

A number of findings clearly stand out. First, a sizable share of the sample (26.2 per cent) did not mention any considerations or concerns they had when making their decision and this is not associated with consent status. Second, 32 per cent of the sample expressed concerns about sharing confidential data with third parties whilst a similar proportion mentioned that they considered being "helpful" with research. Third, as one may expect, there are significant differences in the nature of the considerations expressed by consenters and non-consenters ($p<0.01$). For example, 58 per cent of non-consenters expressed concerns regarding sharing of confidential data compared to 15 per cent of those who consented, and 49 per cent of consenters wanted to be helpful compared to 6 per cent of non-consenters.

It is worth noting that just over one in six people who consented still had concerns about confidentiality. Fourth, when focusing on the reasons mentioned by consenters, just under one in ten of those who gave consent (9.1 percent) said they considered their trust in the fieldwork agency or survey organization whereas 6.3 per cent mentioned it was because they clearly understood why and how the linkage would take place, compared to 1.9 per cent of those who declined to consent. This demonstrates the importance that the reasons why linking survey responses to administrative data helps research, and the process by which the information is linked, are available to the respondent. Their 'duty' as a respondent was mentioned as a consideration by 5 per cent of those who gave consent whilst this aspect was considered by only 1.2 per cent of the non-consenters. Note that some proportion of non-consenters and consenters mentioned other reasons than those anticipated by the research team based on the literature on survey participation and consent, both positive (4 per cent) and negative (6.8 per cent).

7. Conclusions

Data linkage is an increasingly popular survey feature, decisions regarding its implementation, however, are seldom based on empirical evidence. Very few guidelines have been produced and shared practices on how to best implement this complex process, especially in a longitudinal context. For example, we still do not know where to locate the consent question in a questionnaire to maximise consent rates and reduce bias. One of the reasons for this lack of knowledge lies in the scarcity of experimental data available.

This research sets out to evaluate the role that a number of interview features play in the consent process drawing on a unique set of experiments carried out in the framework of a national panel study of the British population; the Innovation Panel. We focused on three aspects of the consent process; the location of the consent question within the questionnaire, the time in the life of the panel in which the consent question should be asked and the question wording of the consent question. These are some of the key issues that survey designers have to face while implementing data linkage. We also collected additional non-experimental information from consenters and non-consenters on reasons for consent.

A number of findings clearly stand out from our analysis. First, drawing on the contextual explanation, we hypothesized that consent rates may vary by the position of the consent question within the questionnaire. In particular, we state that when the consent question to link survey data to economic records is asked after a series of questions on benefits receipts

(i.e., a context where the request is salient), consent rates may be higher. This hypothesis finds some support in the empirical data. When asked "in context" consent rates are 7 per cent point higher than when the consent question is asked at the end of the questionnaire (significant at the 0.1 level).

However, one may argue that this finding is also consistent with the "survey fatigue" explanation. Towards the end of the questionnaire the respondent may want to hurry the interview along because of the length of the questionnaire. They may therefore be less inclined to spend time reading an information leaflet and consent form and giving the matter their full consideration. Unfortunately, we cannot disentangle this explanation from the contextual explanation since in our treatment group the benefits module always appeared at the same place in each interview. If our finding is confirmed by other similar studies, it may be advisable to ask for consent in a relevant context rather than at the end. In our study we focussed on consent to data linkage to benefit records. Further research should investigate, for example, whether the relationship between consent rates and the location of the consent question holds when looking at other domains (e.g., health or education) and should further explore the mechanisms in place. Experiments with manipulations of the relevant questionnaire section may be designed to contribute to an understanding of the reasons that lead respondents to consent when the request to data linkage is asked in a relevant context.

