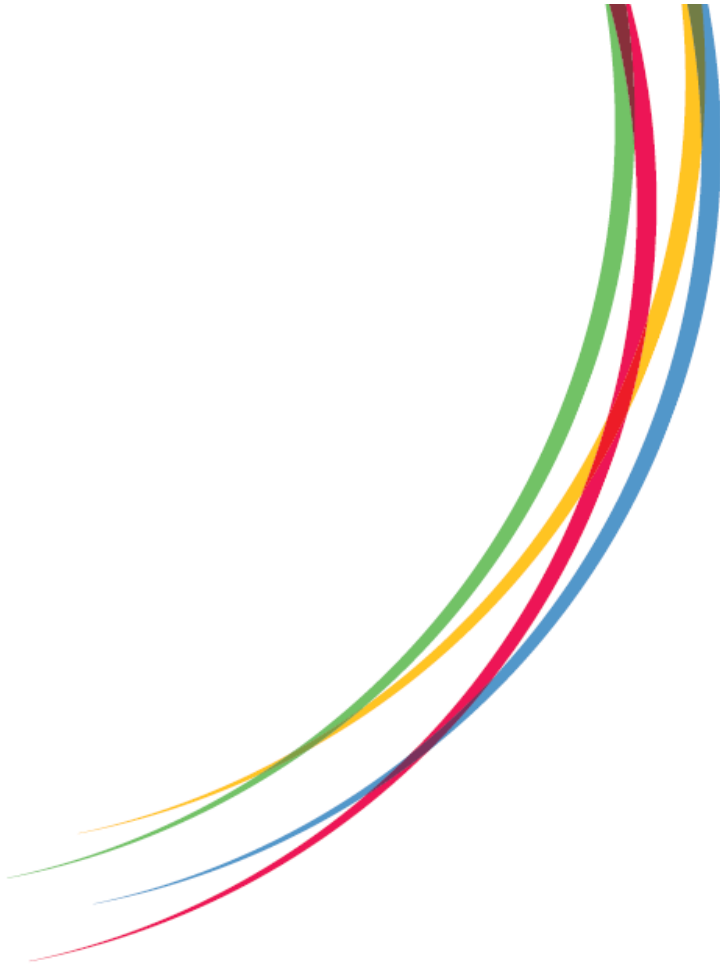




Understanding Society
Working Paper Series

No. 2014 – 09

December 2014

**Linking Administrative Records to Surveys:
Differences in the Correlates to Consent
Decisions**

Tarek Al Baghal, Gundi Knies, & Jonathan Burton

Institute for Social and Economic Research

University of Essex

Non-technical Summary

Fewer people are responding to surveys than ever before. At the same time, there has been an increased demand for the production and availability of data. The linking of administrative records to survey data is seen as an important tool to conduct research. Linking data from a person's administrative records to their survey responses has several possible benefits. For example, it increases the amount of possible information available. Linked data may also be more accurate than asking people more questions. Related to this, it requires fewer questions be asked to the survey participants, making it easier to conduct the survey. However, research on asking for people's consent to link their administrative data to their survey responses is still relatively new.

This research provides additional understanding on consent to data linkage in the survey context by using data from the first wave of Understanding Society: the UK Household Longitudinal Study (UKHLS). The study is a large, nationally representative annual survey in the UK. At the first wave all survey respondents 16 years old and over were asked if they gave consent for to their health and education records to be linked to their survey responses. This study adds to the understanding of obtaining consent for administrative data linkage for health and education records from adults. The goal is to identify possible reasons why people may or may not consent to link their administrative record and to find any differences between types of people more or less likely to consent. By examining multiple data linkage requests, it may be possible to find if there are differences based on the types of records asked for and if there is any consistent results for different consent questions.

Analyses examine whether consent rates differ by people's social and economic backgrounds, as well as if there is any relationship between people's attitudes and the consent decision. Household composition is also examined, to see how living environment influences decisions. Finally, information about the interviewer conducting the survey is included, because the interviewer may influence decisions as well.

There are some results of note. People are less likely to consent to health than to education record linkage. Importantly, no other of the variables consistently appears to be related to consent decisions. Generally, but not always, ethnic minorities are significantly less likely to consent to link their administrative records. Those supporting a liberal political party are frequently more likely to consent. Generally longer interviews lead to greater consent rates, suggesting the possibility that the interviewer and respondent build a rapport than increase the chances of consenting to requests for additional data. Conclusions are made about the possible biases in linked record data, and what the findings suggest about how it may be possible to improve data linkage requests. Importantly, the findings suggest that more can be learned about why or why not people consent to link their administrative records to their survey answers.

Linking Administrative Records to Surveys: Differences in the Correlates to Consent Decisions

Tarek Al Baghal, Gundi Knies, and Jonathan Burton

Abstract

With decreasing survey response rates, increased costs, and tightened survey budgets, the linking of administrative records to survey data is seen as an important tool for research. The current study analyzes consent decisions in the first wave of Understanding Society: the UK Household Longitudinal Study. This study examines the decision to consent through theories of nonresponse such as those of Groves and Couper (1998). People are less likely to consent to health than to education records. However, there are few variables consistently related to consent. Results suggest that there is still much to be learned about the consent decision.

Key words: data linkage, consent decision

JEL classifications: C81, C83

Author contact details: talbag@essex.ac.uk

Acknowledgements: This work was supported by the “Understanding Non-response on Understanding Society” project of the National Centre for Research Methods (NCRM) programme [grant number: DU/512589103]. Additional support was received from the Understanding Society project [grant number: ES/K005146/1]. We are grateful for comments received from our colleague Peter Lynn.

