Session Number: Parallel Session 2B
Time: Monday, August 23, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview and Interviewer Characteristics

Emanuela Sala, Jon Burton, Gundi Knies

For additional information please contact:

Name: Gundi Knies
Affiliation: ISER, University of Essex

Email Address: gknies@essex.ac.uk

CORRELATES OF OBTAINING INFORMED CONSENT TO DATA LINKAGE: RESPONDENT, INTERVIEW AND INTERVIEWER CHARACTERISTICS

Emanuela Sala, Jon Burton, Gundi Knies

ISER, University of Essex

Abstract

In the UK, in order to link individual-level administrative records to survey responses, a respondent needs to give their written consent. Respondents' propensity to consent has been shown to be associated with respondent characteristics and survey design features. This paper explores whether characteristics of the interviewer or the interview process itself also influence consent. We use the BHPS combined with a survey of interviewers to model the probability that respondents consent to adding health and social security records to their survey responses. We find that interviewer characteristics, including personality and attitudes to persuading respondents, are not associated with consent. By contrast, some respondent characteristics and characteristics of the interview process within the household seem to matter.

Keywords: Consent, Data linkage, BHPS, Administrative records, Interviewer characteristics.

Acknowledgments: We are grateful to colleagues at the Institute for Social and Economic Research and to participants at the Institute of Education-ISER Joint Seminar on data linkage for their comments on previous drafts of the paper. Special thanks go to Mark Bryan, Stephen Jenkins, Peter Lynn and Stephanie McFall. This project is funded by the ESRC grant RES-586-47-0002.

INTRODUCTION

Linkage of administrative data to survey data is becoming increasingly popular both in the UK and elsewhere.[1] It is regarded as a powerful tool to overcome some of the main challenges currently facing survey practitioners. Administrative records offer a wealth of information which could significantly enhance research opportunities, help improve data quality, reduce survey costs and ease respondent (and interviewer) burden.

One of the paramount challenges to this end is gaining respondents' informed consent to their data being linked. Willingness to give consent is not universal, reducing the number of observations and potentially introducing bias. In fact, previous studies have shown that consenters and non-consenters vary on socio-economic characteristics (for recent reviews see Dunn et al. 2004; Huang et al. 2007; Tate et al. 2006), and there is some empirical evidence that consent is also associated with features of the data collection process and with study characteristics; including the survey topic, the domain of the data linkage (Jenkins et al. 2006; Singer et al. 2003) and who is asking for consent (Armstrong et al. 2008). Systematic research on these issues is as yet scant, even in the survey methodology literature. In particular, there are very few studies that look at consent bias on general population surveys, or that explore potential differences across different domains of data linkage.[2] Moreover, no study has investigated which role specific interviewer characteristics and survey design features play in the consent process.

Our paper makes an important contribution to the existing literature by systematically examining consent bias not only with respect to respondent characteristics and survey design features but also with respect to interviewer characteristics. We use an innovative study design drawing on a unique dataset; the British Household Panel Study (BHPS) combined with a rich dataset from a survey of the BHPS interviewers. This linked data has not been exploited before but promises to offer plenty of new insights into the processes that lead to differential survey outcomes.

[1] Major social surveys have linked their data with a wide range of administrative data including benefit receipt, adolescent's school performance and health and morbidity (e.g. the US Current Population Survey, the Longitudinal Survey of Young People in England, the UK Millennium Cohort Study).

[2] The most comprehensive research based on the general population sample has been conducted by Jenkins and colleagues and investigates consent to link benefit and employer records of responders to the discontinued 'low-income' European Community Household Panel sub-sample of the BHPS.

Factors that affect consent to data linkage

Patterns of consent have been studied predominantly by health researchers and relate to patients participating in small-scale surveys run by medical researchers at local General Practitioners' practices, where patients are making an informed decision on whether or not to allow the researchers to also review their medical records (Dunn et al. 2004; Kho et al. 2009). However, there are also a small number studies that have investigated consent bias on large-scale surveys and focussing on consent to linking economic records (e.g., Jenkins et al. 2006; Olson 1999). While a plethora of research in the survey methodology research has documented that outcomes such as response to a survey request are affected by respondent characteristics, features of the interview process and characteristics of the interviewers, the research investigating patterns of consent to link administrative data has as yet to explore whether interviewer characteristics or survey design features matter.

Respondent characteristics

The research that explores respondent characteristics and consent to data linkage has produced two major findings. First, consent is associated with respondents' socio-demographic characteristics (age, gender, socio-economic status and ethnicity) and their health (Gerber et al. 2007; Olson 1999). However, the nature of the relationship between the different respondent characteristics and the propensity to consent remains unclear as characteristics that are associated with higher consent in one study are negatively associated with consent in another. For example, Kho et al. (2009)'s review of 17 medical research papers reports that seven papers find significant differences across age strata, and seven do not. Four studies find females less likely to consent, while six find no gender differences.

Second, consent is related to respondent's perception of risk, altruism and community-mindedness. Consent is lower among people who refuse to provide information on income or wealth (Jenkins et al. 2006; Olson 1999; Woolf et al. 2000), who believe that the data may be used for fraud detection (Gray et al. 2008) and higher among those who perceive that the wider society can benefit from the data linkage (Dunn et al. 2004; Jenkins et al. 2006).[3] Consent is also lower for respondents who fear that information may not be kept confidential (Armstrong et al. 2008).

[3] This is a conjecture also put forward for survey participation in general, see Singer et al. (2003).

Survey design features

Research on the impact of the survey design and interview process on other types of consent such as survey (non)response shows that interview length and topic play a role in obtaining respondents' consent to co-operate (for a review see Groves and Couper 1998). This is an area of research that is very much under-researched with respect to consent to data linkage. We are aware of only a single study that explores this. Jenkins et al. (2006) found that consent to data linkage is positively associated with the quality of the interviewer-respondent rapport (as measured by interview length and the interviewer evaluation of the "smoothness" of the interview). It is, then, not implausible that other characteristics of the interview process such as number – and order – of interviews in the household also affect consent.

Interviewer characteristics

A further gap in the empirical literature on patterns of consent to data linkage exists with respect to interviewer characteristics. However, this research field is quite promising: qualitative epidemiological studies suggest that patients' propensity to consent varies on the status of the medical staff who is asking for consent (i.e., consent rates are higher when GPs, rather than receptionists ask for it, see, e.g., Armstrong et al. 2008; Baker et al. 2000) whilst the survey methodology research has consistently documented the occurrence of interviewer effects in a wide range of survey outcomes including (non)response and data quality (amongst others, Fuchs 2009; O'Muircheartaigh and Campanelli 1998; Pickery and Loosveldt 2000).

Unfortunately, the former research strand has not identified what particular characteristics of the medical staff (e.g. age, gender etc.) are likely to be associated to patients' consent and the latter has failed to clearly identify the interviewer characteristics that are driving these interviewer effects. Interviewer socio-demographic characteristics such as gender, age, education and social class are often found to be independent of different indicators of survey response or data quality (Esbensen and Menard 1991; Link 2006; Lipps 2007; O'Muircheartaigh and Campanelli 1999; Pickery and Loosveldt 2000; Pickery and Loosveldt 2001; Pickery and Loosveldt 2004). Exceptions are, for example, Haunberger (2009) who finds that highly educated interviewers obtain better co-operation than others and Hansen (2006) who finds that male and older interviewers achieve higher response rates. The role of interviewer experience remains particularly unclear. For example, Hansen (2006), Pickery and Loosveldt (2000) and Jäckle et al. (2010) find that more experienced interviewers

achieve higher response rates while Kennickell (1999) and, more recently, Durrant et al. (2010) show that long-term interviewers can perform less well than those with less experience. Some empirical evidence shows that the interviewer personality and attitudes, rather than their socio-demographic characteristics, are associated with a number of survey outcomes. For example, interviewer confidence and attitudes towards persuasion have a positive impact on survey response (Kennickell 1999; Lehtonen 1996). Preliminary work by Jäckle at al. (2010) finds that interviewer personality, measured using the so-called 'Big Five' instrument (John and Srivastava 1999), is associated with co-operation. That study also finds, however, that interviewer attitudes to persuading respondents are *not* associated with respondents' co-operation.