Second, we intended to test whether consent rates varied by the wave in which the request to consent to data linkage is asked. Comparing consent rates obtained at wave 1 to those obtained later in the life of the panel, i.e., wave 4, we find some indication that consent rates may be higher when asked later in the life of the panel. The implications of this finding are not straightforward as pros and cons are associated with the decision to ask for consent at the first wave or at a later wave in the life of the panel. Despite the increased consent rate elicited when asking for consent at a later stage in the life of a panel study, it may be advisable to ask for consent as early in the life of the panel as is possible when the larger sample size (before attrition) results in more individuals giving consent, compared to a higher consent rate further into the life of a panel when attrition has reduced the sample. This is at least true as long as asking for consent does not have an impact on attrition. In our case, we did not find that being asked the consent question in wave 1 influenced participation in wave 4 (N=2,399, b-coefficient -.035, S.E.=0.117). There has, however, been a sizable rate of attrition with only about 50 per cent of interviewees at IP1 being re-interviewed in IP4. We believe that a possible strategy for maximising the number of linked data would be to ask at the first wave,

and then to re-ask those who did not give consent at a later wave. As with our first experiment further research is needed before clear guidelines on this issue could be provided. The main limitation of the study is the implementation error at IP1 that may weaken the impact of our findings. Such errors are likely to occur when the data collection is commissioned to third parties and researchers have little control on how, in practice, the experiment is implemented and carried out. To minimise the occurrence of such errors, one may evaluate the introduction of particular norms in the contract that regulate this aspect.

Third, we evaluated whether the question wording, i.e., dependent and independent questions, has an effect on consent rates. When previous consenters and non-consenters are administered the DI question, we find the highest and the lowest consent rates, respectively 94 and 24 per cent. This suggests that respondents tend to be consistent with their previous answers when answering survey questions. We can speculate, according with findings from other studies (Sala et al., 2011), that DI may facilitate the response process. The results from this experiment lead us to formulate the following recommendation. When having to re-collect consent to data linkage, we have shown that reminding people of their earlier decisions prompts them to make the same decision. Thus, to maximise the number of people for whom consent is retained, a strategy would be to remind those who had previously given consent whilst those who have not given consent in the past are asked an independent question. This strategy may not always be implemented as different ethical committees may have different requirements and they may not necessary agree with the suggested recommendation.

Last but not least, when looking at the reasons for agreeing or disagreeing to data linkage, two important findings stand out. First, the results suggest that higher consent rates may be achieved if the consent question wording highlights, for example, the research potential that data linkage opens up. Second, if interviewers are able to reassure respondents, concerns about confidentiality need not lead to a refusal to consent. Concerns about confidentiality are the main reason given by those who withhold consent. Thus, improving messages about data security may be important in easing these concerns. Overall, these findings demonstrate the importance that the reasons why linking survey responses to administrative data helps research, and the process by which the information is linked, are available to the respondent. An effective interviewer training programme, with a focus on how to deal with major concerns on data security, may contribute to increase consent rates.

8. References

Belli, R. F., Traugott, M. W., Young, M., McGonagle, K. A. (1999). Reducing vote overreporting in surveys: Social desirability, memory failure, and source monitoring. *Public Opinion Quarterly,* 63, 90-108.

Budd, S., Gilbert, E., Burton, J., Jäckle, A., Kaminska, O., Uhrig, S.C. N., Brown, M., Calderwood, L. (2012). *Understanding Society Innovation Panel Wave 4: Results from methodological experiments.* Understanding Society Working Paper Series No. 2011 – 05, Colchester, University of Essex.

Burton J., Budd S., Gilbert, E. , Jäckle A., L. McFall, S. L., Uhrig, S. C. N. (2011). *Understanding Society Innovation Panel Wave 4: Results from methodological experiments.* Understanding Society Working Paper Series No. 2012 – 06, Colchester, University of Essex.

Groves, R. M., Cialdini, R.B.,Couper, M.P.(1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, 475-495.

Jenkins, S.P., Cappellari, L., Lynn, P., Jäckle, A., Sala, E. (2006). Patterns of consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society, Series A,* 169(4), 701-722.