1. Introduction

In the era of “big data”, coinciding with decreasing survey response rates and tightening survey budgets, the linking of administrative records to survey data is seen as a possibly important tool to conduct research. Besides lowering survey burden (for both the researcher and respondent) by reducing the demands of information being requested, administrative data expands the research questions that can be explored. Linked administrative-survey data have been used to examine substantive issues such as healthcare spending and economic planning (e.g. Hogan et al. 2001, Scholz et al. 2006), as well as methodological issues, such as survey measurement error (Kreuter et al. 2010, Olson 2006, Sakshaug et al. 2010). However, the literature on data linkage requests within a survey context is still relatively nascent. This is particularly true for longitudinal studies, and for studies where consent is elicited from more than one member of the household.

Although there are a variety of linkage methods, here we focus on directly asking for an individual’s personal records. While the usefulness of such data may be apparent, obtaining it is not direct. Importantly, in most instances informed consent is required to obtain and link administrative data. Many studies asking for this informed consent have found that, like the survey request itself, a substantial portion of the sample do not affirmatively respond, possibly introducing a non-consent bias. The reported rates of consent have varied widely (see da Silva et al. 2012, Kho et al. 2009, and Sakshaug et al. 2012 for reviews). Consent rates obtained have been as low as 19.0% (McCarthy et al. 1999) and as high as 96.5 % (Rhoades and Fung 2004).

Consent rates vary over which types of records are being requested (e.g. health, benefits, tax) and can vary by sociodemographic characteristics. A significant number of studies examine consent to data linkage focusing on sociodemographic variables to understand differences in the likelihood to consent and possible groups that are under-

represented in the linked data. Findings are not always consistent across studies. For example, several studies have found females less likely to consent (Knies et al. 2012, Sala et al. 2012), others have found females more likely to consent (Dunn et al. 2004, Bryant et al. 2006), while others have found no differences between sexes (Huang et al. 2007, Korbmacher and Schroeder 2013, Sakshaug et al. 2012, Sakshaug et al. 2013). Although some studies have found no significant differences across ethnic groups, when ethnicity has been found to be related to consent rates, it is with minority groups providing significantly lower levels of consent (Kho et al. 2009, Knies et al. 2012, Knies and Burton 2014). Several review articles have identified the differing effects found across studies (Bohensky et al. 2010, da Silva et al. 2012, Kho et al. 2009).

Besides basic sociodemographics, several other variables are found to significantly impact the likelihood to consent. Although the effects of health status (examined for health record linkage consent) is varied like other demographics (see Kho et al. 2009), when significant findings are found, these tend to show that those with worse health are more likely to consent (Carter et al. 2010, Knies et al. 2012, Knies and Burton 2014). More clearly impacting consent are expressions relating to privacy and trust. Respondents who are more risk averse, or express more concern over privacy or display lower trust levels, either indirectly (nonresponse to questions such as income) or directly (reports of general trust), are less likely to consent (Korbmacher and Schroeder 2013, Sakshaug et al. 2012, Sala et al. 2012). Sala et al. (2012) have also found that support for a liberal party (compared to conservative or no party support) is related to consent rates. Besides simple party support, it may be that political engagement generally is important to consent agreement, as less politically engaged people have been found to be less likely to accept the survey request (Keeter et al. 2006).

Consent requests within a survey are possibly explicable through the framework of survey response itself, as laid out by Groves and Couper (1998). This framework suggests that multiple factors influence the decision to accept a request or not; not only the respondent, but also the respondent's environment and the survey features. Given the importance of these additional factors, recent studies also examine the influence of environment and survey variables. At the level of the respondent's environment, the focus has been on the household. Studies have included household characteristics as possible predictors, with some finding significant effects while others do not. Sala et al. (2012) find no differences in household composition, while Jenkins et al. (2006) find, in some cases, that single resident households and those with more children are less likely to consent.

Importantly, Sala et al. (2012) argue for a household "contagion" process (also suggested but not examined in Korbmacher and Schroeder (2013)), whereby initial consent outcomes affect later consent decisions within a household. The argument for contagion is twofold. First, respondents consult each other about such decisions (cf. Sala et al. 2012 who find that there is no association between the presence of others during the interview and consent). Second, given members of a household tend to be generally similar to one another in a variety of ways (Gulliford et al. 1999, Lohr 2010), in the case of interviewer-administered surveys, decisions by other household members signal to the interviewer how easy or difficult it is to get consent, which subsequently impacts their decision to ask for consent or to skip the question. However, if the predisposition to consent is a shared trait of all household members, then ordering within a household should not matter, although it does still suggest the importance of the household in consent decisions. This importance may be particularly important in domains of data linkage where one person's consent to data linkage is likely to reveal information about the household context in which she lives (such as entitlements to income maintenance and social security benefits that are means-tested). In the

case where multiple respondents within a household are asked for consent, then controlling for this clustering is necessary in analysis (Lohr 2010).

The interview process and interviewer characteristics also have impacts on consent outcomes, further showing the importance of factors beyond the respondents themselves. Resistance to taking the survey may also be related to resistance to consent (Sakshaug et al. 2012). Conversely, greater interviewer-respondent rapport is likely to increase the likelihood of consent (Jenkins et al. 2006, Korbmacher and Schroeder 2013). Several studies have further shown the influence of interviewers on consent outcomes, but few of the observed interviewer characteristics have a significant impact on the respondent's decision (Sakshaug et al. 2012, Sakshaug et al. 2013, Sala et al. 2012), although some effect for age has been found in other studies (Korbmacher and Schroeder 2013).

Overall interviewer experience somewhat paradoxically appears to have a negative impact on consent outcomes (Korbmacher and Schroeder 2013, Sakshaug et al. 2013, Sala et al. 2012), although Sakshaug et al. (2012) has found no effect of interviewer experience. However, past success appears have a strong positive impact – the more successful an interviewer has been in the past indicates success in the future (Korbmacher and Schroeder 2013, Sala et al. 2012, Sakshaug et al. 2012). As with the case of number of previous consents within households, however, past successes of the interviewer is likely an endogenous measure (Korbmacher and Schroeder 2013). This endogeneity is likely given a dependency of past and current success on some shared underlying trait, rather than past success causing current success.

