Published in final edited form as:

Sociol Methods Res. 2012 November ; 41(4): 535–569. doi:10.1177/0049124112460381.

LINKING SURVEY AND ADMINISTRATIVE RECORDS: MECHANISMS OF CONSENT

Joseph W. Sakshaug^{1,2}, Mick P. Couper³, Mary Beth Ofstedal³, and David R. Weir³

¹Institute for Employment Research, Nuremberg, Germany

²Department of Statistics, Ludwig-Maximilians University of Munich

³Institute for Social Research, University of Michigan

Abstract

Survey records are increasingly being linked to administrative databases to enhance the survey data and increase research opportunities for data users. A necessary prerequisite to linking survey and administrative records is obtaining informed consent from respondents. Obtaining consent from all respondents is a difficult challenge and one that faces significant resistance. Consequently, data linkage consent rates vary widely from study-to-study. Several studies have found significant differences between consenters and non-consenters on socio-demographic variables, but no study has investigated the underlying mechanisms of consent from a theory-driven perspective. In this study, we describe and test several hypotheses related to respondents' willingness to consent to an earnings and benefit data linkage request based on mechanisms related to financial uncertainty, privacy concerns, resistance towards the survey interview, level of attentiveness during the interview, the respondents' preexisting relationship with the administrative data agency, and matching respondents and interviewers on observable characteristics. The results point to several implications for survey practice and suggestions for future research.

Keywords

Consent; Administrative records; Data linkage; Mechanisms; Health and Retirement Study; Interviewer effects; Privacy; Attentiveness

INTRODUCTION

Sample surveys are facing significant challenges due to falling response rates (De Leeuw and Heer 2002; Curtin et al. 2005), noncoverage of the target population (Blumberg and Luke 2007), and increasing survey costs. All of these challenges have the potential to threaten the quality of the collected survey data and reduce confidence in the conclusions drawn from surveys. Despite these challenges, the demand for more detailed and comprehensive data structures has grown in the research and policy realms. To meet these demands, survey organizations may be forced to rely more heavily on complex, but less

proven, methodologies to overcome response rate problems, reduce survey costs, and improve the quality and utility of the data.

One promising method towards achieving these objectives is to link administrative data to survey data. Linking these data sources can significantly enhance the quality and utility of the survey data and enable data users to answer important substantive and methodological research questions that cannot be addressed using a single data source (Calderwood and Lessof 2009; Lillard and Farmer 1997). Administrative data linkage may also lead to shorter interviews, less respondent burden, and an overall reduction in survey costs. Furthermore, administrative data can potentially offer a significant increase in the number of auxiliary variables that may be used for nonresponse bias adjustment.

For most surveys, a necessary prerequisite to directly linking survey and individual-level administrative records is obtaining informed consent from respondents. Informed consent is needed to ensure that respondents are aware of the risks and benefits involved in releasing and linking their administrative records for research purposes (GAO 2001). Obtaining linkage consent from all respondents is a challenging task. Linkage consent rates (conditional on survey response) vary substantially within and between surveys thereby increasing the risk of bias for key estimates. Table 1 shows the variability in consent rates (and corresponding response rates) across different surveys and administrative data types. The range of consent rates is large, ranging from a low of 19.0% (McCarthy et al. 1999) and a high of 96.5% (Rhoades and Fung 2004).

Because respondent consent is not universal, there are concerns that inferences drawn from linked data sets may be biased. Several studies have found systematic differences between consenters and non-consenters based on socio-demographic characteristics (see Kho et al. 2009 and Dunn et al. 2004 for reviews), however, there is little consistency across studies. For example, some studies find that older persons are more likely to consent than younger persons (Bryant et al. 2006; Dunn et al. 2004), while other studies find that *younger* persons are more likely to consent than older persons (Huang et al. 2007; Yawn et al. 1998), and yet some studies find no age effect at all (Buckley et al. 2007; Harris et al. 2005). Gender and income exhibit similar inconsistent relationships with the likelihood of consent: gender and income are significantly related to consent in some studies (Huang et al. 2007; Olson 1999), but not others (Al-Shahi 2005; Harris et al. 2005). The inconsistent findings make it difficult to advance the study of linkage consent for purposes of minimizing bias and informing survey practice.

The inherently different populations that have been studied (e.g., patients, caregivers, older birth cohorts) may partially explain the lack of consistency across studies. The conflicting findings may also be driven by different underlying mechanisms of consent. The existing literature does not adequately develop and test theory-driven hypotheses on why some respondents consent to data linkage and others do not. For example, one hypothesized reason why respondents may consent to administrative data linkage is because they are unable to accurately self-report information contained in their administrative record (e.g., self-reports of unemployment spell duration in a labor market survey linked to employment records) and therefore find it easier to provide this information through data linkage.

Another hypothesis is that respondents refuse to consent due to privacy concerns or general resistance towards the survey interview. These two hypotheses have never been tested jointly on the same survey data set. Other hypothesized mechanisms of consent involve the effects of interviewers on influencing respondent consent, and whether respondents have a preexisting relationship with the administrative data agency, among others.

The purpose of this paper is to develop and test hypotheses of data linkage consent using a national population-based panel survey of older adults who were asked for consent to link earnings and benefit records at the end of the interview. In doing so, this paper fills an important research gap by utilizing theory-driven hypotheses to examine respondents' likelihood of linkage consent.

HYPOTHESIZED MECHANISMS OF CONSENT

Accuracy/Uncertainty

Administrative records (e.g., tax records, medical billing records) often contain detailed and relatively complex information collected over a long period of time. Such information can be burdensome for respondents to recall with great precision and expensive for survey organizations to collect during the survey interview. Still, it is common for surveys to ask respondents to report on *some* variables that can be found in administrative records (e.g., household income, health care utilization) in conjunction with the linkage request. A respondent who is unable to report their administrative information with certainty may be more likely to consent to the subsequent data linkage request in order to provide the survey organization with a more accurate assessment of their situation.^{1,2} This hypothesis of consent has never been tested.

Privacy/Confidentiality Concerns

One of the most frequently-posed hypotheses concerning why many respondents are unwilling to consent to administrative data linkage is because they are concerned about a potential breach of confidentiality, they fear that their records could be used for nefarious purposes, or feel that the information contained in the records is too personal to be released. Studies exploring this notion have the disadvantage of looking at relatively indirect indicators of privacy concerns, such as refusal to answer sensitive items (e.g., income; Jenkins et al. 2006; Sala et al. 2010), rather than direct or interviewer-observed assessments of such concerns. Obtaining more direct assessments of respondents' privacy concerns and relating them to their likelihood of consent would be useful for purposes of understanding when non-consent is likely to occur and the mechanism behind its occurrence.

¹Although it is rare that the consent request is framed in this way, panel survey respondents may be more likely to make this connection given their extensive commitment and engagement in the study.

²An alternative hypothesis posited by a reviewer is that it may not be the case that respondents are unable to report but rather they see the task as too burdensome and therefore are more inclined to allow consent to records. We suspect that this scenario is more plausible when the consent request is administered at the beginning of the interview or prior to the administrative questions, in which case the respondent knows whether the administrative information will be collected via linkage and therefore may opt not to self-report it. The survey that we use in this study asked for consent at the end of the interview, so we do not explicitly test this hypothesis here.