Against the backlight of this literature, this paper explores the role of a broad range of respondent characteristics, survey design features and interviewer characteristics on respondents' propensity to consent to administrative data linkage. Given the mixed and sometimes inconsistent findings yielded by previous empirical research (in particular regarding the association between consent and standard socio-demographic characteristics) we have to be open as to what statistical associations with consent we may find. However, we may expect to find that respondent's propensity to consent is associated with indicators of risk aversion, community-mindedness and interviewer-respondent rapport, as well as with interviewers' personality and their attitudes to persuading respondents. Moreover, we speculate that survey design features such as household-interview specific characteristics (the number – and order – of interviews in the household) affect consent.

DATA

We use a unique dataset; the British Household Panel Study (BHPS) combined with information gathered in a survey of the interviewers who collected the data.

The British Household Panel Study (BHPS)

The BHPS is one of the most important research resources in the UK and is one of the longest running household panel studies in the world. It started in 1991 with a sample of nationally-representative stratified, clustered sample of 5,500 households and roughly 10,000 individuals interviewed face-to-face, with interviewers calling on respondents in their homes. In 1999, booster samples of around 1,500 households each were added in Scotland and

Wales, and in 2001, a sample of 2,000 households was added in Northern Ireland. In Wave 18, the most recent of the survey, 12,971 full interviews were completed. Annual waves of data collection provide a wide range of information including household composition and conditions, education and training, health and use of health services, labour market behaviour, socio-economic values and different income sources.

Data linkage

Asking for consent to data linkage to health, social security benefits and educational administrative records was implemented at Wave 18 of the BHPS. The data linkage module was administered at the end of the individual questionnaire (see Appendix 1 for the question wording and order of consent question). In the UK informed consent must be obtained from respondents in order to link administrative data at the individual level to survey data. If the respondent verbally agreed to give consent, the interviewer then handed them a form that the respondent was asked to read and sign. Different kind of consents were asked to different respondents, depending on their age and whether they were the 'responsible adult' for a child (aged under 16) in the household (see Table 1). All adults were asked for their consent to link to their own health and benefit records. Adults who were responsible for a child in the household were asked to give consent to link to the child's health records. Adults who were aged between 16 and 24 were asked for the consent to link to their own education records. Responsible adults of children aged 4-15 were asked for their consent to link to the child's education records. None of the consents were conditional on other consents being given, so if someone refused to give consent to one data linkage they were still asked about the next data linkage. In this paper we focus on adult consent to link to health and benefit records data only.

Table 1 Consent rates by consent type

Consent type	Written consent asked from	Consent rate
Health	All adults (age 16+)	41%
	All "responsible adults" of children (age 0-15)	38%
Education	All young adults aged 16-24	39%
	All "responsible adults" of school-age children (age 4-15)	39%
Economic records	All adults (age 16+)	32%

Source: BHPS Wave 18.

Table 1 gives the consent rate for each of the five consent questions. Forty-one percent of adult respondents gave consent to health data linkage whereas 32 percent consented to the linkage to economic records. Interestingly, then, while ethics regulations in the UK suggest that health data is more sensitive, consent to health data linkage is higher than to economic record linkage. Note that, compared to consent rates obtained in similar surveys, the BHPS consent rates are much lower. For example, a similar request for consent to data linkage to benefit records using the British sample of the former European Community Household Panel resulted in a consent rate of 77 percent. We believe two main causes are driving the low BHPS consent rates: the well-publicised security breaches leading to loss of government data at around the time that the BHPS was going into the field and the announced change in the fieldwork agency[4] which may have discouraged interviewers from learning a new survey task and motivating respondents to give consent.

The BHPS interviewer survey

The interviewer survey consisted of a self-completion questionnaire administered during Wave 18 BHPS briefings. Researchers from the Institute for Social and Economic Research (the institute that runs the survey) attended a number of these in-person briefings and administered a questionnaire to all 180 interviewers present at those briefings (68 percent of all interviewers at Wave 18)[5]. At briefings at which a researcher was present, all interviewers (100 percent) completed the questionnaire and returned it to the researcher in a sealed envelope.

The interviewer questionnaire collects four types of information (see Appendix II for the interviewer survey questionnaire): basic information on socio-demographic characteristics (age, sex, educational qualification, presence of children at home, household composition), interviewer experience (type and duration), interviewers' views on different aspects of their job, interviewers' personality traits measured by the "Big Five" taxonomy (John and Srivastava 1999), their attitudes to persuading and contacting respondents measured by five items from the Lehtonen scale (Lehtonen 1996) and three items used by Blohm et al. (2007).

[4] After Wave 18, the BHPS was incorporated into *Understanding Society*: The UK Household Longitudinal Study. The fieldwork contract for this latter study was won by a different fieldwork agency to that used on the BHPS.

[5] We used a dataset of interviewers provided by the survey agency to check for bias between interviewers who completed the interviewer survey and those who did not. Although we did not find any evidence for bias with respect to interviewer age, we did find that men were more likely to be overrepresented in the interviewer survey.

The level of item non-response was very low and varied from 1 percent or less for the questions on interviewer experience to about 2 percent on questions on personality traits and attitudes to persuading respondents. The highest level of item non response was on for education (5 percent).

A STATISTICAL MODEL OF CONSENT

Model specification

There are two consent outcomes which are available for all adult respondents to the BHPS, namely the consent to link administrative health records and the consent to link administrative benefit records. The data space allows us to focus on three different probabilities, i.e., the probability to consent to health data linkage only, the probability to consent to the benefit data linkage only, and the probability to consent to both data linkage requests. From the point of view of substantive analysis all three probabilities are interesting because they tell us how much bias we may expect if we use BHPS linked with health records only, with benefit records only or with records from both domains (assuming there exists a record for each consenter and it can be linked successfully), respectively.

In our empirical analysis we will focus on the bivariate probability to consent, mainly because we are concerned not only with consent bias but also with modelling consent in the most comprehensive way possible, and achieving a very high degree of generalisability. We will estimate respondent's propensity to consent on the basis of both outcomes using multivariate bivariate probit models, which can be written as:

$$\begin{cases} y_{1i}^{*} = \beta_1^{'} x_{1i} + u_{1i} \\ y_{2i}^{*} = \delta_2^{'} z_{2i} + u_{2i} \end{cases}, i = 1, 2, \dots, n$$

where y_{1i}^{*} and y_{2i}^{*} are latent variables so that the observed dichotomous outcomes y_{1i}, i.e., the health record linkage request, and y_{2i}, i.e., the benefit record linkage request, are given by:

$$\begin{cases} y_{ki} = 1 \; if \; y_{ki}^{*} > 0 \\ y_{ki} = 0 \; if \; y_{ki}^{*} \leq 0 \end{cases}, k = 1,2$$

In the model, x_{1i} and z_{2i} are vectors of observed exogenous variables that have been suggested to affect consenting, and β_1 and δ_2 are the respective parameter vectors. The error terms in this model are distributed as standard bivariate normal variables with correlation coefficient ρ. More detailed information on this standard model can be found, for instance, in

Greene (2003). To our end it is important to note that *Rho* may be interpreted as the respondents' unobserved propensity to consent (see Jenkins et al. 2006). If the parameter is statistically significant, modelling the consent outcomes jointly is more efficient than using univariate probit models (or indeed univariate linear probability models), which otherwise yield the same substantive results. Estimation of the model is straightforward using Stata's *biprobit* command (StataCorp 2009). Note that since many of our respondents live in the same household and members of the household are interviewed by the same interviewer, we adjust standard errors for clustering on interviewers.

Choice of predictor variables

The BHPS offers plenty of information on respondents, their households, and the interview situation, both for the present and the past. Our choice of variables is guided by the literatures on consent bias, survey co-operation and interviewer effects on data quality. We organise the variables in three blocks, i.e., respondent characteristics, survey design features and interviewer characteristics.

Respondent characteristics

Like most other research on consent bias, our models consider respondent demographic characteristics (age, gender, and ethnicity) and socio-economic characteristics (education, household income, and household context). In addition, from the UK-based literature on survey co-operation we know that people in the South East are more likely to participate in surveys (see, e.g., Lynn 2003). Following Jenkins et al. (2006) we merged Londoners (who tend to have lower survey response rates, largely due to higher non-contact rates) with this group.[6] Moreover, we include a number of characteristics which we believe tap into the respondent's perceptions of the risk of data linkage. In particular, we include information on refusing to provide information on income from investments to proxy for the respondent's general attitudes to sharing information.[7] To allow for the differential levels of saliency the respective data linkage request may be for the respondent, we include indicators for whether or not respondents have been to hospital in the previous 12 months, had any of 15 types of

[6] There is no statistically significant association between living in London and consenting to data linkage.
[7] Consent to data linkage is positively associated with response on income and wealth indicators, see, e.g., Woolf et al. (2000) and Olsen (1999).

health problems,[8] whether they currently receive income support payments from the government, and how many means-tested benefits they receive.[9] Respondents may, however, have a more 'community-minded' attitude which drives them to engage in the research. We will throw some light on this by adding dummy variables for (i) whether or not the respondent supports a left-wing/liberal party[10], (ii) whether or not they do voluntary work without receiving pay, and (iii) whether or not they generally trust others.