The current study adds to the understanding of record linkage consent in a survey context by extending analyses in a several ways as well as adding a new data point using a new data set. Understanding Society: the UK Household Longitudinal Study is a large, nationally representative panel survey in the UK. At the first wave of data collection eligible

adult respondents were asked for consent to link their health and education records to their survey responses. We examine the response to two requests as a joint process, suggesting that there are some who are likely to consistently consent to requests, some who consistently do not consent, and some individuals who consent for some requests but not others. This method is somewhat different than Jenkins et al. (2006) and Sala et al. (2012), who use techniques to estimate separate models with correlated binary outcomes. This study also accounts for the household clustering more appropriately than previous studies, while attempting to reduce the use of possibly endogenous variables in the statistical analyses. We conclude with a discussion about what the results suggest about data linkage in the survey context, and possible paths for improvement.

2. *Data and Methods*

Understanding Society is a large (~40,000 households at Wave 1) annual longitudinal survey intended to collect quality data about a variety of issues in order to understand long-term effects of social and economic change in the UK. It covers all parts of the UK (England, Scotland, Wales, and Northern Ireland). The design is such that each wave of the survey will be conducted over a two-year period. The first wave of the survey was conducted over 2009 and 2010.

There are three samples used for the Understanding Society survey, a large General Population Sample (GPS) plus the Ethnic Minority Boost (EMB) Sample and the former British Household Panel Survey sample (BHPS). The BHPS sample was not integrated until the second wave of Understanding Society, and is not considered further here. The EMB sample was designed to provide at least 1,000 interviewed adults from each of five groups: Indian, Pakistani, Bangladeshi, Caribbean, and Africans. Understanding Society employs a complex survey sample, employing a stratified-clustered design selected through probability

proportionate to size (PPS) methods, with the goal of representing the population of the UK as well the ability to examine minority populations in depth. More information of the sample design and compositions can be found in Lynn (2009) and in Knies (2014).

The survey was conducted using face-to-face computer assisted personal interviewing (CAPI). Households were contacted first by a notification letter sent to the selected address, addressed to “The Occupier” along with a leaflet about the survey. An interviewer then called at the household within one week of the mailing. If no contact was made, a minimum of six calls were required before a household could be classified as a non-contact (interviewers could do more than six). A special conversion leaflet was sent to those households not considered “hard refusals”, and an interviewer returned to attempt the survey again.

If an adult at the household verbally consented, first a roster of all household members was collected. Interviews were then attempted for all members of the household aged 16 and over. Proxy interviews were also accepted for those not able to be interviewed in person. The household response rate for the GPS (including households providing at least one survey) is 57.3% (57.1% in Great Britain, 60.9% in Northern Ireland). The individual response rate conditional on household acceptance to be part of the panel is 81.8% (82% in Great Britain, 77.3% in Northern Ireland). The EMB sample had lower response rates: a 39.9% household response rate and a 72.4% conditional individual response rate.

At the end of the survey, respondents were asked if they would consent to link their administrative data to their survey responses (exact wording can be found in Appendix A). They were first asked about linking their health records from the National Health Service (NHS) and then asked about consent for linkage to their education records. Due to the nature of the records maintained at the administrative level, education record-linkage consent was asked only to those born after 1981 and attended school in the UK.

Several variables are also examined for potentially important relationships. Variables which are both of theoretical and empirical interest are included (complete description of variables is included in Appendix A). Sociodemographic variables include: sex, whether the respondent was born in the UK or not, ethnicity, age, employment status, educational attainment, whether they currently cohabit with a partner or not, whether they have children under 15 they are currently responsible for or not, and whether or not they receive any state benefits. Although many of these have been found unrelated to consent, some have, and a different survey context may provide different outcomes.

Besides sociodemographics, several other respondent factors may impact the likelihood to consent. One factor is the respondent's risk aversion and trust levels. Variables are therefore included regarding the respondent's general inclination to take or avoid risks and the respondent's attitude towards trusting others (both on 11-point scales). Health status has been shown to impact consent rates, although not always in a consistent direction (Kho et al. 2009). As such, SF-12 physical health scores are included as a covariate. Expanding the partisanship idea of Sala et al. (2012), indicators of explicit support for both left-leaning and right-leaning party support are included, to see if the driving factor is simply political engagement rather than actual party support.¹

Given the importance of household factors in consent, several household level variables are included. Home ownership and household income (in thousands of GBP) are included, as is the total number of members in the household (capped at 10 to control outliers), which may indicate differences in household composition and environment. Finally, an indicator for the household being in London is included, given the noted relationship between urbanicity and consent (Korbmacher and Schroeder 2013).

¹To avoid missing data problems with this variable, anyone not explicitly stating support for a party is coded as a 0, hence a measure of explicit support.

The last set of variables is related to the interview process and the interviewer, given the potential importance these have on consent. First is the number of calls to the household², which may indicate survey resistance, with more expected to be related to lower consent likelihood (Sakshaug et al. 2012). Second is the estimated length of the interview (by the interviewer)³, possibly indicating respondent-interviewer rapport, with greater rapport (longer interviews) expected to increase consent likelihood (Jenkins et al. 2006). Third, an indicator of whether others are present during the interview is included, as it may impact decisions, although there is not a significant effect in other studies (Sala et al. 2012). Finally, although previous consents in a household may be endogenous, the serial order of interviews may reflect the interviewer's experience within a household environment.