Jenkins, S.P., Lynn, P., Jäckle, A., Sala, E. (2008) The feasibility of linking household survey and administrative record data: New evidence from Britain. *International Journal of Social Research Methodology*, 11(1), 29-43.

Kho, M. E., Duffett, M., Willison, D. J., Cook, D. J., Brouwers, M. C. (2009). Written informed consent and selection bias in observational studies using medical records: Systematic review. *British Medical Journal*, 338:b866.

Knies, G., Burton, J., Sala, E. (2012). Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population. *BMC Health Services Research*, 12(1): 52.

Korbmacher, J. M., Schroeder M. (2013). Consent when linking survey data with administrative records: The role of the interviewer. *Survey Research methods*, 7 (2), 115-131.

Lynn, P., Jäckle, A., Jenkins, S.P., Sala, E. (2006). The effects of dependent interviewing on responses to questions on income sources. *Journal of Official Statistics*, 22(3), 357-384.

Mathiowetz, N., McGonagle A. (2000). An assessment of the current state of dependent interviewing in household surveys, *Journal of Official Statistics*, 16, 401-418.

McClendon, M.J. and O'Brien, D.J., (1988). Question-order effects on the determinants of subjective well-being. *Public Opinion Quarterly*, 52, 351-364.

McKay, S. (2012). Evaluating approaches to Family Resources Survey data linking, DWP Working Paper, No 110, London: Department for Work and Pensions.

Prohaska, V., Brown, N. R., Belli, R. (1998). Forward telescoping: The question matters. *Memory, 6,* 455-465.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., Weir, D. (2012). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods and Research*, 41, 535-569.

Sakshaug, J. W., Kreuter, F. (2012). Assessing the magnitude of non-consent biases in linked survey and administrative data. *Survey Research Methods*, 6 (2), 113-122.

Sakshaug, J. W., Tutz, V., Kreuter, F. (2013). Placement, wording and interviewers: Identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7 (2), 33-144

Sala, E., Burton, J., Knies, G. (2012) Correlates of obtaining informed consent to data linkage. Respondent, interview, and interviewer characteristics. *Sociological Methods Research,* 41 (3), 414-439.

Schuldt, J. P., Konrath, S. H., Schwarz, N. (2011). "Global warming" or "climate change"?: Whether the planet is warming depends on question wording. *Public Opinion Quarterly* 75(1): 115-124.

Schuman, H., Presser, S. (1981). *Questions and answers in attitude surveys: experiments on question form, wording, and context*. New York: Academic Press.

Singer, E., Couper, M. P., Raghunathan, T. E., Antonucci, T. C., Burmeister, M., Van Hoewyk, J. (2010). The effect of question framing and response options on the

relationship between racial attitudes and beliefs about genes as causes of behavior. *Public Opinion Quarterly* 74(3): 460-476.

StataCorp (2012). *Statistical Software: Release 12.1*. College Station, TX, Stata Corporation.

Tourangeau, R. (1984). Cognitive science and survey methods: A cognitive perspective. In T. Jabine, M. Straf, J. Tanur, R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.

Tourangeau, R., Rips, L. J., Rasinski K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

Yang, S., and Hinkle, J.C. (2012). "Issues in Survey Design: Using Surveys of Victimization and Fear of Crime as Examples" in Handbook of Survey Methodology for the Social Sciences, Springer, New York.

9. Appendix

Appendix 1

Consent question wording for experimental treatment groups

	Independent question	Dependent question
Gave consent at IP1	Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. Are you happy for us to do so?	Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. According to our records, when we interviewed you in 2008, you gave us permission to do so. Are you still happy for us to do so?
Did not give consent at IP1		Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. According to our records, when we interviewed you in 2008, you declined that we do this. Are you willing to give your consent now?
Not asked for consent at IP1		Not applicable

Appendix 2

Logistic regression of consent on demographic characteristics, and some interview features.