Survey design features

With respect to survey design features potentially affecting consent, we include a number of proxy measures for rapport (the number of years the respondent has been participating in the BHPS, and whether or not the interviewer in the current wave interviewed the respondent in the previous year[11]). To capture potential influences of others we include a dummy for whether or not others were present at any time during the interview (i.e., not specifically when the consent was asked). Moreover, we consider how many interviews had already taken place in the household for the present BHPS wave, and the number of consents that had already been given by other household members at the time the respondent is asked. This exploits information about the time of the interviews with other members of the household and the respective consent outcomes. We believe the measures will pick up what we might refer to as 'household contagion', i.e., the influence of the respondent's and the interviewer's knowledge of how easy/difficult it has been to get consents from the people already interviewed in this household.

Interviewer characteristics

With respect to interviewer characteristics, we use standard interviewer socio-demographic characteristics (sex, age, education). In addition, we include three different measures of 'experience'; (i) job experience, i.e., the number of years that the person has been an interviewer, (ii) survey experience, i.e., the number of interviews on this survey the interviewer has carried out this wave and (iii) task experience, i.e., the information about what has already occurred when asking for consent within interviews this wave. The idea

[8] That is, health problems relating to arms, legs, hands, etc.; sight; hearing; skin conditions; chest and breathing; heart and blood pressure; stomach and digestion; diabetes; anxiety and depression etc.; alcohol or drug etc.; epilepsy; migraine; cancer; stroke; and other.
[9] The latter two also indirectly measure respondent's economic prosperity.
[10] Labour Party, Liberal-Democrats, Plaid Cymru (Wales), Scottish National Party, or Green Party.
[11] We also tested whether it mattered for consent how many minutes the interviewhas taken until the consent question was asked. Since this did not show any statistically significant association with consent, in contrast to the findings of Jenkins et al. (2006), we dropped the indicator.

here is that interviewers may accumulate not only knowledge about how easy or difficult it is to obtain consents within a given household, but also across households. This could pan out either positively, namely, if interviewers learn from their past task-specific experience and manage to adjust the way in which they ask consents, or negatively, if they do not.

We include a series of measures aimed at picking up any influence of interviewer behaviour on consent. First, to capture the interviewer's likely behaviour while asking for consent, we include five proxy measures for how much effort the interviewer exerts in trying to persuade respondents to consent. Second, to see whether interviewer attitudes and behaviour affect consenting in a more comprehensive way (i.e., not related necessarily to the way interviewers perceive their role as interviewers), we include the interviewers' "Big Five" personality traits. All these variables have been collected in the BHPS interviewer survey (described above).

Summary statistics and variable descriptions for all variables used in the analysis are presented in Appendix 3.

RESULTS

Table 2 reports the results of joint estimation of consent to health and benefit data linkage[12], first, including only respondent characteristics (Model 1), followed by a model that controls for respondent characteristics and survey design features (Model 2) and a model that controls for respondent characteristics, survey design features and interviewer characteristics (Model 3). To ease comparison of the results, Models 1 and 2 are nested in Model 3. First, note that the cross-equation correlation *Rho* is highly statistically significant in all three models suggesting that there is an unobserved factor that affects both decisions. We interpret this as the respondents' unobserved propensity to consent. It cannot be ruled out, however, that despite the large number of regressors included in our analysis, there are other unobserved characteristics that influence both consents.

[12] i.e., the probability that consent to both data linkages are given. Results from the univariate probit models, which are not discussed here, are presented in Appendix 4.

Table 2. Propensity to consent, by BHPS respondent, interview and interviewer characteristics (bivariate probit regressions).

	Coefficients					
	Model I		Model 2		Model 3	
	Health	Benefit	Health	Benefit	Health	Benefit
Respondent characteristics						
Male	0.03	0.06*	0.04	0.09*	0.04	0.09**
Ethnicity (British/Irish White)						
Other White	-0.27*	-0.22*	-0.26*	-0.22*	-0.29**	-0.25*
Mixed	-0.05	0.02	-0.31	-0.29	-0.29	-0.24
British Asian/Black	-0.55***	-0.53**	-0.38*	-0.38*	-0.38**	-0.37*
Other ethnicity	-0.12	-0.07	-0.10	-0.06	-0.13	-0.09
Age group (16-24 years old)						
25-39 years old	-0.24**	-0.34***	-0.21*	-0.29**	-0.21*	-0.29**
40-49 years old	-0.23**	-0.38***	-0.14	-0.30**	-0.17	-0.33***
50-59 years old	-0.24**	-0.36***	-0.12	-0.25*	-0.12	-0.25*
60+ years old	-0.18	-0.41***	-0.03	-0.28*	-0.07	-0.30**
Education degree or beyond	0.11*	0.16**	0.10	0.16**	0.09	0.15**
Household type (Single)						
Couple, no children	0.00	-0.02	-0.02	-0.06	0.01	-0.04
Couple with children	-0.09	-0.13	-0.03	-0.12	0.00	-0.10
Lone parent	-0.01	-0.04	0.00	-0.04	-0.01	-0.06
Other household type	0.12	0.05	0.13	0.03	0.16	0.03
Household size	0.00	-0.01	-0.01	-0.02	-0.01	-0.02
England	0.13	0.13	0.07	0.06	0.03	0.03
London/Southeast	0.22*	0.18	0.20*	0.18*	0.16*	0.15*
Household income (log)	0.00	0.02	0.01	0.04	0.02	0.04
Refused question: Income from investment	-0.61***	-0.73***	-0.58***	-0.70***	-0.62***	-0.75***
Generally trusts others	0.21***	0.17***	0.22***	0.19***	0.22***	0.19***
Supports leftwing/liberal party	0.19***	0.22***	0.19***	0.21***	0.17***	0.19***
Does unpaid voluntary work	0.13**	0.11*	0.14**	0.12*	0.12*	0.11*
Has health problems	0.05	0.06	0.06	0.08*	0.06	0.08
Has been to hospital	0.16*	0.05	0.15*	0.02	0.14*	0.01
Receives any state benefits	0.02	0.05	-0.02	0.01	-0.01	0.02
Number of means-tested benefits received	0.06*	0.07**	0.05*	0.07***	0.04	0.07**

(continues)

***Table 3**. (continued)*

	Coefficients					
	Model I		Model 2		Model 3	
	Health	Benefit	Health	Benefit	Health	Benefit
Survey design features						
Interview sequence within household			-0.83***	-0.79***	-0.75***	-0.71***
Number of previous health consents in household			1.25***	0.69***	1.14***	0.60***
Number of previous benefit consents in household			0.44***	0.99***	0.40***	0.93***
Others present during interview			0.06	0.11*	0.03	0.08*
Number of years in the BHPS			-0.02**	-0.02**	-0.02**	-0.02***
Same interviewer as previous wave			0.21*	0.24*	0.09	0.15
Interviewer characteristics						
Male interviewer					-0.03	-0.11
Interviewer age group (40-49 years old)						
50-59 years old					0.05	0.05
60-69 years old					0.12	0.09
70+ years old					-0.04	-0.13
Interviewer has degree or above					0.07	0.03
Interviewing experience in years					0.00	-0.02
Number previous interviews by interviewer					-0.02***	-0.02***
Number of health consents already obtained					0.03**	0.01
Number of benefit consents already obtained					0.01	0.03**
Attitudes to persuading						
All can be persuaded					-0.05	-0.09
Should persuade					-0.01	-0.01
Should respect privacy					-0.07	-0.11
Should accept refusal					0.06	0.07
Emphasise voluntary nature					-0.06	-0.04
Personality traits						
Agreeableness					0.03	0.03
Conscientiousness					0.02	0.00
Openness					-0.02	0.00
Extraversion					-0.05	-0.05
Neuroticism					-0.03	-0.02
Cross-equation correlation	0.96		0.95		0.94	
Log(pseudo)Likelihood	-5408.1		-4734.1		-4541.3	
N	5825		5825		5825	

Significant at *** .001, ** .01, * .05.
[#] Standard errors adjusted for 148 clusters of interviewers.
Source: BHPS Wave 18 matched with interviewer data.