The interviewer demographics available and included are age, sex, and ethnicity. About 91% of interviewers are white, so ethnicity is coded as white or not. Experience as an interviewer at the research company is also included.⁴ Two additional derived measures are included. First is interviewer experience on Understanding Society, indicated by the number of interviews completed. Second is the interviewer's achieved response rate in Understanding Society, used as an indicator of interviewer ability. Although this measure is also possibly endogenous, it is more general and less obviously linked than previous consent outcomes. After the interview, the interviewer answered several questions about their observations about the interview, such as the presence of others and the estimated duration of the survey. In addition, call records and interviewing timing data was captured in the survey program, while basic demographics for the interviewer were also obtained from the research company.

2 There was missingness in this data, and thus mean imputation of calls (mean = 4.71) is used.

3 Recorded timings in the data had numerous apparent errors, thus the estimated interview length is used.

4 Several interviewers had no available data for these demographics, and mean imputation was used for these.

3. Results

There are 47732 individual interviews completed in the first wave of Understanding Society (not including proxy interviews). For the health consent question, 173 (0.36%) cases are missing consent values through errors in the survey process, while 1527 cases (16.35%) are similarly missing for the education consent question, and are not considered in further analyses of these questions. This leaves 47559 respondents asked for their consent to link their health records, and 6078 asked for consent to link their education records, and 6075 who were asked for consent to both. Table 1 presents overall consent rate for linkage to health and education records, broken down by age, ethnicity, sex, and number of children. It can be seen that the only statistically significant differences for these demographics is that of ethnicity, where, as found in previous consent research, minorities consent to record linkage at a lower rate. It is also evident that respondents are more likely to consent to the education record linkage than to the health record linkage. This higher consent rate for education linkage holds across all demographic characteristics in the table. Examining only those who answered both consent questions shows that in general these result remains, as seen in Table 2, below. Considering the consent rates jointly allows for restriction to one population, removing the possibility of differences due to sample composition.

Table 1: Consent Rates for Health and Education Record Linkage, by Demographics

	Health Records	Education Records
<i>Age</i>		
16-19	71.7% (n = 2976)	79.1% (n=2871)
20-29	67.6% (n=7447)	76.7% (n=3207)
30-39	64.1% (n=8714)	-
40-49	66.6% (n=9061)	-
50-59	67.8% (n=7214)	-
60-69	67.4% (n=6388)	-
70 +	67.1% (n=5759)	-
<i>Ethnicity</i>		
British/Irish White	70.4% (n=36525)	79.8% (n=4499)
Black	54.7% (n=3054)	70.9% (n=477)
South Asian	54.8% (n=4430)	71.6% (n=740)
Other Race	57.7% (n=3524)	75.1% (n=361)
<i>Sex</i>		
Male	67.5% (n=20959)	77.9% (n=2722)
Female	66.51% (n=26600)	77.8% (n=3356)
<i># of Children</i>		
0	66.9% (n=38839)	78.8% (n=5401)
1	67.7% (n=4179)	69.9% (n=472)
2+	66.2% (n=4541)	70.7% (n=205)
Total	66.9% (n=47559)	77.9% (n=6078)

n represent the total number of respondent in that classification

Table 2: Joint Health and Education Record Linkage, by Demographics

	Neither	Education Only	Health Only	Both Health and Education
Age				
16-19 (n=2870)	17.3%	10.4%	3.6%	68.8%
20-29 (n=3205)	17.5%	11.2%	5.8%	65.6%
Ethnicity				
White (n=4498)	15.1%	9.8%	5.1%	70.1%
Black (n=477)	24.3%	13.2%	4.8%	57.7%
South Asian (n=739)	25.4%	14.8%	3.0%	56.8%
Other Race (n=360)	20.3%	11.7%	4.7%	63.3%
Sex				
Male (n=2720)	17.7%	10.4%	4.4%	67.5%
Female (n=3355)	17.2%	11.1%	5.0%	66.7%
# of Children				
0 (n=5398)	17.1%	11.4%	4.1%	67.4%
1 (n=472)	20.6%	9.5%	6.4%	63.6%
2+ (n=205)	18.5%	10.7%	4.4%	66.3%
Total (n=6075)	17.4%	10.8%	4.8%	67.1%

n represent the total number of respondent in that classification

The breakdown in Table 2 is consistent with the findings of Table 1. Consent for education (total of education only and both column) is higher than consent for health (health only and both column). Also, since the majority of all respondents consented to both education and health, it suggests some people are more predisposed to consent, regardless of the content of records requested. Again, minority ethnic groups are less likely to consent to either request. The differences across ethnic groups in the consent to both is driven mostly by refusal to both, although it appears that some of this difference is attributable to minority groups having more education-only consenters (noting the small cell sizes).

In order to more clearly understand the patterns and mechanisms of consent in the adult population, mixed-effects logistic regressions are employed. Mixed-effects models are employed to capture the impact of both respondent, household and interviewer effects. For the health and education records consent questions, separate binary logistic random-intercept models are used. Combining the decisions as a joint outcome provides a four-outcome variable with no ordering. This implies the use of a multinomial random-intercepts model. The multinomial model allows for comparison of those who did not consent to either question as the baseline to those who consented to the education record request only, the health record request only, and those who consented to both requests. Using fixed-effects multinomial logistic regression models using variables included in the final analyses, the Hausman test of independence of irrelevant alternatives fail to reject the null hypotheses, suggesting the decisions can be included as separate decisions (i.e. uncorrelated errors) and use of the multinomial model is acceptable (Greene 2000).

In terms of estimation, respondents are clustered within households, which are clustered within interviewers. The survey was conducted with mainly one interviewer per cluster, and the lack of interpenetration does not allow for separation of interviewer and PSU effects (Campanelli and O'Muircheartaigh 1999). However, the inclusion of interviewers as

a variance component in the model controls for much of the clustering in the sampling design, and examination of possible interviewer effects. Stratification is not incorporated, but given that the expectation that stratification reduces variance estimates, significance tests are likely conservative. As a first step, random-intercepts only (i.e. null) models are estimated to calculate variance components and the intra-class correlation (ICC) coefficients. Only cases also used in the full models are included; however, it is important to note that nested multilevel models using the same categorical outcome are not strictly comparable (Bauer 2009, Hox 2010). The variance components and ICC coefficients of the null model are presented in Table 3.