General Resistance towards the Survey Interview

During the survey recruitment process, sample members may be resistant to the interview request due to lack of interest, burden, and/or distrust of the survey sponsor, among other reasons (Groves and Couper 1998). Despite their initial resistance, many persons eventually agree to participate in the survey interview. It has been hypothesized that these initially reluctant respondents are less thoughtful (and less accurate) in answering survey questions (Fricker 2007; Olson 2006), it is also plausible that these resistant respondents may be less willing to comply with in-survey requests, such as linkage consent. For example, there is evidence in the biomarker consent literature that respondents who required the most contact attempts (an indirect indicator of resistance) or were observed by interviewers to be uncooperative during a prior-wave interview were less likely to consent to biomarker participation in the current wave (Sakshaug et al. 2010). Determining whether respondents who are initially resistant to the survey interview, but ultimately participate, are less likely to give linkage consent is an important step that could help survey practitioners to preidentify (and possibly intervene on) respondents who are potentially resistant to the linkage request.

Preexisting Relationship with the Administrative Data Agency

A respondent's relationship with the administrative data agency may also influence whether they consent to the linkage request. One hypothesized reason why respondents are more likely to consent to a specific linkage request is because they receive services or benefits from the administrative data agency. For example, respondents who are currently receiving social security payments may feel incentivized to consent to link their benefit records due to the saliency of the information being requested, familiarity with the government agency charged with administering their benefit, and reciprocation for the benefits received. There is some evidence in the medical literature that tangentially supports this hypothesis. Dunn et al. (2004) found that survey responders who reported the health symptoms under investigation were more likely to consent to a follow-up interview. Woolf et al. (2000) reported that people in poorer health were more likely to consent to review of medical records. Petty et al. (2001) found that people receiving more repeat prescriptions were more likely to consent to medication review. These studies suggest that respondents who receive services or benefits from the administrative data agency or who have experience or familiarity with the information being requested (e.g., social security records) are more likely to give consent. How this relates to linkage consent and whether respondents with a vested interest in the type of administrative information being requested are overrepresented in data linkage studies is unclear.

A counter hypothesis is that service or benefit receipt from the administrative data agency may actually reduce the likelihood of consent. Respondents who have a preexisting relationship with the agency may be more concerned about the potential linkage, especially if the relationship has been negative (e.g., loss of eligibility due to an adverse assets assessment by the agency). Furthermore, benefit recipients may be concerned about someone 'checking up' on them and examining whether they are claiming benefits fraudulently. This would motivate some respondents to refuse the linkage and negate the reciprocation effect put forward in the first hypothesis.

Acquiescence/Attentiveness

Several studies have shown that respondents have a tendency to acquiesce when presented with difficult survey questions (Gage et al. 1957; Hanley 1962; Trott and Jackson 1967). In general, this notion is related to the theory of survey “satisficing”, which suggests that subsets of respondents take various cognitive shortcuts that lead to suboptimal response formation (Krosnick 1991, 1999). Given the complexities involved with describing the purpose, process, and implications of data linkage, it is plausible that acquiescent or inattentive respondents may give consent without taking the time to consider all of the information provided to them about the linkage request. In a sense, these respondents are choosing the simpler task, which is the provision of consent over the more demanding task of carefully reading all of the linkage materials before they make a decision. An alternative hypothesis is that acquiescent respondents may be more likely to reject the consent request because they perceive that to be the easier task. The notions of acquiescence and satisficing have not been studied in the context of linkage consent.

Interviewer Effects

A relatively unexplored source of consent variation is that due to the interviewer. Interviewers are charged with administering the consent request, helping respondents understand what is being asked for, and addressing any related respondent concerns. Given that interviewers are usually not incentivized to obtain linkage consent from respondents nor are they instructed to attempt consent refusal conversions, it is unclear how effective they are in obtaining consent. It is also unclear whether interviewers vary in their ability to obtain consent in the same way that interviewers vary in their ability to recruit survey respondents (O’Muircheartaigh and Campanelli 1999), though this seems quite plausible. Furthermore, it is unknown whether the same interviewer attributes that have been linked to increases in survey participation are also successful in obtaining linkage consent. Some studies have shown that interviewer attitudes towards survey refusal conversion and interviewer personality traits [measured using the ‘Big Five’ personality inventory; (John and Srivastava 1999)] are positively related to recruitment of survey respondents (Kennickell 1999; Lehtonen 1996; Jäckle et al. 2010). Whether these interviewer traits are also related to the likelihood of consent is an open question, but most evidence points to the lack of a relationship. In a study of consent to link health and benefit records in the U.K., Sala et al. (2010) found no relationship between interviewers’ attitudes toward persuading reluctant persons to participate in the survey and their likelihood of obtaining consent from respondents. The authors also found no relationship between interviewer personality traits and respondents’ likelihood of consent. If, as the evidence suggests, interviewer personality and persuasion traits do not explain the significant interviewer-level variation in consent rates (Sala et al. 2010), then other potential interviewer-level attributes should be considered. Identifying such factors may lead to improved interviewer training resulting in higher consent rates.

One hypothesis of consent is that interviewers who possess critical views toward data privacy are less likely to obtain linkage consent from respondents, especially if those respondents possess similarly critical views. Interviewer demographic characteristics and their interaction with respondents may also play a role in the likelihood of obtaining consent.

It has been shown that matching respondents and interviewers on demographic characteristics can have a positive influence on survey participation (Webster 1996). But what is unknown at this point is whether matching interviewers to respondents based on common demographic characteristics can lead to a higher likelihood of consent. Alternatively, matching interviewers and respondents on demographic characteristics could backfire if they both share similarly sensitive views on data privacy. Interviewer experience may also influence whether a respondent provides consent. Prior interviewing experience has been shown to be beneficial for the recruitment of survey participants (Lipps and Pollien 2010), but whether it helps interviewers obtain consent from respondents is an open question. The expectation is that experienced interviewers perform better than their inexperienced counterparts in obtaining consent from respondents due to their greater familiarity with administering the consent request and addressing respondent concerns.

In the following sections of this paper, we address these hypotheses using data from the 2008 Health and Retirement Study (HRS) (<http://hrsonline.isr.umich.edu/>), a nationally-representative panel survey of older adults. This rich longitudinal data source allows us to test the aforementioned hypotheses using indicators derived from current- and prior-wave data collections, interviewer observations, and call record histories.

METHODS

Sample and Data Source

The HRS is a federally-funded longitudinal survey of adults over the age of 50 conducted by the Institute for Social Research (ISR) at the University of Michigan. The study began in 1992 with a cohort of then preretirement-aged individuals born between 1931 and 1941. New cohorts were added in 1993 and 1998 to round out the sample over age 50, and additional cohorts are enrolled every 6 years (e.g., in 2004, 2010) to refresh the sample at the younger ages. Response rates range from 70% to 82% in the baseline wave (depending on birth cohort and entry year), and from 87% to 89% at each follow-up wave.³ HRS conducts about 20,000 interviews every 2 years using a combination of telephone and in-person interviews. In 2006, HRS launched what is referred to as an enhanced face-to-face (E-FTF) interview, which includes a set of physical measures and biomarkers and a self-administered questionnaire on psychosocial topics. A random half of the sample was assigned to receive the E-FTF interview in 2006 and the other half in 2008. The plan is to repeat the E-FTF interview every other wave (or every 4 years) on the alternating half sample.