Respondent characteristics

When looking at respondents' characteristics that influence consent, a clear pattern emerges: respondents' propensity to consent to data linkage does not seem to be strongly associated with their demographic or socio-economic characteristics but it appears to be related to their attitudes to privacy, community-mindedness and data linkage salience.

Table 4 shows that, on the whole, demographic and socio-economic respondent characteristics are only mildly associated with consent. Members of UK minority ethnic groups[13] as well as older respondents are less likely to consent (note, in this case, the change in the level of statistical significance for the health equation in Model 2). These findings are consistent with previous studies (see, e.g., Hockley et al. 2007; Tate et al. 2006; Woolf et al. 2000). Characteristics describing the respondent's household context (including household income) do not appear to be associated with consent.

On the other hand, indicators of respondents' attitudes to privacy[14] and community-mindedness show a marked association with consent. In particular, refusing to answer the question on income from investment is a strong predictor of not giving consent, while generally trusting others positively affects consent. Indicators of saliency of the data linkage such as being in the hospital in the last 12 months or receiving a larger number of means-tested benefits are positively associated with consent. However, for the BHPS respondents salience does not appear to influence consent as much as has been suggested in previous research on health data linkage requests (Dunn et al. 2004).

The pattern of consent that we have just described is true for both types of consent (though some of the effects seem to have more influence on consent to benefit record linkage) and, with only a few exceptions, is robust to the inclusion of interview and interviewer characteristics (compare upper panels of Models 1-3, Table 2). For example, for the salience indicators, the statistical associations with consent vary for the health and benefit consent outcomes and across specifications. Whilst the health measure is only significant for the

[13] However, we do have to add the caveat that the ethcic minority sample size in the BHPS is quite small, even when groups are combined.

[14] It was suggested that this indicator may tell us more about how wealthy the respondent is than about his or her reluctance to share sensitive information. To assess this we also included, from the 15th wave of the BHPS, information on whether or not the respondent has savings worth £10k or more, debt amounting to £5k or more, and investments worth £5k or more. The inclusion of these variables did not affect the propensity to consent, nor did it change the effect of refusing the income from interest question on consent. Thus, we are more confident that this is an effect of privacy, rather than wealth.

health linkage, the benefit measure is significant on both (but more significant on the benefit linkage).

Survey design features

As Model 2 clearly shows, respondents' propensity to consent to data linkage also is strongly associated with some survey design features; the sequence of interviews within the household, the effects of previous consent requests ('household contagion') and survey "fidelity". The direction of some of these relationships, however, is not always as expected. Interviewer-respondent rapport does not seem to play a role in the consent process.

There is a negative association between household interview sequence and consent. Later interviewees are less likely to consent to data linkage compared to household members who are interviewed earlier. We interpret this variable as an indicator of survey resistance, However, it could also be interpreted as an indicator of interviewer burden (e.g., due to time pressure interviewers rush through the later interviews[15]). Paradata on timing and contact attempts are needed to explore this issue further.

Interestingly, respondents' probability to consent is positively associated with the number of household members who have already consented to data linkage. This is evidence for a 'household contagion' effect (i.e., household members consult each other and take joint decisions). Note that mere presence of others during the interview does not appear to be associated with consent.

Contrary to our expectations, respondents' consent to data linkage is negatively associated with the survey "fidelity" indicator: respondents' propensity to consent decreases with the number of years they have been in the panel. We speculate that BHPS respondents who have been part of the panel for longer may feel they have provided so much information already over the past (up to 18 years) that they do not see why access to administrative data may be needed. Another reason may be that they are suspicious of a survey innovation which comes about after so many years.

[15] In the BHPS all members of the household are eligible for interview and these interviews often take place sequentially within a single visit by the interviewer (although interviewers are briefed to make multiple trips to the household to interview everyone).

We also do not find strong evidence to support the claim that interviewer-respondent rapport has an impact on consent. Respondents' consent is only very weakly associated with having been interviewed by the same interviewer in the previous year.[16]

As with respondent characteristics, the pattern of consent that we have just described is true for both types of consent and, on the whole, is robust to the inclusion of interviewer characteristics.

Interviewer characteristics

Model 3 shows that interviewer socio-demographic characteristics are not associated with respondents' propensity to consent. Furthermore, somewhat unexpected, our proxies for interviewer behaviour (i.e., interviewer personality traits and attitudes to persuading respondents) show no statistically significant association with consent.[17] The only interviewer characteristic that appears to matter is experience. Albeit, it is not the length of time they have worked as interviewers that matters, but rather their more specific survey experience in the current wave and their task-specific experience. The more BHPS interviews an interviewer has already carried out during the wave, the less likely she/he is to obtain respondents' consent. However, the more successful she/he has been in obtaining respondents' consent in one particular domain, the more likely she/he is to gain respondents' consent. We theorise that an interviewer who has already asked the consent questions, and the consent has been withheld, is likely to feel less optimistic about asking an additional person in the same household. This may affect the effort made by the interviewer to explain the purpose of the data linkage, and may even affect whether or not the interviewer actually asks the question – rather than just assuming a refusal. A respondent who is aware that others in the household have already withheld their consent may find it easier to withhold their own consent – and thus save themselves a couple of minutes in the interview – and may even feel some 'peer pressure' to refuse to maintain a consistent household response (and not appear inconsistent).

[16] In contrast to what was found by Jenkins et al. Jenkins, Stephen P., Lorenzo Cappellari, Peter Lynn, Annett Jäckle, and Emanuela Sala. 2006. "Patterns of consent: evidence from a general household survey." *Journal of the Royal Statististical Society, Series A* 169: 701-722., we also did not find an association with the interview length (results are, therefore, not reported here).

[17] We also investigated whether it may in fact be congruence of interviewers and respondents on key characteristics such as age, gender and education as well as personality that matter for the probability to get consent – are respondents more likely to agree to data linkage when they are asked by someone who is "like them"? Results from regression analysis did not show any statistically significant relationships (available from the authors on request).

Magnitude of the influences on consent

To assess the relative strength of the effects, we estimated the relative marginal effects (MEs) for the most comprehensive model, i.e., Model 3. These are shown in Table 3. MEs are easier to interpret as they express by how many percentage points the average probability to consent would change if the explanatory characteristic changes by a unit, holding all else constant. For categorical variables, MEs express how much the probability would change if we were to observe a discrete change away from the base category. The baseline predicted probability of consent, calculated at the mean of the explanatory variables, and the means provide a reference point for whether or not the MEs are small or large.

Table 3 shows that the average predicted probability to consent to both data linkages is 25 percent. It also allows identifying the factors that are most positively/negatively associated with consent. Indicators of household contagion are the strongest positive predictors of respondents consenting to data linkage (ME: 0.26/0.23). Other things being equal, an additional consenting respondent to health (benefit) data linkage in the household increases to 48 (51) percentage points the propensity for another household member to consent. Other important positive predictors of consent are respondents' community-mindedness (ME: 0.06) and trust (ME: 0.06). At the other end of the spectrum, the indicators of survey resistance and privacy concerns are the strongest negative predictors of consent (ME: -0.23 and -0.1, respectively).