	Model 1.1		Model 1.2		Model 1.3		Model 1.4		Model 1.5	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Asked for consent in context	0.26+	0.13	0.27*	0.13	0.27*	0.13	0.32*	0.14	0.26	0.15
Age			-0.01	0.01	-0.01	0.02	-0.02	0.02	0.00	0.02
Age squared			0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00
Female			-0.04	0.10	-0.03	0.10	-0.06	0.11	-0.08	0.13
Employment status (comparison group: employed)										
self-employed					-0.22	0.42	-0.19	0.42	-0.19	0.58
unemployed					0.11	0.40	0.08	0.40	-0.45	0.57
pensioner					-0.04	0.39	-0.10	0.40	-0.07	0.53
other					-0.20	0.33	-0.18	0.35	-0.38	0.49
Usual monthly net earnings (in £)					-0.01	0.03	-0.01	0.03	-0.01	0.04
Number of times interviewed							-0.00	0.06	-0.05	0.07
Respondent was suspicious							-2.39*	0.23	-1.74*	0.23
Good understanding of questionnaire							1.03*	0.25	1.26*	0.32
Considerations in decision about to data linkage										
being helpful with the research									1.96*	0.27
clarity over what is requested									0.77	0.62
duty as a respondent									-0.03	0.57
dl_influ									-0.96	0.89
trust in the survey agency									1.51*	0.43
concerns over data sharing									-1.60*	0.21
Other negative consideration									-0.80*	0.28
Other positive consideration									0.33	0.39
Constant	0.34*	0.12	0.88*	0.37	0.98*	0.44	0.42	0.54	-0.15	0.68
Number of observations	2,157		2,157		2,157		2,157		2,157	

Notes: Standard errors adjusted for the complex survey design. Results weighted for unequal selection probabilities and non-response. Sample includes all IP4 adult respondents. Significant at * 95%, + 90%.

Source: Innovation Panel, Waves 1-4, IP4 release.

Appendix 3: Logistic regression of asking for consent dependently on demographic characteristics (N=1,091).

	Model 3.1		Model 3.2		Model 3.3		Model 3.4		Model 3.5	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
IP1 consent X INDI (Non-consenters asked INDI)										
Consenters asked INDI	0.88*	0.29	0.85*	0.28	0.86*	0.28	0.80*	0.28	0.63+	0.34
Non-consenters asked DI	-1.03*	0.34	-1.08*	0.33	-1.21*	0.32	-1.08*	0.32	-1.07*	0.38
Consenters asked DI	2.91*	0.32	2.86*	0.32	2.88*	0.32	2.95*	0.34	2.55*	0.42
Age			-0.04	0.05	-0.03	0.05	-0.07+	0.04	-0.02	0.04
Age squared			0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00
Female			-0.03	0.17	-0.07	0.16	-0.20	0.17	-0.42*	0.20
Employment status (employed)										
self-employed					-0.34	1.05	-0.50	1.22	-1.46	1.48
unemployed					1.08	1.07	0.68	1.18	0.50	1.40
pensioner					0.85	0.97	0.55	1.15	0.33	1.42
other					-0.40	0.91	-0.37	1.15	-0.82	1.45
Usual monthly net earnings (in £)					-0.00	0.08	-0.01	0.09	-0.03	0.12
Number of times interviewed							0.03	0.23	-0.02	0.24
Respondent was suspicious							-3.37*	0.70	-2.72*	0.70
Good understanding of q'aire							1.78*	0.53	1.73*	0.57
Considerations about to linkage										
being helpful with the research									1.69*	0.42
clarity over what is requested									0.25	0.63
duty as a respondent									15.08*	0.73
dl_influ									-18.95*	1.42
trust in the survey agency									2.47*	0.62
concerns over data sharing									-1.95*	0.32
Other negative consideration									-0.86	0.56
Other positive consideration									0.03	0.51
Constant	-0.13	0.23	1.07	1.21	1.12	1.28	0.75	1.45	0.51	1.44

Notes: See previous table.