Table 3. Variance and ICC parameters from Intercepts-Only Models

	Variance	ICC
<i>Health Records Model</i>		
Household	9.935	0.790
Interviewer	2.414	0.154
<i>Education Records Model</i>		
Household	11.021	0.826
Interviewer	4.605	0.243
<i>Joint Consent Model</i>		
Household	10.253	0.809
Interviewer	3.655	0.213

The results show that first, respondents within a household are very similar in their decision to consent, suggesting that consent is largely driven by factors shared among household members. Second, interviewers have an important role in the consent decision, although less prominent role than household factors. The null model ICC for the health records consent question is similar to that for employment record consent in Sakshaug et al. (2013) (ICC = 0.154), with the education and joint consent interviewer ICCs being somewhat

larger. All three ICCs are substantially lower than found for consent in Korbmacher and Schroeder (2013).

In order to better understand these effects and respondent characteristics in the consent decision, full models are estimated by including a number of potentially important variables. Respondent sociodemographics, attitudes to risk and trust, perceived health, and political affiliation are included, as are several household characteristics. Interviewer demographics are included at the third-level, along with interviewer response rates and previous experience with the Understanding Society survey. Results for the models including these variables are included in Table 4.

Table 4. Odds Ratios from Multilevel Models Predicting Consent

	Health	Education	Multinomial Model (Comparison: Non-consenters)		
			Education Only	Health Only	Both
<i>Respondent Characteristics</i>					
Female	0.881*	1.412	1.585*	1.417	1.379
UK Born	1.066	1.026	1.425	1.068	1.086
No Benefits	0.880	0.396*	0.648	1.052	0.456*
Partner	0.940	0.911	1.544	1.994	1.331
Age	0.993*	0.868*	0.943	1.061	0.937
No Children	0.979	1.626*	1.178*	0.937	3.064*
SF-12 Physical	0.997	0.991	0.994	0.972	0.983
Risk Taking	1.020	0.990	1.007	1.088	1.028
Trust in People	1.106*	1.068	1.068	1.090	1.082
<i>Employment Status (Not in Labour Force)</i>					
Employed	0.951	1.124	0.978	0.764	0.940
Unemployed	1.092	0.790	0.909	1.415	0.941
<i>Ethnicity (Other ethnicity)</i>					
British/Irish White	2.424*	1.633	1.642	2.134	1.822
Black	0.678*	0.541	0.704	0.562	0.620
South Asian	0.611*	0.458	0.690	0.564	0.416
<i>Education (Less than professional)</i>					
College Degree	0.721*	1.320	1.721	1.244	1.169
Professional	0.925	0.542	0.668	0.851	0.595
<i>Political Support (No Party)</i>					
Right-Leaning	1.400*	1.670	1.669	1.594	1.741
Left-Leaning	1.853*	2.891*	1.962*	1.127	2.401*

<i>Household Characteristics</i>					
Household Size	1.029	1.073	1.041	0.996	1.072
London	0.569*	0.766	0.896	0.893	0.804
Household Income	1.083	1.101	1.169*	1.164*	1.151*
Own Home	0.605*	0.697	0.584*	0.509*	0.552*
<i>Survey Environment</i>					
Serial Position HH	1.059	1.069	1.026	0.933	1.018
Number of Calls	0.916*	0.963	0.995	0.997	0.976
Others Present	1.083	0.857	0.942	1.146	0.902
Interview Length	1.012*	1.024*	1.023*	1.003	1.018
Interviewer-White	0.816	0.557	0.393	0.403	0.513
Interviewer-Age	1.005	1.027	1.034	1.043*	1.035*
Interviewer-Female	1.108	1.440	1.296	1.007	1.279
Years as Interviewer	0.971	0.928*	0.931*	0.964	0.932
Number of UKHLS Interviews Completed	0.999	1.001	1.002	1.001	1.002
Interviewer Response Rate	0.447*	1.116	1.372	1.032	1.084
<i>Random-effects Parameters</i>					
Household Variance	9.635	11.376		11.404	
Interviewer Variance	2.229	4.669		3.549	
n Respondents	29413	4446		4446	
n Households	17053	3433		3433	
n Interviewers	694	612		612	

There are several important findings based on these results. First is that even with the large number of independent variables included in the model, both the interviewer and particularly the household levels contribute largely to the model variance. The increases in these variances relative to the null models for the education only and joint models is reasonable given the estimation procedures of categorical multilevel models and the non-comparability of these nested models (Bauer 2009, Hox 2010). Korbmacher and Schroeder (2013) also found that including additional variables still left a considerable interviewer component. Along with that finding, results here suggest that there is still a large unobserved heterogeneity in both households and interviewers that matters in consent decisions.

The second finding of interest is that there is no consistent pattern in which variables are significant across models. This lack of consistency suggests the possibility that characteristics affecting the consent decision depends on what records will be linked. Although the sample differs in who is included in the models for health records and education records, corresponding to the first two columns of Table 4, as seen in Table 1, when examining only the youngest respondents, health records are significantly lower than consent rates for education records.

Further, the multinomial model presented in Table 4, corresponding to the final three columns, show a similar heterogeneity in effects, while restricting estimation to only those who were asked both consent questions, and is also the same respondents in the education records model. These findings indicate the possibility that the decision to consent not only differs across requests, but also that characteristics influence these outcomes in different ways. It is important to note when discussing these differences, that the given the age restriction for the education records and joint consent models, variables may measure aspects

differently. For example, employment may mean something different to someone between 16 and 29, compared to the full age spectrum.