Consent Procedures

In accordance with an agreement with the Social Security Administration (SSA), HRS respondents are periodically asked to grant ISR permission to obtain respondents' earnings and benefit histories as reported to the SSA. Until 2004, the SS linkage consent was

³The response rates reported for the baseline wave are the product of the household screening response rate and the interview response rate among sample members who are identified as eligible in the household screener. The denominator used in calculating the response rates for the second and later waves includes only those who were themselves respondents at Wave 1, or whose spouse or partner was a respondent at Wave 1. In addition, individuals who have died or who have requested to be permanently removed from the sample are excluded from the denominators for the follow-up waves. More details of the response rate calculations can be found in HRS (2011).

retrospective, that is it covered linkage of past earnings histories and benefits up to the date on which the consent was granted. Starting in 2006, HRS began asking for consent for prospective linkage of SS records. HRS has had the best success when the SS linkage request is administered during an in-person interview (as opposed to a telephone interview). As a result, the SS linkage request was folded in to the E-FTF interview starting in 2008. In that wave, a random half of respondents (the 2008 E-FTF sample) were asked for prospective consent to release their SSA records. The consent form contains the following statements:

“We would like to obtain a history of your earnings and any benefits from programs administered by the Social Security Administration applied for or received through 2023. Since most people cannot recall this information very well, we are asking for your permission to obtain from government records the following: 1) Your earnings reported to Social Security. 2) Any information about benefits from programs administered by the Social Security Administration applied for or received through 2023.”

SSA requires separate signatures for earnings records and benefits records. Consenting respondents were asked to provide their social security number to facilitate the linkage, but this piece of information was not required to perform the linkage. The consent request was administered at the end of the interview. A small minority of respondents in the E-FTF sample (about 5%) declined the in-person interview but agreed to complete their interview by telephone. These individuals were still asked for SS linkage consent at the end of the interview. The permission form was mailed to them and they were asked to complete the form and send it back to the home office. Overall, about 68% of respondents consented to the SS linkage.

Indicators of Consent Mechanisms

Accuracy/Uncertainty—One hypothesis of consent to be examined in this study asserts that respondents who don’t know or are unable to report their administrative information with high accuracy are more likely to consent to the linkage request in order to facilitate the transaction of this information. The HRS asks respondents to report on several variables that can be found in the linked SSA earnings records, including social security income (“Do you currently receive any income from Social Security?”), supplemental security income (“Did you receive any income last month from Supplemental Security Income, also called SSI?”), and other sources of government income receipt, including welfare (“Did you receive any income from welfare in the last calendar year, not including SSI?”), veteran benefits or military pension (“Are you currently receiving any income from veteran benefits or a military pension?”), and food stamps (“Did you receive government food stamps at any time since month of previous interview?”). An additive index was created that is the sum of “don’t know” responses given to these five income items (range: 0–5). This *uncertainty* index is used as an indicator of respondents’ uncertainty about their earnings. The hypothesized expectation is that the index will be positively related to the likelihood of consent.

Privacy/Confidentiality Concerns—Another hypothesis of consent is that respondents are less likely to consent to data linkage if they have concerns regarding privacy and data confidentiality. We test this hypothesis using interviewer observations collected during the prior-wave HRS interview in 2006.⁴ In addition, we construct an indicator which is the number of item refusals given to the five financial questions used to construct the uncertainty index as described in the preceding paragraph. At the end of each interview, interviewers were asked to report on whether respondents ever inquired about the need to know the answers to some of the survey items, whether respondents expressed concerns about the confidentiality of their responses, and whether they believed respondents answered truthfully to the financial questions. The specific questions asked of interviewers were:

1. “During the interview, how often did the respondent ask you why you needed to know the answer to some questions? (never, seldom, or often)”
2. “During the interview, how often did the respondent express concern about whether his/her answers would be kept confidential? (never, seldom, or often)”
3. “How truthful do you believe the respondent was regarding his/her answers to financial questions? (completely truthful, mainly truthful, about half and half, mainly untruthful)”

We treat these items as indicators of respondents having concerns regarding privacy and data confidentiality.⁵ A *confidentiality concern* index (range: 0–3) was created by summing the number of times any indication of a confidentiality concern (denoted by the underlined responses) was given to these three items.⁶ Our expectation is that this index, as well as the number of item refusals to financial questions, will be negatively related to the likelihood of consent.

General Resistance towards the Survey Interview—This hypothesis states that some respondents express initial resistance towards the survey interview due to a variety of different reasons (e.g., lack of interest, burden, etc.) and may be less likely to fully cooperate during the interview and less likely to comply with in-survey requests, such as data linkage. The HRS collects several indicators of resistance and uncooperativeness based on call records and interviewer observations measured at the end of each interview. The call record data includes the number of call attempts needed to complete the interview, whether the respondent initially refused the survey request, whether the respondent failed to participate in any of the previous wave interviews, and whether the respondent refused the enhanced face-to-face interview and opted instead for a telephone interview. Interviewer observations of cooperation include an assessment of how often the respondent asked how much longer

⁴All interviewer observations used in this study were collected in the prior wave interview (2006), as opposed to the current wave interview (2008). This decision was based on the desire to avoid the potential contaminating effects of the current-wave consent outcome on interviewers’ assessments of respondents.

⁵A reviewer noted that items 1 and 3 could be motivated by other reasons besides privacy and confidentiality, including lack of knowledge or social desirability, and may have less face validity than item 2. We conducted a sensitivity analysis by removing items 1 and 3 from the index and re-running the analysis. The conclusions did not change and the three items were kept in the final analysis.

⁶The rationale for deciding which categories to use for all indicators was based on several factors, including examination of the distribution of the responses and whether different categorizations produced different results. Some of the response options, particularly at the extremes, did not yield enough cases to retain the original categorizations, and recoding was necessary. We tried different formulations of the categorizations and came to the same conclusions.

the interview would last, a rating of the respondent's cooperation during the interview, a rating of the respondent's resistance level, and a rating of the respondent's enjoyment with the interview. The exact wording of the relevant interviewer observations are as follows:

1. "During the interview, how often did the respondent ask how much longer the interview would last? (never, seldom, often)"
2. "How was respondent's cooperation during the interview? (excellent, good, fair, poor)"
3. "How would you describe the level of resistance from the respondent? (low/passive, moderate, high)"
4. "How much did the respondent seem to enjoy the interview? (a great deal, quite a bit, some, a little, not at all)"

We use these 4 items, collected in the prior-wave interview, to create an additive *uncooperativeness* index (range: 0–4). A value of 1 was assigned to each item if the interviewer observation yielded an indication of uncooperative behavior (denoted by the underlined responses), and 0 otherwise.⁷ This index was constructed using prior-wave interview observations (2006) instead of the current-wave (2008) observations to avoid the interviewer impressions being influenced by whether the respondent complied with the consent request. The hypothesized expectation is that the uncooperativeness index will be negatively related to the likelihood of consent.