Table 3. Estimated Marginal Effects. Consent to health and benefit record linkage (bivariate probit regression).[#]

	Margin		S.E.	Mean
Estimated probability (at means)				0.25
Respondent characteristics				
Male	0.03	*	0.01	0.47
Ethnicity (British/Irish White)				
Other White	-0.08	**	0.03	0.03
Mixed	-0.07		0.06	0.01
British Asian/Black	-0.10	**	0.04	0.02
Other ethnicity	-0.03		0.03	0.05
Age group (16-24 years old)				
25-39 years old	-0.09	**	0.03	0.22
40-49 years old	-0.10	**	0.03	0.19
50-59 years old	-0.07		0.04	0.16
60+ years old	-0.08	*	0.04	0.30
Household type (Single)				
Couple, no children	-0.01		0.02	0.32
Couple with children	-0.03		0.03	0.42
Lone parent	-0.01		0.03	0.09
Other household type	0.02		0.04	0.03
Household size	-0.01		0.01	2.78
England	0.01		0.02	0.54
London/Southeast	0.05	*	0.02	0.22
Education degree or beyond	0.04	*	0.02	0.16
Household income (log)	0.01		0.01	7.87
Refused question: Income from investment	-0.18	***	0.02	0.04
Generally trusts others	0.06	***	0.01	0.33
Supports leftwing/liberal party	0.06	***	0.01	0.50
Does unpaid voluntary work	0.04	*	0.02	0.24
Has health problems	0.02		0.01	0.61
Has been to hospital	0.01		0.02	0.09
Receives any state benefits	0.01		0.02	0.57
Number of means-tested benefits received	0.02	**	0.01	0.47
Survey design features				
Number of previous interviews in household	-0.23	***	0.02	1.55
Number of previous health consents in household	0.23	***	0.02	0.22
Number of previous benefit consents in household	0.26	***	0.02	0.17
Others present during the interview	0.02		0.01	0.32
Number of years in the BHPS	-0.01	***	0.00	15.42
Same interviewer as previous wave	0.04		0.02	0.88

(continues)

Table 3*. (continued)*

	Margin		S.E.	Margin
Interviewer characteristics				
Male interviewer	-0.03		0.03	0.17
Interviewer age group (40-49 years old)				
50-59 years old	0.02		0.04	0.26
60-69 years old	0.03		0.04	0.59
70+ years old	-0.03		0.04	0.09
Interviewer has degree or above	0.01		0.02	0.31
Interviewing experience in years	0.00		0.01	2.57
Number previous interviews by interviewer	-0.01	***	0.00	33.27
Number of health consents interviewer already obtained	0.00		0.00	12.91
Number of benefit consents interviewer already obtained	0.01	*	0.00	9.90
Attitudes to persuading				
All can be persuaded	-0.03		0.02	2.69
Should persuade	0.00		0.01	2.64
Should respect privacy	-0.03		0.02	1.28
Should accept refusal	0.02		0.01	2.26
Emphasise voluntary nature	-0.01		0.02	1.56
Personality traits				
Agreeableness	0.01		0.01	5.81
Conscientiousness	0.00		0.01	5.82
Openness	0.00		0.01	5.11
Extraversion	-0.02		0.01	5.14
Neuroticism	-0.01		0.01	3.16
Cross-equation correlation	-0.03		0.03	0.17

Significant at *** .001, ** .01, * .05.
[#] Standard errors adjusted for 148 clusters of interviewers.
Source: BHPS Wave 18 matched with interviewer data.

DISCUSSION

Despite the increasing number of British and international surveys that ask respondents for permission to data linkage, there is currently very little knowledge on what drives consent. Research on consent to data linkage is mainly constituted by descriptive analyses of the variation in respondents' propensity to consent and the potential resulting respondent bias. The assumption underlying these studies is that the mechanisms that govern consent are located mainly in survey respondents (or, in case of medical studies, patients) and, in particular, in their socio-demographic characteristics. However, factors relating to the survey process may be equally important. Interviewers and their characteristics, the survey topic as

well as the point in time during the life of a panel survey when the respondents are asked for their consent to link administrative data and the survey mechanics within the household could play a major role in the respondents' decision whether or not to consent.

This paper advances the knowledge about consent and consent bias in important ways. It is the first empirical analysis to explore the role of respondents, survey design features and the interviewer in obtaining respondents' consent to perform data linkage. Consent bias is examined for different types of administrative data, comparing consent to link economic records and health records. The research looks at the issue of consent to data linkage using a general population sample, rather than a specialised medical-based sample. Moreover, it draws on a large pool of interviewer-level characteristics, making this the first research investigating whether interviewer attitudes to persuading respondents and interviewer personality affect respondents' consent to data linkage. It is also the first time that the mechanics of interviewing within a household context are explored.

We find that respondent socio-demographic characteristics are mildly associated with consent, but there is a much greater effect of attitudes towards privacy and community-mindedness. Contrary to other research, we do not find that survey design features, such as the length of the interview was significantly associated with consent (Jenkins et al. 2006). We also find that the length of time in the panel was significant, but not positively, as we expected; those who had been in the panel for longer are *less* likely to give consent, other things being equal. We find that interviewer characteristics, including interviewer attitudes and personality traits, which have been suggested as a source for variation in survey response in the survey methodology literature, are not associated with consent to data linkage. We do, however, find a significant effect of the intra-household dynamics on consent, suggesting that the decision by an individual is located within the interaction between the individual, the interviewer and the wider household-context.

So what are the implications of this and how might future research shed further light on these issues? We would like to know more about what happens during the interaction between the interviewer and respondent when the consent question is asked. One possibility is that this section of the interview could be recorded and behaviour-coded. Such a resource could be used to analyse actual interviewer behaviour (rather than using their attitudes and personality traits as proxies for behaviour). We might find indications of household-level interactions which lead to order-effects, or indications of time pressure on the interviewer which may lead to a tendancy to skip past the consent question, particularly if other members of the

household have already withheld their consent. Future studies on data linkage and errors of non-observation could also aim to provide a better understanding of the reasons *why* respondents consent or do not consent to data linkage, for example by carrying out qualitative studies on groups of consenters and/or non-consenters or adding a follow up question in a quantitative study that collects respondents' reasons for withholding consent. Only when a substantial body of empirical research has been collected, a theoretical model that explains the complex processes that lead respondents to consent can be formulated.

For the time being, what we can take away from this analysis is that adjustments for non-response on analyses which incorporate data from administrative sources may need to vary, depending on what administrative sources are being used. Evidence suggests that the bias increases with the number of data sources used. We have discussed just the consent bias for data drawing on three sources; the survey, the administrative health data and the administrative benefit data (for biases of survey data and health data, and survey data and benefit data, see Appendix 4). Whichever linked data is being used it is important to note that sample sizes for analysis are seriously reduced, and biases exist on characteristics that one may be particularly interested in. For instance, the fact that people who refuse to provide information on their income from investment are less likely to consent to benefit data linkage means that their administrative records cannot be used to substitute for information missing from the survey data. We are confident, nevertheless, that, subject to the appropriate adjustments to account for both response and consent bias, linked survey-administrative data provide a rich resource for future substantive analysis.

REFERENCES

Armstrong, Victoria, Julie Barnett, Helen Cooper, Michelle Monkman, Jo Moran-Ellis, and Richard Shepherd. 2008. "Public Perspectives on the Governance of Biomedical Research: A qualitative study in a deliberative context." Pp. 128, edited by Wellcome Trust: University of Surrey.

Baker, Richard, Christopher Shiels, Keith Stevenson, Robin Fraser, and Margaret Stone. 2000. "What proportion of patients refuse consent to data collection from their records for research purposes?" *British Journal of General Practice* 50: 655-556.

Blohm, M., Joop J. Hox, and A. Koch. 2007. "The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face household surveys." *International Journal of Public Opinion Research* 19: 97-111.

Dunn, Kate M., Kelvin Jordan, Rosie J. Lacey, Mark Shapley, and Clare Jinks. 2004. "Patterns of Consent in Epidemiologic Research: Evidence from Over 25,000 Responders." *American Journal of Epidemiology* 159: 1087-1094.

Durrant, Gabriele B., Robert M. Groves, Laura Staetsky, and Fiona Steele. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys." *Public Opin Q* 74: 1-36.

Esbensen, Finn-Aage, and Scott Menard. 1991. "Interviewer-Related Measurement Error in Attitudinal Research: A Nonexperimental Study." *Quality and Quantity* 25: 151-165.

Fuchs, Marek. 2009. "Gender-of-Interviewer Effects in a Video-Enhanced Web Survey Results from a Randomized Field Experiment." *Social Psychology* 40: 37-42.

Gerber, Y, SJ Jacobsen, JM Killian, SA Weston, and VL Roger. 2007. "Participation bias assessment in a community-based study of myocardial infarction, 2002-2005." *Mayo Clinic Proceedings* 82: 933-938.

Gray, Michelle, Rebecca Constantine, Joanna d'Ardenne, Margaret Blake, and S.C. Noah Uhrig. 2008. "Cognitive testing of Understanding Society. The UK Household Longitudinal Study Questionnaire." *Understanding Society Working Paper* 2008.