These differences generally arise from characteristics apparently influencing consent in some cases, but not in others, rather than directional differences. The health records model has the several more significant effects overall, which may be due to the greater power. Examining the health records model to understand consent in for this request, as well as being the request made to the whole sample, shows the importance of several individual characteristics. First, females are less likely to consent than males, and health records are the only model with a significant effect for sex. Older respondents and university education also have lower odds to consent relative to younger and less educated. Ethnicity is significant, showing that British or Irish are the most likely to consent, whereas black and south Asians are similarly less likely to consent. Attitudes also have some influence, with those expressing more trust in people, and those politically engaged are more likely to consent. Political engagement, with support for any party as opposed to none, appears to increase consent, and the effect is potentially not simply partisanship as suggested in Sala et al. (2012).

Household environment also has some relationship with the consent decision. Being in London (i.e. an urban centre) decreases odds of consent, consistent with other findings (e.g. Korbmacher and Schroeder (2013)). The family owning the home (as opposed to renting or other circumstance) is also related to lower odds of consent. Taking together the effect of home ownership with the findings of lower odds for university graduates, these findings suggest that, at least for this sample, higher socioeconomic status may play a role in lower consent rates.

Survey factors, as found in other studies, also play a role in the consent process for health records in Understanding Society. A greater number of calls, a potential measure of survey resistance, are related to lower odds of consent as expected. Interview length, a

potential measure of interviewer-respondent rapport, is also significant in the expected direction, with longer interviews (greater rapport) related to higher odds of consent. Interviewer response rate, included as a measure of interviewer quality, however, is significant in the opposite direction of expectation as lower response rates are related to higher odds in consent. This opposite direction is unexpected as it suggests that there is possibly a different interviewer quality driving consent to conduct the survey and consent to link health records.

Both the education records and joint consent multinomial model are estimated on the same set of sample members, and contain fewer significant effects than the health records model. Even with the same sample subset, differences in effects are found across models. In the education records models, sex has no effect, but in the joint consent model, females have significantly higher odds of consenting to education only compared to those who refused both requests. Receiving benefits is associated with higher odds of consent in the education records model and for consent to both request in the joint consent model, but has no discernible impact on education or health only consent among these respondents.

Older respondents (again constrained by the 16-29 year range) are less likely to consent in the education records models, but no significant impact is found in the joint consent model. Having no children is related to higher odds of consent for everyone in these two models except health only consenters; this variable is similarly not significant in the health records model. Unlike the health consent model, it does appear that partisanship rather than simple political engagement is of importance, as those supporting left-leaning parties are more likely to consent for all except health-only consenters, but supporting right-leaning parties has no apparent impact across models.

No impacts are found in the education records model for household characteristics, but some are found in the joint consent model. Home ownership is negatively related all

levels of consent in the joint consent models, but household income has a significant effect in the opposite direction. Respondents in higher household incomes more likely to consent to either one or both requests compared to those who did not consent to either request. These results are somewhat oppositional, as both are intended to measure underlying economic status factors.

As with the health records model, the education and joint consent models suggest the influence of survey factors on consent as well. However, like with other characteristics, the findings are not consistent across models. Like the health records model, interviews estimated to be longer are related to higher odds of consent in the education records and education only component of the joint consent model, suggesting that greater rapport increasing consent rates in some instances. Years of experience as an interviewer are negatively related with consent only in the education records and education only component of the consent model. The findings here suggest that the impact of experience may affect decisions differently, in this particular instance, the education decision. Finally, older interviewers are more likely to obtain consent for health only and for both health and education in the joint records model, but not significant elsewhere. It is possible that interviewer age influences only younger respondents but only for some consent decisions (e.g. health).

4. Discussion

The current research adds to the nascent literature on consent within a survey context by extending previous findings and shedding light on consent decisions across different requests for linkage. Consent rates among observed sample members are generally high, constituting the majority of respondents. Consent rates are significantly lower for health records than for education records across all respondent divisions. However, consent rates among British or Irish whites are significantly higher than among other ethnicities,

particularly for health records. These differences in consent rates suggest potential differences in obtained records, particularly across ethnic groups.

To further understand the mechanisms of consent, models predicted consent outcomes using respondent characteristics, their environment, and survey factors. Importantly, analyses included random effects for both the household and interviewer components. While some previous studies included an interviewer component, none have included a household random effect. The results show that, similar to previous findings, consent is affected by several factors, including the respondent, but also their environment and survey-related factors. A large part of the variation in consent outcomes is related to household homogeneity, and should be controlled for and further explored in further studies. A variety of variables are found to be related to the different consent decisions which may help further the understanding of the correlates of consent to data linkage. The importance of survey environment is positive for researchers wanting to obtain consent for data linkage, as unlike respondent characteristics and their environment, the researcher has some control over these factors.

This outcome is tempered by the inconsistency in effects found, not only in this study, but across all studies on consent to data linkage. Unlike most previous studies, the current research examines two (rather than a single) consent decision. Models are employed that assume independence between these decisions, and analyses suggest that these decisions can be treated as independent. Findings further show that what is significant for one decision is not significant for another, and factors relating to consent to only one of the requests differ from those relating to consent to both requests. There is a lack of consistent effects for respondent characteristics, conforming with the general inconsistencies identified in the consent to data linkage literature (Bohensky et al. 2010, da Silva et al. 2012, Kho et al. 2009, Knies and Burton 2014). Similarly, survey variables that are related to consent are not

consistent across models in the current study. While the few studies examining survey features have found similar effects in some cases (e.g. experience), in others there is a lack of consistency (e.g. age, number of prior interviews).

Although the health records model is comprised of a different sample composition (education records are age restricted), the differences across these sample compositions further shows that there is not one set of correlates that extend to everyone. The lack of consistent effects found across decisions within respondents or across studies suggests that there is still much to learn about the mechanisms for consent to data linkage. Recent work suggests that some survey features can be altered to possibly influence consent, such as placement of the question (Sakshaug et al. 2013, Sala et al. 2013). Further work should identify the best placement of questions and other features under the researcher's control.