Preexisting Relationship with the Administrative Data Agency—Another mechanism of consent under consideration relates to respondents' relationship with the agency responsible for maintaining the administrative data being requested. In particular, respondents who have a preexisting relationship with the administrative data agency may be more likely to comply with the linkage request if they receive services or benefits from the agency. Conversely, receiving services from the agency may actually decrease the likelihood of consent if the respondent's relationship with the agency has been negative, or if respondents fear that the linkage would lead to examination of whether they are claiming benefits or services fraudulently. Two important services that the Social Security Administration provides are the distribution of social security and supplemental security income. HRS asks respondents to report whether they are currently receiving these two income sources. These two variables are combined into a single binary indicator of whether benefits are being received from SSA. We also create a binary indicator of whether a respondent currently receives any other government benefits, including welfare, veteran benefits or military pension, and/or food stamps. Our hypothesis is two-sided. If receiving services or benefits from SSA (or another government agency) incentivizes respondents to consent to the linkage, then we would expect the likelihood of consent to be higher for those receiving such benefits. Alternatively, if respondents who receive these benefits are

⁷The bivariate correlations of consent and the resistance items ranged from -0.11 to -0.12 with the exception of the enjoyment item, which yielded a correlation of -0.08 . Since the enjoyment item was the weakest of the four, we removed it from the index and re-ran the analysis. Based on this sensitivity analysis, there was no evidence that including the enjoyment item attenuated the relationship between uncooperation and consent. Therefore the enjoyment item was kept in the index.

concerned about the linkage for the reasons discussed above, then we would expect these respondents to have a lower likelihood of consent.

Acquiescence/Attentiveness—We test the notion that inattentive or acquiescent respondents are either more or less likely to consent based on their perception of what the easier task might be. Without any direct measures of respondents' acquiescent or inattentive behavior during the survey interview or at the time of the linkage request, we utilize a proxy indicator of respondents' attentiveness from the prior wave interview. At the conclusion of the prior wave interview, interviewers were asked to rate respondents' level of attentiveness during the interview ("How attentive was the respondents to the questions during the interview? (not at all attentive, somewhat attentive, very attentive).") Respondents who were distracted during the interview by such things as television or radio, children or other household members, or a cell phone or PDA are likely to have been rated as somewhat or not at all attentive, depending on the frequency or duration of the distractions. Alternatively, if the interviewer felt that the respondent wasn't listening carefully to the questions then this may lead to a poor rating of attentiveness. We created a binary indicator indicating whether respondents were not "very attentive" during the interview (denoted by the underlined response options). The hypothesis for this indicator is also two-sided, depending on whether respondents consider consent or non-consent to be the simpler task. Thus, we hypothesize that the least attentive respondents are either more likely or less likely to consent to the linkage request compared to the most attentive respondents.

Interviewer Effects—There are several interviewer-related hypotheses that we attempt to address in this study. The first hypothesis is that interviewers who possess critical views towards data privacy are less likely to obtain consent from respondents. We do not have direct measurements of interviewers' attitudes toward data privacy. Instead, we use interviewer education as a weak proxy. We assert that interviewers with the lowest levels of education possess the most negative attitudes toward data privacy. Our rationale is based on studies that find that persons with the lowest levels of education are more likely to express concerns about privacy and data confidentiality in survey research studies (Westin 1990; Singer et al. 1993). In addition to main effects, we test the interaction between interviewers' and respondents' level of education to assess whether matching respondents with interviewers who share the same education level (a weak proxy for shared attitudes on data privacy) produce lower or higher rates of linkage consent relative to other interviewer-respondent education combinations.

The second interviewer-related hypothesis is that matching interviewers and respondents on demographic variables increases the likelihood of consent as it seems to increase the likelihood of survey participation (Webster 1996). We examine the effects of matching based on age, gender, and race/ethnicity variables collected on both HRS respondents and interviewers. Our expectation is that matching on any one of these variables will increase the likelihood of consent to a greater extent than that of not matching. Finally, the third interviewer-related hypothesis we consider claims that experienced interviewers obtain consent at a higher rate than inexperienced interviewers. We operationally define interviewer experience to be dichotomous, where "inexperienced" interviewers are those who have never

worked on the HRS prior to the current wave and “experienced” interviewers are those who have worked on at least one HRS wave prior to the current one.⁸ In addition, two running count variables indicating the number of consent requests administered by the interviewer in all interviews prior to the current one, as well as the number of consents obtained in those interviews, are used as dynamic measures of interviewer experience accrued during the field period.⁹ We expect that the number of consent requests administered by the interviewer and number of consents that they obtain prior to the current interview will have a positive effect on obtaining linkage consent.

Outcome Measure

The primary outcome measure is a dichotomous indicator of whether HRS respondents consented to the SSA data linkage request. Consent is operationally defined as returning the signed consent form to the HRS interviewer or to ISR. Consenting respondents were asked to provide a social security number to facilitate the linkage, but this was not required. In the preliminary analysis, two different definitions of consent were considered: 1) consent with social security number (conditional) and; 2) consent with or without social security number (unconditional).¹⁰ Both outcomes yielded similar results with the unconditional consent definition producing slightly stronger effects than the conditional definition. We use the unconditional definition of consent for all analyses presented below.

Statistical Analyses

A multilevel random effects logistic regression model was used to quantify respondents’ likelihood of consenting to the data linkage request conditional on respondent- and interviewer-level covariates. The logistic regression was performed using the NLMIXED procedure in SAS 9.2 (Littell et al. 1996). HRS utilizes complex sample survey design features which were incorporated into the statistical analysis. The sample consists of 112 primary sampling units paired within 56 sampling strata. Sampling weights were used to adjust for differential probabilities of selection, nonresponse, and sample noncoverage. The weights were incorporated into the NLMIXED procedure to obtain valid point estimates. Standard errors were computed using a jackknife variance procedure (with 56 replications) that account for the effects of stratification and clustering (Berglund 2002). A total of 6,613 randomly selected respondents belonging to the target population of noninstitutionalized adults over the age of 50 received the consent request. Interviewer observations from the prior wave interview were missing for 229 respondents and these cases were excluded from the analysis. This yielded a final analytical sample size of 6,384. No significant differences were found between cases that were included or excluded based on socio-demographic variables.

⁸About 31% (or n=51) of HRS interviewers were labeled as “inexperienced” and 69% (or n=113) were labeled as “experienced”.

⁹The median number of consent requests administered by the interviewers was 41 with an interquartile range of 35. The median number of consents obtained by interviewers was 23.5 with an interquartile range of 27.5.

¹⁰The conditional and unconditional consent rates (weighted) were 49.1% and 67.8%, respectively.

RESULTS

Sample Characteristics

The overall consent rate was 67.8% (weighted and unweighted). Table 2 shows weighted estimates of the target population of adults over the age of 50 who were eligible for the data linkage consent request. Overall, males were more likely to provide consent than females (69.1% vs. 66.9%), white respondents were more likely to give consent than blacks and other racial groups (69.1% vs. 60.4% vs. 65.8%), and married respondents provided consent at a higher rate than separated/divorced, widowed, and never married ones (69.5% vs. 64.9% vs. 65.3% vs. 65.9%). There were no statistically significant consent rate differences by birth cohort, Hispanic ethnicity, and education.