Greene, William H. 2003. *Econometric Analysis*. Upper Saddle River, New York: Prentice Hall.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in household interview surveys*. New York: Wiley.

Hansen, Kasper M. 2006. "The Effects of incentives, interview length, and interviewer characteristics on response rates in a CATI study." *International Journal of Public Opinion Research* 19: 112-121.

Haunberger, Sigrid. 2009. "The effects of interviewer, respondent and area characteristics on cooperation in panel surveys: a multilevel approach." *Quality & Quantity*.

Hockley, Christine, Maria Quigley, Jon Johnson, Rachel Rosenberg, Carol Dezateux, and Heather Joshi. 2007. "Millennium Cohort Study: Birth Registration and Hospital Episode Statistics Linkage. A Guide to the Dataset." London: Centre for Longitudinal Studies, Bedford Group for Lifecourse & Statistical Studies, Institute of Education, University of London.

Huang, N., S. F. Shih, H. Y. Chang, and YJ. Chou. 2007. "Record linkage research and informed consent: who consents?" *BMC Health Services Research* 7: 59.

Jäckle, Annette, Peter Lynn, Jennifer Sinibaldi, and Sarah Tipping. 2010. "The Effect of Interviewer Personality, Skills and Attitudes on Respondent Co-operation with Face-to-Face Surveys." *mimeo*.

Jenkins, Stephen P., Lorenzo Cappellari, Peter Lynn, Annett Jäckle, and Emanuela Sala. 2006. "Patterns of consent: evidence from a general household survey." *Journal of the Royal Statististical Society, Series A* 169: 701-722.

John, Oliver P., and Sanjay Srivastava. 1999. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." in *Handbook of Personality: Theory and Research*, edited by O. P. John and L. A. Pervin. New York: Guilford Press.

Kennickell, Arthur B. . 1999. "Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances." *Journal of Official Statistics* 15: 283-303.

Kho, Michelle E., Mark Duffett, Donald J. Willison, Deborah J. Cook, and Melissa C. Brouwers. 2009. "Written informed consent and selection bias in observational studies using medical records: systematic review." *British Medical Journal*: 822.

Lehtonen, Risto. 1996. "Interviewer attitudes and unit nonresponse in two different interviewing schemes." in *Sixth International Workshop on Household Survey Nonresponse*, edited by S. Laaksonen. Helsinki: Statistics Finland.

Link, Michael W. 2006. "Predicting the Persistence and Performance of Newly Recruited Telephone Interviewers." *Field Methods* 18: 305-320.

Lipps, Oliver. 2007. "Interviewer and Respondent Survey Quality Effects in a CATI Panel." *Bulletin de Methodologie Sociologique* 97: 5-25.

Lynn, Peter. 2003. "PEDASKY: Methodology for Collecting Data about Survey Non-respondents." *Quality & Quantity* 37: 239-261.

O'Muircheartaigh, Colm, and Pamela Campanelli. 1998. "The relative impact of interviewer effects and sample design effects on survey precision." *Journal of the Royal Statistical Society, Series A* 161: 63-77.

—. 1999. "A multilevel exploration of the role of interviewers in survey non-response." *Journal of the Royal Statistical Society, Series A* 162: 437-446.

Olson, Janice A. 1999. "Linkages With Data From Social Security Administrative Records in the Health and Retirement Study." *Social Security Bulletin* 62.

Pickery, Jan, and Geert Loosveldt. 2000. "Modeling Interviewer Effects in Panel Surveys: An Application." *Survey Methodology* 26: 189-198.

—. 2001. "An exploration of question characteristics that mediate interviewer effects on item non-response." *Journal of Official Statistics* 17: 337-350.

—. 2004. "A Simultaneous Analysis of Interviewer Effects on Various Data Quality Indicators with Identification of Exceptional Interviewers." *Journal of Official Statistics* 20: 77-89.

Singer, Eleanor, John Van Hoewyk, and Randall J. Neugebauer. 2003. "Attitudes and behaviour. The impact of privacy and confidentiality concerns in the 2000 Census." *Public Opinion Quarterly* 67: 368-384.

StataCorp. 2009. *Statistical Software: Release 11.0*. College Station, TX: Stata Corporation.

Tate, A. Rosemary, Lisa Calderwood, Carol Dezateux, and Heather Joshi. 2006. "Mother's Consent to Linkage of Survey Data with her Child's Birth Records in a Multi-ethnic National Cohort Study." *International Journal of Epidemiology* 35: 294-298.

Woolf, Steven H., Stephen F. Rothemich, Robert E. Johnson, and David W. Marsland. 2000. "Selection Bias From Requiring Patients to Give Consent to Examine Data for Health Services Research." *Archives of Family Medicine* 9: 1111-1118.

APPENDIX

Appendix 1. Consent questions

In this appendix, we set out the actual question wording for the consent module, along with the type of information contained in the administrative records. Q1 and Q2 refer to health linkage; Q1 was asked of all adults, Q2 was asked of the responsible adult of children (aged under 16) in the house. Q3 is for adults aged 16-24 who were asked for consent to link to both benefit and education records, Q4 is for adults aged 25 and above who were just asked about benefit records and Q5 is for responsible adults of 4-15 year old children to ask about education linkage.

The consent questions read as follows:

Q1 Finally, we would like to add some information from administrative health records to the answers you have given. We have sent you an information leaflet which details this and here is a permission form. Please read it, ask me any questions you may have and sign the form if you are happy for us to do this.

Q2 We would also like to add further information on your child's health and use of health services. Could you read through this form and sign it if you wish to give permission.

Q3 We would also like to add some information from educational and economic records to the answers you have given. Here is another information leaflet which details this and here is a permission form. Please read these, ask me any questions you may have and sign the form if you are happy for us to do this.

Q4 We would also like to add some information from economic records to the answers you have given. Here is another information leaflet which details this and here is a permission form. Please read these, ask me any questions you may have and sign the form if you are happy for us to do this.

Q5 We would also like to add further information on your child's education. Could you read through this form and sign it if you wish to give permission.

The information leaflet and the consent form set out what type of information was contained in the administrative records that would be linked. Health linkage covered information about admissions or attendance to hospital (including dates, diagnoses, treatments, surgical procedures and waiting times), records of specific conditions such as cancer or diabetes, prescriptions, health registration information, cause and date of death. The term "economic records" was used to cover National Insurance contributions, state benefits received, employment and earnings information, savings and pensions information and participation in government schemes run by the Department for Work and Pensions. Education records covered national tests and assessment dates and scores, school-level information on pupils and courses taken in further education.

Appendix 2. Interviewer questionnaire

Your answers to this questionnaire are strictly confidential and are for use in statistical analysis only. Your responses will not be passed back to Gfk-NOP.

Please write in your interviewer number:_____

1. Have you ever worked as an interviewer for an organisation other than Gfk-NOP? Please include any work you might be doing now.

   Please indicate your answer by circling the appropriate code.

   Code first that applies.

   Yes, currently..... 1
   Yes, but not now..... 2
   No, never 3

2. How long in total have you worked as an interviewer on social surveys which involve interviewing people in their own homes?

   It is not important whether this is your main employment or just part-time. Please include your current work and work you may have done earlier, including for other organisations.

   Years ☐☐ and months ☐☐

3. Apart from interviewing people in their own homes for a social survey, have you ever done any of the following activities?

   Please indicate your answer by circling the appropriate code for each activity.

		Yes	No
3.1	Other survey interviewing (over the phone or market research)	1	2
3.2	Other non-survey interviewing (for recruitment, benefit office, etc)	1	2
3.3	Activities involving interaction with the general public	1	2
3.4	Activities involving 'cold calling' at people's homes	1	2
3.5	Activities where you needed to persuade people (sales job, fund raising, etc)	1	2

4. Thinking about jobs in general, how important do you personally think each of the following aspects are in a job?
   Please indicate your answer by circling the appropriate code for each aspect.
   How important is . . .

		Very important	Important	Not important	Not important at all
4.1	a job that offers good pay?	1	2	3	4
4.2	a job that is interesting?	1	2	3	4
4.3	a job that allows me to work independently?	1	2	3	4
4.4	a job that involves interaction with people?	1	2	3	4
4.5	a job that allows me to decide my times or days of work?	1	2	3	4

(*continues*)

5. Below follows a series of statements on persuading respondents. To what extent do you agree or disagree with each of these. Interviewers may differ in their opinions about these strategies. There are no right or wrong answers. We are interested in your opinion, based on your experience as an interviewer.