Efforts should also focus on the personal mechanisms causing respondent to consent or decline. The lack of consistent sociodemographic effects suggest that studies should focus on psychological factors. Psychological factors are theorized to drive the decision to consent to a survey request generally (Groves et al. 1992), while personality is found to be related panel attrition (Lugtig 2014). It may be, however, that the psychological factors underlying nonresponse is not exactly the same as consent refusal. The current results find that for health records, there is an inverse relationship between an interviewer's response rate and their consent rate, suggesting the possibility of different mechanisms. Differences may also be indicated by the fact that the person who refuses to consent in a survey is still a respondent, and the psychology of nonresponse is distinct.

If there is a better understanding of the psychological factors leading to declining data linkage, possible question designs could be considered to confront the problem. For example, expediency was highlighted in Sakshaug et al. (2013) to no effect, but confidentiality, purpose for linkage, or benefits to society may have a more positive influence. Initial

research does indeed suggest that confidentiality and benefits may be important reasons why people choose to consent or not (Sala et al. 2013). Additionally, further understanding of how these psychological differences across the population may be illuminating, as there are likely multiple reasons people refuse requests. If such differences are found, findings may allow for question tailoring in survey design. Without fully understanding why respondents decline, design choices will be led by supposition.

With lower consent rates, bias becomes an increasing possibility. One issue in discussing consent is that the rates presented in most survey linkage studies are conditional. For example, Table 1 presents the consent to health data linkage at 66.9%; however this conditional on the fact that a person consented to conduct the survey first. If the population of inference is observed sample members, this is the correct metric. However, if the population of inference is what the total sample is expected to represent, then the obtained records are significantly less.

In the current case, for health records, given a 57.2% household response rate, and 81.8% respondent within household response rate, and a 66.9% consent rate, the overall percentage of records obtainable for the total sample is 31.4%. This reduction is even further pronounced among the ethnic minority boost sample. Not only are they significantly less likely to consent, they are significantly less likely to respond. In the ethnic minority sample, 39.9% of households and 72.4% within households responded, and 56.1% of these consented, for an overall 16.2% of obtainable records over that sample. As found by Sakshaug and Kreuter (2012), both non-response and non-consent can lead to bias in administrative record data, although not necessarily for all variables.

Questions arise as how to best correct for these missing observations, such as how to best weight (if at all) administrative data to be representative to the population of inference. Weighting already is used to adjust for unit nonresponse (Brick 2013). However, like unit

nonresponse, as demonstrated through this and other studies, understanding of the causes of non-consent is still somewhat limited. However, for propensity-based weighting techniques such as those used for nonresponse, the predictors of nonresponse should be included in the model to correctly adjust for missing data (Brick 2013). The current research continues to add information about the consent process that may be useful for such techniques, but also underscores the continuing need for further research on the mechanisms of consent.

5. References

- Bauer, D.J. (2009). A Note on Comparing the Estimates of Models for Cluster-Correlated or Longitudinal Data with Binary or Ordinal Outcomes, *Psychometrika*, 74, 97-105.
- Bohensky, M.A., Jolley, D., Sundararajan, V., Evans, S., Pilcher, D.V., Scott, I., & Brand, C.A.(2010). Data Linkage: A powerful research tool with potential problems *BMC Health Services Research* 10, 346-353.
- Brick, J.M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review *Journal of Official Statistics*, 29, 329–353.
- Bryant H, Robson PJ, Ullman R, Friedenreich C, & Dawe U. (2006). Population-based cohort development in Alberta, Canada: A feasibility study. *Chronic Diseases in Canada* 27, 51-59.
- Campanelli, P. & O’Muircheartaigh, C. (1999). Interviewers, Interviewer Continuity, and Panel Survey Nonresponse *Quality & Quantity*, 33, 59–76.
- Carter, K., Shaw, C., Hayward, M. & Blakely, T. (2010). Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey *Kotuitui: New Zealand Journal of Social Sciences Online*, 5, 53-60.
- da Silva M.E., Coeli C.M., Ventura M., Palacios M., Magnanini M.M., Camargo T.M., and Camargo K.R. Jr. (2012). Informed consent for record linkage: a systematic review. *Journal of Medical Ethics*, 10, 639-642.
- Dunn K.M., Jordan K., Lacey R.J., Shapley M., Jinks C. (2004). Patterns of consent in epidemiologic research: Evidence from over 25,000 responders. *American Journal of Epidemiology*, 159, 1087-1094.
- Greene, W.H. (2012) *Econometric Analysis*, 7th Edition. Upper Saddle River, N.J.: Prentice Hall.
- Groves, R. M. & Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

- Groves, R. M., Cialdini, R.B., Couper, M.P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.
- Gulliford, M.C., Ukoumunne, O.C., & Chinn, S. (1999). Components of Variance and Intraclass Correlations for the Design of Community-based Surveys and Intervention Studies: Data from the Health Survey for England 1994, *American Journal of Epidemiology*, 149, 876-883.
- Hogan, C., Lunney, J., Gabel, J., & Lynn, J. (2001). Medicare beneficiaries' costs of care in the last year of life. *Health Affairs*, 20, 188-195.
- Hox, J.J. (2010). *Multilevel Analysis: Techniques and Applications*, 2nd Edition, New York: Rutledge.
- Huang N., Shih S-F, Chang H-Y, & Chou Y-J. (2007). Record linkage research and informed consent: who consents? *BMC Health Services Research*, 7, 18-23.
- Jenkins, S. P., Cappellari, L., Lynn, P., Jaekle, A., & Sala, E. (2006). Patterns of Consent: Evidence from a General Household Survey *Journal of the Royal Statistical Society (Series A)* 169, 701-22.
- Keeter, S., Kennedy, C., Dimock, M., Best, J. & Craighill, P. (2006) Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey *Public Opinion Quarterly* 70, 759-779.
- 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, Gundi (ed.) (2014). *Understanding Society –UK Household Longitudinal Study: Wave 1-4, 2009-2013, User Manual*. Colchester: University of Essex.
- Knies, G., & Burton, J. (2014). Analysis of four studies in a comparative framework reveals: health linkage consent rates on British cohort studies higher than on UK household panel surveys, *BMC Medical Research Methodology*, 14, 125.
- 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, 52-58.
- Korbmacher, J. M., & Schroeder M. (2013). Consent when linking survey data with administrative records: The role of the interviewer. *Survey Research methods*, 7, 115-131.
- Kreuter, F., Muller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: making use of administrative data. *Public Opinion Quarterly*, 74, 880-906.