Model Building and Evaluation

A weighted logistic regression model for consent was fit by incorporating covariates related to respondent background characteristics and indicators of the hypothesized mechanisms. To assess the relative impact of each mechanism on consent the set of indicators associated with each mechanism were removed from the full model one at a time. Then the reduced model was fit and contrasted against the full model. Three diagnostic measures were used to assess the impact of each mechanism. The first diagnostic measure is the likelihood ratio (LR) statistic, which is computed as,

$$LR = -2 \log_e \left(\frac{\text{likelihood for reduced model}}{\text{likelihood for full model}} \right)$$

The second diagnostic measure is the relative Akaike's information criterion (AIC) computed as,

$$\text{Rel AIC} = 100 \cdot \frac{AIC_{\text{reduced}} - AIC_{\text{full}}}{AIC_{\text{full}}}$$

Finally, the third diagnostic measure is the relative Schwarz criterion (SC), which is computed as,

$$\text{Rel SC} = 100 \cdot \frac{SC_{\text{reduced}} - SC_{\text{full}}}{SC_{\text{full}}}$$

All of these measures produce values that are greater than or equal to zero. The diagnostic values increase as a function of the relative importance of the mechanism (and its corresponding set of indicators) in terms of improving model fit. Table 3 presents the diagnostic results. In general, the Interview Resistance indicators had the biggest impact on model fit based on the diagnostic measures (LR, 283.62, $p < 0.01$; Rel AIC, 3.63; Rel SC, 2.73), followed by the Privacy/Confidentiality indicator (LR, 43.48, $p < 0.01$; Rel AIC, 0.56; Rel SC, 0.44). All other mechanisms had relatively minimal impact on model fit.

Likelihood of Consent Based on Indicators of Hypothesized Mechanisms

Table 4 shows weighted odds ratios of the likelihood of linkage consent. After controlling for respondent background characteristics and the indicators of each hypothesized mechanism, education was the only respondent demographic characteristic that yielded a statistically significant relationship with consent: college graduates tended to be more likely to consent relative to persons who did not complete high school (16+ years: OR, 1.32; 95% CI, 1.02–1.70). A marginally significant result ($P < 0.10$) was observed for black respondents who were less likely to consent relative to white respondents (OR, 0.80; 95% CI, 0.63–1.01). A respondent's foreign citizenship status, self-reported health rating, and numeracy skill (based on a numeracy assessment module) were unrelated to the likelihood of consent.

Substantive financial variables had a mixed relationship with consent. Household income¹¹ was found to be unrelated to consent, whilst net worth¹² was significantly related to consent. As net worth increased, the likelihood of consent decreased (2nd quartile: OR, 0.75; 95% CI, 0.62–0.92; 3rd quartile: OR, 0.63; 95% CI, 0.48–0.82; 4th quartile: OR, 0.54; 95% CI, 0.38–0.76).

Effect of Financial Uncertainty on the Likelihood of Consent to Social Security Record Linkage

We did not find any support for the *uncertainty* hypothesis of consent. The number of “don't know” responses given to financial questions in either the prior wave (2006) or the current wave (2008) did not appear to provoke respondents to consent to the SSA linkage request in the current wave (2006: OR, 1.21; 95% CI, 0.52–2.83; 2008: OR, 0.83; 95% CI, 0.45–1.52). We also examined the percentage of “don't know” responses to financial questions across consent propensity strata. Based on the final consent model, consent propensity scores were generated for all cases. Following suggestions from Rosenbaum and Rubin (1983), five equal-sized propensity score groups were formed, ordered from low-to-high propensity of consent. Respondents were assigned the value 1 in the new formed variable if their propensity to consent lay between 0.0006 and 0.60, the value 2 if their estimated propensity fell within the interval of 0.60 and 0.69, the value 3 for propensities between 0.69 and 0.74, 4 if the estimated propensity to consent ranged from 0.74 and 0.78, and 5 for the remaining respondents indicating the highest consent propensity range of 0.78 and 0.91. Table 5 shows the percentage of “don't know” responses across the propensity stratum. Contrary to expectations, a roughly negative trend is observed for each year: the largest percentages of “don't know” responses to financial questions occur among the lowest consent propensity groups and the smallest percentages occur among the highest propensity groups (2006: trend test, $P = 0.333$; 2008: trend test, $P = 0.083$). Overall, we conclude that there is no evidence supporting the hypothesis that persons who are less familiar with their administrative information are more likely to give linkage consent.

¹¹Household income reflects total income for the last calendar year. It is set to the sum of respondent and spouse earnings, pensions and annuities, Supplemental Security Income and Social Security Disability, Social Security retirement, unemployment and workers compensation, and other government transfers, and household capital income and other income.

¹²Net worth is a generated value based on total wealth and is calculated as the sum of all assets less the sum of all debt.

Effect of Privacy/Confidentiality Concerns on the Likelihood of Consent

In contrast to the uncertainty hypothesis, we find strong support for the privacy/confidentiality hypothesis. The confidentiality concern index is negatively related to the likelihood of consent (OR, 0.62; 95% CI, 0.51–0.74). That is, as the number of confidentiality-related concerns expressed by the respondent during the 2006 wave interview increases, the likelihood of consent in 2008 decreases. A second indicator of privacy concerns is the number of times a respondent refused to answer relevant financial questions in the survey. This indicator, measured in both 2006 and 2008, is significantly related to consent in the current wave (2006: OR, 0.88; 95% CI, 0.79–0.97; 2008: OR, 0.68; 95% CI, 0.47–0.99); as the number of item refusals increase, the likelihood of consent decreases. In addition, we examined the percentage of item refusals across consent propensity strata. Table 5 shows a negative trend indicating that the largest percentages of item refusals occur among the lowest consent propensity groups and the smallest percentages among the highest propensity groups (2006: trend test, $P = 0.017$; 2008: trend test, $P = 0.083$). All indications suggest that there is strong support for the hypothesis that privacy and confidentiality concerns are negatively related to the likelihood of consent.

Effect of Interview Resistance on the Likelihood of Consent

We find support for the notion that respondents are more likely to refuse the consent request if they exhibit initial resistance towards the survey interview. The *uncooperativeness* index – constructed from interviewer assessments of respondents' cooperativeness in the prior-wave interview – is significantly related to the likelihood of consent (OR, 0.76; 95% CI, 0.70–0.82). More specifically, interviewer observations of uncooperative behavior in the prior-wave interview tend to be negatively correlated with the likelihood of consent in the current wave. Other indicators of interview resistance were negatively related to the likelihood of consent, including initial refusal to be interviewed in the current wave (OR, 0.57; 95% CI, 0.38–0.85) and at least one noninterview in a previous wave of HRS data collection (OR, 0.64; 95% CI, 0.50–0.82). Respondents who refused the enhanced face-to-face interview in favor of a telephone interview were also less likely to consent to data linkage (OR, 0.34; 95% CI, 0.23–0.48). Overall, we find strong support for the notion that interview resistance is negatively related to the likelihood of consent.