Please indicate your answer by circling the appropriate code for each statement.

		Strongly agree	Agree	Disagree	Strongly disagree
5.1	Reluctant respondents should always be persuaded to participate.	1	2	3	4
5.2	With enough effort, even the most reluctant respondent can be persuaded to participate.	1	2	3	4
5.3	An interviewer should respect the privacy of the respondent.	1	2	3	4
5.4	If a respondent is reluctant, a refusal should be accepted.	1	2	3	4
5.5	One should always emphasise the voluntary nature of participation.	1	2	3	4
5.6	It does not make sense to contact reluctant target persons repeatedly.	1	2	3	4
5.7	If you catch them at the right time, most people will agree to participate	1	2	3	4
5.8	Respondents persuaded after great effort do not provide reliable answers.	1	2	3	4

6. .....The following questions are about how you see yourself as a person. Please circle the number which best describes how you see yourself where 1 means 'does not apply to me at all' and 7 means 'applies to me perfectly'.

I see myself as someone who . . .

		1=Does not apply to me at all						7=Applies perfectly to me
6.1	is sometimes rude to others	1	2	3	4	5	6	7
6.2	does a thorough job	1	2	3	4	5	6	7
6.3	is talkative	1	2	3	4	5	6	7
6.4	worries a lot	1	2	3	4	5	6	7
6.5	is original, comes up with new ideas	1	2	3	4	5	6	7
6.6	has a forgiving nature	1	2	3	4	5	6	7
6.7	tends to be lazy	1	2	3	4	5	6	7
6.8	is outgoing and sociable	1	2	3	4	5	6	7
6.9	gets nervous easily	1	2	3	4	5	6	7
6.10	values artistic, aesthetic experiences	1	2	3	4	5	6	7
6.11	is considerate and kind to almost everyone	1	2	3	4	5	6	7
6.12	does things efficiently	1	2	3	4	5	6	7
6.13	is reserved	1	2	3	4	5	6	7
6.14	is relaxed and handles stress well	1	2	3	4	5	6	7
6.15	has an active imagination	1	2	3	4	5	6	7

(*continues*)

ABOUT YOU

7. Please tick whether you are male or female

Male ☐ Female ☐

8. Write in your Age _____

9. What is your highest educational qualification:

9.1	Degree/Higher degree	1
9.2	Teaching/Nursing/Other higher qualification	2
9.3	A-Levels/Scottish Higher/equivalent	3
9.4	O-Levels/GCSE A-C/equivalent	4
9.5	Clerical/Commercial qualifications	5
9.6	CSE Grade 2-5/O-Level D-E/GCSE D-G/equivalent	6
9.7	Other qualifications	7
9.8	No qualifications	8

10. Do you have any children who are living at home?
CODE ALL THAT APPLY

10.1	At home, under 16	1
10.2	At home, 16 or over	2
10.3	Children left home	3
10.4	Have no children	4

11. Which of the following best describes your current household. If you have some children under 16 and some over 16 at home, please use the code for under 16.
CODE ONE ONLY

11.1	Single Non-Elderly (under 65)	1
11.2	Single Elderly (over 65)	2
11.3	Couple No Children	3
11.4	Couple: children under 16 years	4
11.5	Couple: children 16 plus	5
11.6	Lone parent: children under 16 years	6
11.7	Lone parent: children 16 plus	7
11.8	Other (Write in)	9

_____

12. If you worked in some other job before starting interviewing can you write in the job title and what you did below.

Previous job before started interviewing

_____

_____

No previous job ☐

MANY THANKS FOR COMPLETING THE QUESTIONNAIRE.

PLEASE HAND IT TO THE ESSEX UNIVERSITY REPRESENTATIVE ATTENDING YOUR BRIEFING

Source: BHPS Interviewer survey.

Appendix 3. Summary statistics of variables used in the multivariate analysis of consent (N=8,525). Variable description included in the table notes.

Domain	Content block	Variable label	Mean	S.D.	Min	Max
Respondent characteristics	Socio-economic and socio-demographic characteristics[1]	Male	0.47	0.499	0	1
		Ethnicity (British/Irish White)	0.89	0.312	0	1
		Other White	0.03	0.179	0	1
		Mixed	0.01	0.073	0	1
		British Asian/Black	0.02	0.132	0	1
		Other ethnicity	0.05	0.224	0	1
		Age group (16-24 years old)	0.12	0.328	0	1
		25-39 years old	0.22	0.416	0	1
		40-49 years old	0.19	0.395	0	1
		50-59 years old	0.16	0.366	0	1
		60+ years old	0.30	0.460	0	1
		Household type (Single)	0.14	0.347	0	1
		Couple, no children	0.32	0.467	0	1
		Couple with children	0.42	0.494	0	1
		Lone parent	0.09	0.284	0	1
		Other household type	0.03	0.168	0	1
		Household size	2.78	1.380	1	16
		England	0.54	0.499	0	1
		London/Southeast	0.22	0.414	0	1
		Degree or higher	0.16	0.366	0	1
		Household income (log)[2]	7.87	0.751	-2.3	10.4
	Risk-aversion/Community-mindedness[1]	Refused question: Income from investment	0.04	0.190	0	1
		Generally trusts others	0.33	0.469	0	1
		Supports leftwing/liberal party[3]	0.50	0.500	0	1
		Does unpaid voluntary work	0.24	0.426	0	1
	Saliency of the data linkage[1]	Has health problems[4]	0.61	0.488	0	1
		Has been to hospital	0.09	0.286	0	1
		Receives any state benefits	0.57	0.495	0	1
		Number of means-tested benefits received[5]	0.47	1.027	0	9

(continues)

Notes on respondent-level variables:
1 Unless otherwise stated, these variables are standardly included in the individual respondent's data file (indresp) of the BHPS. We collated some of the categories of the ethnicity, household type, region and education classification.
2 Total household income in the month before the interview. Standardly derived variable included in the household respondent's data file (hindresp) of the BHPS.
3 See Footnote 10.
4 See Footnote 8.
5 Count of how many of the following benefits a respondent receives: disability allowances, incapacity benefit, income support, job seeker's allowance, return to work credit, working tax credit, housing benefit, council tax benefit, any other state benefit. It excludes all pensions, child benefit, child tax credit.

Appendix 3. Continued

Domain	Content block	Variable label	Mean	S.D.	Min	Max
Survey design features	Household contagion	Number of previous interviews in household[6]	1.55	0.752	1	10
		Number of health consents given in household[7]	0.22	0.504	0	5
		Number of benefit consents given in household[7]	0.17	0.456	0	4
		Others present during the interview[8]	0.32	0.466	0	1
	Survey fidelity	Number of years in the BHPS[9]	15.42	4.621	1	18
	Rapport	Same interviewer as previous wave[10]	0.88	0.323	0	1

(continues)

Notes on survey design features variables:
6 Count of the number of interviews previously conducted in the household. Uses information on date and time of all interviews conducted in the household.
7 Count of the number of consents to health (benefit) data linkage given by other household members who have been interviewed before. Uses information on date and time of all interviews within the household.
8 Interviewer's report of whether any other person has been present at any time during the interview.
9 Uses information from the xwaveid.dta file of the BHPS. Number of years since respondent first gave a full interview on the BHPS. This is based on the earliest wave for which the respondent gave a full interview.
10 Uses information from survey participation in Wave 17. The indicator will be zero if the respondent is interviewed for the first time or if a change of interviewers has occurred.