- Lohr, S. L. (2010). *Sampling: Design and Analysis*, 2nd edition. Pacific Grove, CA: Duxbury Press.
- Lynn, P. (2009). Sample Design for Understanding Society *Understanding Society Working Paper Series No. 2009 – 01*, Institute for Social and Economic Research, University of Essex.
- Lutig, P. (2014) Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers *Sociological Methods & Research*, 43, 699-723
- McCarthy, D. B., Shatin, D., Drinkard, C.R., Kleinman, J.H. & Gardner, J.S. (1999). Medical Records and Privacy: Empirical Effects of Legislation. *Health Services Research* 34, 417-25.
- Olson, K. M. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly* 70, 737-58.
- Rhoades, A. E.& Fung, K. (2004). Self-Reported Use of Mental Health Services Versus Administrative Records: Care to Recall? *International Journal of Methods in Psychiatric Research* 13, 165-75.
- Sakshaug, J. W., Couper, M.P. & Ofstedal, M.B. (2010). Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults. *Medical Care* 48:64-71.
- Sakshaug, J. W., & Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods*, 6, 113-122.
- Sakshaug, J.W., Couper, M. P., Ofstedal, M. B., & Weir, D. (2012). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research*, 41, 535-569.
- 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, 133-144.
- Sala, E., J. Burton, & G. Knies. (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research* 41, 414-439.
- Sala E, Knies G, & Burton J: Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology* 2014, 17(5):455-473.
- Scholz, J., Seshadri, A., & Khitatrakun, S. (2006). Are Americans saving 'optimally' for retirement? *Journal of Political Economy*, 114, 607-643.

6. Appendix

A: Measures Used

Health Records Linkage Question

Finally, we would like to add some information from administrative health records to the answers you have given. This leaflet gives you information about what we would like to do. Please read it, ask me any questions and sign the form if you are happy for us to do this.

Education Record Linkage Question

We would also like to add information from your education records. Here is a permission form and information leaflet. Please read this, ask me any questions and sign the form if you are happy for us to do this.

Independent variables

Female = 1 if female, 0 if male

UK Born = 1 if born in UK, 0 if born anywhere else

No benefits = 1 if reported obtaining any of a list of benefits, 0 if any benefit received

Partner = 1 if reported currently cohabitating with a spouse/partner, 0 if not currently cohabitating

Age = Continuous measure of age, range 16-98

No Children = 1 if no children, 0 if 1 or more children

SF-12 Physical = SF-12 Physical Component Summary (PCS). This measure converts valid answers to the origin questions into a single physical functioning score, resulting in a continuous scale with a range of 0 (low functioning) to 100 (high functioning).

Risk Taking = "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (0= Avoid Taking Risks, 10 = Fully prepared to take risks)

Trust in People = "Are you generally a person who is fully prepared to take risks in trusting strangers or do you try to avoid taking such risks?" (0= Avoid taking risks in trusting strangers, 10 = Fully prepared to take risks in trusting strangers)

Employed = 1 if employed (full or part-time), 0 otherwise

Unemployed=1 if indicated unemployed but in labour force, 0 otherwise

British/Irish White = 1 if white from Great Britain or Ireland, 0 otherwise

Black = 1 if Mixed African, Mixed Caribbean, African, Caribbean, or Any other black background, 0 otherwise

South Asian = 1 if Indian, Pakistani, Bangladeshi, 0 otherwise

College Degree = 1 if has University Higher Degree (e.g. MSc, PhD), First degree level qualification including foundation degrees, graduate membership of a professional Institute, PGCE, 0 otherwise

Professional = 1 if Diploma in higher education, Teaching qualification (excluding PGCE), Nursing or other medical qualification, HNC/HND, 0 otherwise

Right-leaning = 1 if favours Conservative, Ulster Unionist, Democratic Unionist, 0 otherwise

Left-leaning = 1 if favours Labour, Liberal Democrat, Scottish National Party, Plaid Cymru, Green Party, SDLP, Alliance Party, Sinn Fein, 0 otherwise

Household Size = number of members living in household, capped at 10

London = 1 if household in London, 0 otherwise

Household income = Total reported household income, in thousands of GBP.

Own home = 1 if home is owned by household, 0 if not owned

Serial position in household = ordering of individual's completion of survey

Number of calls = number of calls to household until survey achieved

Others present = 1 if anyone else present during interview, 0 if no one else

Interviewer length = length of interview in minutes, estimated by interviewer

Interviewer White = 1 if interviewer is white, 0 otherwise

Interviewer age = Continuous measure of interviewer age, range 23-82

Interviewer female = 1 if interviewer is female, 0 if male

Years as interviewer = number of years as interviewer at research company

Number of UKHLS interviews completed = number of interviews, prior to the current one, that the interviewer has completed in the current survey

Interviewer response rate = proportion of successfully completed surveys of total outcomes (successfully completed surveys plus refusals and noncontacts at eligible households)