Effect of Preexisting Relationship with the Administrative Data Agency on the Likelihood of Consent

We find mixed support for the preexisting relationship hypothesis. Respondents receiving social security or supplemental security income were neither more nor less likely to consent to the SSA data linkage request compared to persons not receiving income from the Social Security Administration (OR, 1.05; 95% CI, 0.87–1.28). However, respondents receiving other types of government income (including welfare, veteran benefits, food stamps) were more likely to consent to SSA data linkage (OR, 1.34; 95% CI, 1.10–1.65). Thus, there is mixed support for the hypothesis that respondents receiving services from a government agency are more likely to consent to a government data linkage request.

Effect of Respondent Attentiveness on the Likelihood of Consent

There is some evidence that the level of attentiveness -- a proxy indicator for acquiescence -- of respondents during the survey interview influences their likelihood of consent. Specifically, we find that respondents who were rated by the interviewer as being not “very attentive” in the prior-wave interview were more likely to consent to the linkage in the current wave (OR, 1.22; 95% CI, 0.96–1.56). Although marginally significant ($P < 0.10$), this suggests that inattentive respondents may acquiesce during the linkage request and give consent because that is perceived to be the easier task relative to not giving consent.

Interviewer Effects

We find little support for the interviewer-related hypotheses. Interviewer education – a weak proxy for attitudes towards data confidentiality and privacy – was not related to the consent outcome. The interaction between respondent and interviewer education levels (not shown) was also not statistically significant. In general, we find no evidence that matching respondents and interviewers on demographic characteristics leads to a higher likelihood of consent based on gender, race/ethnicity, age, and education. Furthermore, inexperienced HRS interviewers were not more or less successful at obtaining consent relative to experienced interviewers. The number of consent requests administered by the same interviewer in previous interviews had no effect on the likelihood of consent in the current interview, but the number of previous consents obtained by an interviewer had a marginally significant ($P < 0.10$) and positive effect on the likelihood of consent (OR, 1.03; 95% CI, 1.00–1.06). This suggests that interviewers tend to benefit from early success when obtaining consent in subsequent interviews over the course of the field period. We also find a marginally significant interviewer variance component (σ^2_{int} , 1.52; 95% CI, 0.97–2.38), which suggests that there are other interviewer-level factors that contribute to interviewers’ likelihood of obtaining consent from respondents, but were not accounted for in this study.

DISCUSSION

Sample surveys are increasingly being linked to administrative data sources to enhance the quality of the survey data and increase research opportunities for data users. However, obtaining linkage consent from respondents remains a persistent challenge that threatens the validity of inferences obtained from linked data sources. Little is known about the underlying mechanisms driving the consent decision. We investigated this issue by addressing several hypotheses of linkage consent in the Health and Retirement Study, a large nationally-representative panel survey of older adults linked to Social Security Administration records.

We did not find any support for the hypothesis that respondents are more likely to consent to the linkage if they are unable to self-report items contained in their administrative record. That is, it does not appear that respondents compensate for their uncertainty by increasing their likelihood of consent. The fact that the consent request was not framed this way may explain the lack of support for this hypothesis. That is, respondents who were unable to answer certain administrative items were not encouraged to overcome their uncertainty by consenting to the data linkage. Indeed, it may matter whether this connection is brought to

the respondent's attention or not during the consent request. Experimental research on the framing of the consent request is needed to further explore this possibility.

An alternative post-hoc hypothesis is that respondents choose not to release their records to survey organizations because they are uncertain of what the record contains. We find some evidence supporting this hypothesis: the highest frequency of "don't know" responses to administrative financial items occurred among respondents who had the lowest likelihood of consent. On the other hand, a "don't know" response may in fact be a passive refusal motivated not by uncertainty but rather unwillingness to divulge confidential information to the survey organization. If so, this is likely to decrease the likelihood of consent, as is evidenced by our finding that explicit refusal to answer financial items led to a lower likelihood of consent.

We find strong support for the privacy and interview resistance hypotheses based on prior-wave interviewer observations of respondent behavior and item refusals to income questions in the prior- and current-wave interviews. The fact that privacy is related to the consent outcome is not surprising and is consistent with the literature (Bates 2005). Regarding interview resistance, the evidence suggests that uncooperative behavior exhibited in the prior-wave interview tends to be associated with a lower likelihood of consent. These findings have important practical implications for ongoing surveys planning future data linkages. Indications of interview resistance or confidentiality concerns observed prior to the linkage request may be used to preidentify (and possibly intervene on) persons who are less likely to consent. For cross-sectional surveys, respondents' refusal to answer income questions could be used for similar preidentification and intervention strategies for subsequent linkage requests.

A respondent's preexisting relationship with the administrative data agency denoted by whether they currently receive income from the Social Security Administration – was not related to consent. That is, we find no evidence that social security income recipients are overrepresented in the consent group. However, respondents who received other types of government income tended to consent at a higher rate than those who were not receiving government income. This finding is somewhat consistent with the medical literature, which finds that patients with the characteristic of interest tend to consent to medical record review and follow-up interview at a higher rate than those who don't have the characteristic of interest (Dunn et al. 2004; Woolf et al. 2000; Petty et al. 2001). This finding raises potential overrepresentation issues that could introduce bias. Accounting for the selectivity of consent through weighting adjustments or other statistical procedures may need to be considered for data linkage studies. HRS is currently in the process of developing weighting adjustments for non-consent.

We found that inattentive respondents tend to provide linkage consent at a higher rate than attentive respondents. This acquiescent behavior begs the question of whether respondents take the time to consider all of the information about the linkage request that is provided to them by the survey organization. Linkage requests often contain technical and relatively complex administrative details that may be too burdensome for some respondents to pore over. Rather than expend the necessary effort needed to read through and fully understand

the request, some respondents may simply give consent because that is the easier task. Future research on respondents' comprehension of the linkage request is needed to determine whether current methods of administering the consent request are effective, or if alternative methods should be considered.

Although interviewers play a crucial role in administering the consent request and addressing respondent concerns, we found little evidence of systematic interviewer effects. Specifically, we found no relationship between interviewers' demographic characteristics and respondents' likelihood of consent. Matching respondents and interviewers based on demographic characteristics did not increase the likelihood of consent. This result conflicts with the survey participation literature, which finds that "matching" interviewers with respondents has a positive effect on participation (Webster 1996). We also found no strong effect of interviewer experience on the likelihood of consent. However, the number of consents obtained by an interviewer in previous interviews tended to have a positive effect on obtaining consent in the current interview, which suggests that interviewers who have success obtaining consent from respondents early in the data collection period also tend to be successful later on.

The results also indicated unexplained consent variation attributable to the interviewer. Indeed, a wide range of consent rates were observed between interviewers: the 25th, 50th, and 75th quartiles for interviewer-level consent rates were 35.4%, 56.5%, and 69.1%, respectively. It may be of interest to survey organizations to monitor these rates and intervene when interviewers are performing below expectation. In addition, it is worthwhile to explore other interviewer-level factors that may influence a respondent's likelihood of consent. One potential factor, interviewers' attitudes toward data privacy, may play a role in how effective interviewers are in administering the consent request and addressing respondents' concerns. Findings based on interviewer assessments at the U.S. Census Bureau indicate that some interviewers share the same concerns about data privacy and confidentiality as their respondents, and are even skeptical of the promise of confidentiality that they themselves give to respondents during survey recruitment (Mayer, 2001). More direct measures of interviewer attitudes are needed to explore how their attitudes interact with those of respondents and influence their likelihood of linkage consent. A key challenge for future work will be to identify interviewer-level factors that can be leveraged during interviewer recruitment and training in ways that could improve respondents' likelihood of consent and minimize selection bias.