Appendix 3. continued

Domain	Content block	Variable label	Mean	S.D.	Min	Max
Inter-viewer characteristics	Socio-demographics[11]	Male interviewer	0.17	0.374	0	1
		Interviewer age group (40-49 years old)	0.05	0.224	0	1
		50-59 years old	0.26	0.440	0	1
		60-69 years old	0.59	0.491	0	1
		70+ years old	0.09	0.288	0	1
	Interviewer experience	Interviewer has degree or above	0.31	0.463	0	1
		Interviewing experience in years[12]	2.57	0.733	0.3	3.6
		Number previous interviews[13]	33.3	27.4	1	153
		Number of health consents already obtained[14]	12.9	12.1	0	67
		Number of benefit consents already obtained[14]	9.9	9.6	0	62
	Attitudes to persuading respondents[15]	All can be persuaded	2.69	0.624	1	4
		Should persuade	2.64	0.678	1	4
		Should respect privacy	1.28	0.480	1	4
		Should accept refusal	2.26	0.619	1	4
		Emphasise voluntary nature	1.56	0.619	1	3
	Personality traits[16]	Agreeableness	5.81	0.732	3	7
		Conscientiousness	5.82	0.903	1	7
		Openness	5.11	0.921	1	7
		Extraversion	5.14	1.067	2	7
		Neuroticism	3.16	1.110	1	6.3
N			5825			

Notes on interviewer-level variables:
11 Standard information collected in the interviewer survey.
12 Indicator of job experience.
13 Indicator of survey experience. Number of interviews the interviewer has already conducted in the current wave of the BHPS. Uses information on date and time of all interviews conducted by an interviewer.
14 Indicators of task experience. Count of the number of consents given by respondents who have been interviewed earlier. Uses information on date and time of all interviews conducted by an interviewer.
15 Responses to Question 5.1-5.5 of the interviewer survey (see Appendix 2).
16 Derived from responses to Question 6 of the interviewer survey (see Appendix 2). For each trait, three item responses are added up and divided over three. Agreeableness uses 6.1 (reverse-scored), 6.6, and 6.11; Conscientiousness uses 6.2, 6.7 (reverse-scored) and 6.12; Extraversion uses 6.3, 6.8 and 6.13 (reverse-scored); Neuroticism uses 6.4, 6.9 and 6.14 (reverse-scored); Openness uses 6.5, 6.10 and 6.15.
Source: BHPS Wave 18 matched with interviewer survey.

Appendix 4. Estimated coefficients and marginal effects. Consent to health and benefit record linkage (Univariate probit regression; N=8,525).[#]

	Pr(Health consent)		Pr(Benefit consent)		Pr(Health consent)		Pr(Benefit consent)			
	Coefficient	***S.E.***	***Coefficient***	***S.E.***	***Margin***		***S.E.***	***Margin***		***S.E.***
Male	0.05	0.03	0.10**	0.04	0.02		0.01	0.03	*	0.01
Ethnicity (British/Irish White)										
Other White	-0.29**	0.11	-0.26*	0.1	-0.10	**	0.04	-0.07	**	0.03
Mixed	-0.25	0.27	-0.27	0.25	-0.10		0.08	-0.07		0.06
British Asian/Black	-0.39**	0.14	-0.37*	0.18	-0.13	**	0.04	-0.11	*	0.04
Other ethnicity	-0.13	0.09	-0.11	0.1	-0.05		0.03	-0.03		0.03
Age group (16-24 years old)										
25-39 years old	-0.17	0.1	-0.32***	0.09	-0.08	*	0.04	-0.10	**	0.03
40-49 years old	-0.14	0.1	-0.35***	0.1	-0.06		0.04	-0.11	**	0.04
50-59 years old	-0.11	0.11	-0.26*	0.11	-0.05		0.04	-0.09	*	0.04
60+ years old	-0.06	0.12	-0.32**	0.12	-0.03		0.05	-0.11	*	0.04
Household type (Single)										
Couple, no children	0.01	0.06	-0.03	0.06	0.00		0.02	-0.01		0.02
Couple with children	0	0.08	-0.09	0.09	0.00		0.03	-0.03		0.03
Lone parent	0.02	0.09	-0.05	0.09	0.00		0.03	-0.02		0.03
Other household type	0.16	0.12	0.05	0.13	0.06		0.05	0.01		0.04
Household size	-0.01	0.02	-0.03	0.02	0.00	***	0.01	-0.01		0.01
England	0.02	0.06	0.04	0.07	0.01		0.02	0.01		0.02
London/Southeast	0.15*	0.07	0.16*	0.07	0.06		0.03	0.05	*	0.02
Degree or higher	0.09	0.05	0.16**	0.05	0.03	*	0.02	0.05	**	0.02
Household income (log)	0.02	0.03	0.04	0.03	0.01		0.01	0.01		0.01

(continues)

Appendix 4. continued

	Pr(Health consent)		Pr(Benefit consent)		Pr(Health consent)			Pr(Benefit consent)		
	Coefficient	***S.E.***	***Coefficient***	***S.E.***	***Margin***		***S.E.***	***Margin***		***S.E.***
Refused question: Income from investment	-0.62***	0.13	-0.75***	0.14	-0.20		0.03	-0.19	***	0.02
Generally trusts others	0.22***	0.04	0.18***	0.05	0.08	***	0.02	0.06	***	0.02
Supports leftwing/liberal party	0.17***	0.04	0.20***	0.04	0.06	***	0.01	0.06	***	0.01
Does unpaid voluntary work	0.12**	0.05	0.10*	0.05	0.05	***	0.02	0.04	*	0.02
Has health problems	0.07	0.04	0.07	0.04	0.02	*	0.02	0.03		0.01
Has been to hospital	0.13	0.07	0.03	0.07	0.05		0.03	0.00		0.02
Receives any state benefits	-0.01	0.05	0.03	0.05	0.00	*	0.02	0.01		0.02
Number of means-tested benefits received	0.04*	0.02	0.06**	0.02	0.01		0.01	0.02	**	0.01
Survey design features										
Number of previous interviews in household	-0.75***	0.06	-0.69***	0.07	-0.28		0.02	-0.23	***	0.02
Number of previous health consents in household	1.25***	0.1	0.43***	0.09	0.42	***	0.03	0.20	***	0.03
Number of previous benefit consents in household	0.28**	0.09	1.07***	0.09	0.15	***	0.03	0.30	***	0.02
Others present during the interview	0.02	0.04	0.08	0.04	0.01	***	0.01	0.03	*	0.01
Number of years in the BHPS	-0.02***	0.01	-0.02***	0.01	-0.01		0.00	-0.01	***	0.00
Same interviewer as previous wave	0.08	0.08	0.13	0.09	0.03	**	0.03	0.05		0.03

(continues)

Appendix 4. continued

	Pr(Health consent)		Pr(Benefit consent)		Pr(Health consent)			Pr(Benefit consent)		
	Coefficient	***S.E.***	***Coefficient***	***S.E.***	***Margin***		***S.E.***	***Margin***		***S.E.***
Interviewer characteristics										
Male interviewer	-0.04	0.09	-0.11	0.09	-0.01		0.03	-0.03		0.03
Interviewer age group (40-49 years old)										
50-59 years old	0.04	0.12	0.03	0.12	0.02	***	0.05	0.02		0.04
60-69 years old	0.1	0.12	0.08	0.12	0.04		0.05	0.03		0.04
70+ years old	-0.06	0.15	-0.12	0.15	-0.01		0.06	-0.04		0.05
Interviewer has degree or above	0.06	0.06	0.03	0.06	0.02		0.02	0.01		0.02
Interviewing experience in years	0.01	0.05	-0.02	0.05	0.00	***	0.02	-0.01		0.01
Number previous interviews by interviewer	-0.02***	0.00	-0.02***	0.00	-0.01		0.00	-0.01	***	0.00
Number of health consents interviewer already obtained	0.03**	0.01	0.00	0.01	0.01	***	0.00	0.00		0.00
Number of benefit consents interviewer already obtained	0	0.01	0.03**	0.01	0.00	**	0.00	0.01	**	0.00
All can be persuaded	-0.05	0.05	-0.08	0.05	-0.02		0.02	-0.03		0.02
Should persuade	-0.01	0.04	-0.01	0.04	0.00		0.02	0.00		0.01
Should respect privacy	-0.07	0.06	-0.09	0.06	-0.03		0.02	-0.04		0.02
Should accept refusal	0.06	0.04	0.08	0.05	0.02		0.02	0.02		0.01
Emphasise voluntary nature	-0.06	0.05	-0.04	0.05	-0.02		0.02	-0.01		0.02
Agreeableness	0.03	0.04	0.03	0.04	0.01		0.01	0.01		0.01
Conscientiousness	0.03	0.04	0.00	0.04	0.01		0.01	0.00		0.01
Openness	-0.01	0.03	0.01	0.03	-0.01		0.01	0.00		0.01
Extraversion	-0.05*	0.02	-0.05	0.03	-0.02		0.01	-0.02		0.01
Neuroticism	-0.03	0.03	-0.02	0.03	-0.01		0.01	-0.01		0.01

Notes: The Log(Pseudo)Likelihood in the health and benefit consent models are -3069.8 and -2862.1, respectively.
Source: BHPS Wave 18 matched with interviewer survey.