A major strength of the present study is the use of rigorously collected population-based data collected from respondents and interviewers over multiple waves of data collection. In addition, our study benefits from the joint modeling of respondent- and interviewer-level characteristics of consent which is rare in studies of consent. There are several limitations of this study that should be discussed. The observational nature of the data sources is informative for hypothesis generation and exploration, but is limiting for making definitive statements about the factors affecting consent. Although our models were extensively built to control for possible confounders, it is possible that extraneous factors were not accounted for. Another potential limitation is due to the fact that we studied a sample of panel respondents who comprise the older adult population. Panel respondents may be viewed as

being more cooperative than a cross-sectional sample of respondents and this may reduce the generalizability of our findings. Furthermore, the consent mechanisms of older age groups may be quite different than for younger age groups. We suspect that older people are more civic minded, yield higher response rates, and may be less concerned about privacy than younger adults, in which case these data may underestimate the consent problem. On the other hand, these panel respondents have demonstrated a vested interest in the survey (some respondents have participated in the HRS for almost 2 decades) and yet almost 35% of them refused the linkage request. Their persistent cooperation in the HRS yet non-trivial resistance towards linkage requests comprise an intriguing case study for studying the underlying mechanisms of consent.

REFERENCES

- Al-Shahi R, Vousden C, Warlow C. Bias from Requiring Explicit Consent from All Participants in Observational Research: Prospective, Population Based Study. *BMJ*. 2005; 331(7522):942. [PubMed: 16223793]
- Baker R, Shiels C, Stevenson K. What Proportion of Patients Refuse Consent to Data Collection from their Records for Research Purposes? *British Journal of General Practice*. 2000; 50:655–656. [PubMed: 11042920]
- Banks, J.; Lessof, C.; Taylor, R. Linking Survey and Administrative Data in the English Longitudinal Study of Aging. Presented at the Meeting on Linking Survey and Administrative Data and Statistical Disclosure Control; Royal Statistical Society; London. 2005 May.
- Bates, N. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2005. Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records; p. 3786-3793.
- Bates, N.; Pascale, J. Report of the Demographic Surveys Division, Statistical Research Division. Washington, D.C: U.S. Census Bureau; 2006. Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records.
- Berglund PA. Analysis of Complex Sample Survey Data Using SURVEYMEANS and SURVEYREG Procedures and Macro Coding. In: Proceedings of the 27th Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc. 2002; 263:1–5.
- Blumberg SJ, Luke JV. Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults. *Public Opinion Quarterly*. 2007; 71(5):1–16.
- Bryant H, Robson PJ, Ullman R, Friedenreich C, Dawe U. Population-Based Cohort Development in Alberta, Canada: A Feasibility Study. *Chronic Diseases in Canada*. *Chronic Diseases in Canada*. 2006; 27(2):51–59. [PubMed: 16867239]
- Buckley B, Murphy AW, Byrne M, Glynn L. Selection Bias Resulting from the Requirement for Prior Consent in Observational Research: A Community Cohort of People with Ischaemic Heart Disease. *Heart*. 2007; 93(9):116–1120.
- Calderwood, L.; Lessof, C. Enhancing Longitudinal Surveys by Linking to Administrative Data. In: Lynn, P., editor. *Methodology of Longitudinal Surveys*. New York: Wiley; 2009.
- Christoph, B.; Mueller, G.; Gebhardt, D.; Wenzig, C.; Trappmann, M.; Achatz, J.; Tisch, A.; Gayer, C. Institute for Employment Research. Nuremberg, Germany: 2006. Codebook and Documentation of the Panel Study 'Labour Market and Social Security (PASS): Vol. 1: Introduction and Overview, Wave 1 (2006/2007). No. 200805_en, FDZ Datenreport.
- Cleary PD, Jette AM. The Validity of Self-Reported Physician Utilization Measures. *Medical Care*. 1984; 22:796–803. [PubMed: 6492908]
- Curtin R, Presser S, Singer E. Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*. 2005; 69:87–98.
- Dahlhamer, JM.; Cox, CS. Proceedings of the Federal Committee on Statistical Methodology Research Conference. Washington, D.C.: 2007 Nov. Respondent Consent to Link Survey Data with

Administrative Records: Results from a Split-Ballot Field Test with the 2007 National Health Interview Survey.

- De Leeuw, E.; de Heer, W. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Groves, RM.; Dillman, DA.; Eltinge, JL.; Little, RJA., editors. *survey Nonresponse*. New York: John Wiley & Sons; 2002. p. 41-55.
- Dunn KM, Jordan K, Lacey RJ, Shapley M, Jinks C. Patterns of Consent in Epidemiologic Research: Evidence from Over 25,000 Responders. *American Journal of Epidemiology*. 2003; 159(11):1087–1094. [PubMed: 15155293]
- Edwards WS, Winn DM, Collins JG. Evaluation of 2-Week Doctor Visit Reporting in the National Health Interview Survey. *Vital Health Statistics*. 1996; 2(122):1–46. [PubMed: 8756998]
- Finkelstein MM. Do Factors Other than Need Determine Utilization of Physician's Services in Ontario? *Canadian Medical Association Journal*. 2001; 165:565–570. [PubMed: 11563208]
- Fricker, S. Unpublished Doctoral Dissertation. College Park: University of Maryland; 2007. The Relationship between Response Propensity and Data Quality in the Current Population Survey and the American Time Use Survey.
- Gage NL, Leavitt GS, Stone GC. The Psychological Meaning of Acquiescence Set for Authoritarianism. *Journal of Abnormal and Social Psychology*. 1957; 55:98–103.
- General Accounting Office (GAO). Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information. Report of the General Accounting Office, GAO-01-126SP. 2001
- Groves, RM.; Couper, MP. *Nonresponse in Household Interview Surveys*. New York: Wiley; 1998.
- Haider, S.; Solon, S. Non Random Selection in the HRS Social Security Earnings Sample. Santa Monica, Calif.: RAND Corporation; 2009. DRU-2254-NIA. <<http://www.rand.org/pubs/drafts/DRU2254/>>
- Hanley C. The 'Difficulty' of a Personality Inventory Item. *Educational and Psychological Measurement*. 1962; 22:577–584.
- Harris T, Cook DG, Victor C, Beighton C, Dewilde S, Carey I. Linking Questionnaires to Primary Care Records: Factors Affecting Consent in Older People. *Journal of Epidemiology and Community Health*. 2005; 59(4):336–338. [PubMed: 15767389]
- Health and Retirement Study (HRS). Technical Report. Institute for Social Research, University of Michigan; 2011. Sample Sizes and Response Rates. <<http://hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf>>
- Huang N, Shih SF, Chang HY, Chou YJ. Record Linkage Research and Informed Consent: Who Consents? *BMC Health Services Research*. 2007; 7:18. [PubMed: 17291357]
- Jäckle A, Lynn P, Sinabaldi J, Tipping S. The Effect of Interviewer Personality, Skills, and Attitudes on Respondent Cooperation with Face-to-Face Surveys. *ISER Working Paper Series*, No. 2011-14. 2011 Jun.
- Jay, MG.; Belli, RF.; Lepkowski, JM. Proceedings of the Joint Statistical Meetings. Toronto, Ontario: American Statistical Association; 1994 Aug. Quality of Last Doctor Visit Reports: A Comparison of Medical Record and Survey Data.
- John, OP.; Srivastava, S. The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In: John, OP.; Pervin, LA., editors. *Handbook of Personality: Theory and Research*. New York: Guilford Press; 1999.
- Jenkins SP, Cappellari L, Lynn P, Jaekle A, Sala E. Patterns of Consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society (Series A)*. 2006; 169(4):701–722.
- Kennickell AB. Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances. *Journal of Official Statistics*. 1999; 15:283–303.
- Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written Informed Consent and Selection Bias in Observational Studies using Medical Records: A Systematic Review. *British Medical Journal*. 2009; 338:b866. [PubMed: 19282440]
- Klassen A, Lee S, Barer M, Raina P. Linking Survey Data with Administrative Health Information: Characteristics Associated with Consent from a Neonatal Intensive Care Unit Follow-Up Study. *Canadian Journal of Public Health*. 2005; 96:151–154. [PubMed: 15850038]

- Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Helenius H, Koskenvuo M. Non-Response and Related Factors in a Nation-Wide Health Survey. *European Journal of Epidemiology*. 2001; 17:991–999. [PubMed: 12380710]
- Krosnick JA. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*. 1991; 5(3):213–236.
- Krosnick JA. Survey Research. *Annual Review of Psychology*. 1999; 50(3):537–567.
- Lehtonen, R. Interviewer Attitudes and Unit Nonresponse in Two Different Interviewing Schemes. In: Laaksonen, S., editor. *Sixth International Workshop on Household Survey Nonresponse*. Helsinki: Statistics Finland; 1996.
- Lillard LA, Farmer MM. Linking Medicare and National Survey Data. *Annals of Internal Medicine*. 1997; 127:691–695. [PubMed: 9382381]
- Lipps O, Pollien A. Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey. *Field Methods*. 2010
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD. *SAS System for Mixed Models*. Cary, NC: SAS Institute; 1996.
- Mayer TS. Interviewer Attitudes about Privacy and Confidentiality. U.S. Census Bureau, Research Report Series, Survey Methodology, 2001-02. 2001
- McCarthy DB, Shatin D, Drinkard CR, Kleinman JH, Gardner JS. Medical Records and Privacy: Empirical Effects of Legislation. *Health Services Research*. 1999; 34(1, pt 2):417–425. [PubMed: 10199685]
- Murdoch M, Pietila DM, Partin MR. Randomized Trial Showed That an ‘Embedded’ Survey Strategy Optimized Authorization Rates Compared with Two After Survey Strategies in Veterans with PTSD. *Journal of Clinical Epidemiology*. 2010; 63:665–671. [PubMed: 20304608]
- Nagelkerke NJ. A Note on a General Definition of the Coefficient of Determination. *Biometrika*. 1991; 78:691–692.
- Olson JA. Linkages with Data from Social Security Administration Records in the Health and Retirement Study. *Social Security Bulletin*. 1999; 62(2):73–85.
- Olson KM. Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly*. 2006; 70(5):737–758.
- O’Muircheartaigh C, Campanelli P. A Multilevel Exploration of the Role of Interviewers in Survey Non-response. *Journal of the Royal Statistical Society (Series A)*. 1999; 162(3):437–446.
- Petty DR, Zermansky AG, Raynor DK, Lowe CJ, Vail A, Freemantle N. Clinical Medication Review by a Pharmacist of Patients on Repeat Prescriptions in General Practice. *International Journal of Pharmacy Practice*. 2001; 9(S1):47.
- Rhodes AE, Fung K. Self-Reported Use of Mental Health Services Versus Administrative Records: Care to Recall? *International Journal of Methods in Psychiatric Research*. 2004; 13(3):165–175. [PubMed: 15297900]
- Rosenbaum PR, Rubin DR. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983; 70(1):41–55.
- Sakshaug JW, Couper MP, Ofstedal MB. Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults. *Medical Care*. 2010; 48(1):64–71. [PubMed: 20050351]
- Sala, E.; Burton, J.; Knies, G. Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview and Interviewer Characteristics. No. 2010-28. ISER Working Papers, Institute for Social and Economic Research. 2010. <http://econpapers.repec.org/RePEc:ese:iserwp:2010-28>
- Shah S, Harris TJ, Rink E. Do Income Questions and Seeking Consent to Link Medical Records Reduce Survey Response Rates? A Randomised Controlled Trial among Older People. *British Journal of General Practice*. 2001; 51:223–225. [PubMed: 11255905]
- Silva SM, Smith WT, Bammer G. The Effect of Timing When Seeking Permission to Access Personal Health Services Utilization Records. *Annals of Epidemiology*. 2002; 12:326–330. [PubMed: 12062920]
- Singer E, Mathiowetz NA, Couper MP. The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 Census. *Public Opinion Quarterly*. 1993; 57:465–482.

- Tate RA, Calderwood L, Dezateux C, Joshi H. Mother's Consent to Linkage of Survey Data with her Child's Birth Registration Records in a Multi-Ethnic National Cohort Study. *International Journal of Epidemiology*. 2006; 35(2):294–298. [PubMed: 16373376]
- Trott DM, Jackson DN. An Experimental Analysis of Acquiescence. *Journal of Experimental Research in Personality*. 1967; 2:278–288.
- Webster C. Hispanic and Anglo Interviewer and Respondent Ethnicity and Gender: The Impact on Survey Response Quality. *Journal of Marketing Research*. 1996; 33(1):62–72.
- Westin, AF. The Equifax Report on Consumers in the Information Age. Atlanta: Equifax; 1990.
- Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research. *Archives of Family Medicine*. 2000; 9:1111–1118. [PubMed: 11115216]
- Yawn BP, Yawn RA, Geier GR, Xia Z, Jacobsen SJ. The Impact of Requiring Patient Authorization for Use of Data in Medical Records Research. *Journal of Family Practice*. 1998; 47(5):361–365. [PubMed: 9834771]
- Young AF, Dobson AJ, Byles JE. Health Services Research Using Linked Records: Who Consents and What is the Gain? *Australian and New Zealand Journal of Public Health*. 2001; 25(5):417–420. [PubMed: 11688619]
- Zell Ezzati-Rice ERTM, Battaglia MP, Wright RA. National Immunization Survey: The Methodology of a Vaccination Surveillance System. *Public Health Reports*. 2000; 115:65–77. [PubMed: 10968587]

Table 1

List of studies linking survey records and administrative data sources.

Citation	Target population	Country	Type of administrative data	Response rate (%)	Consent rate (%)
Baker et al. 2000	Patients	UK	Medical records	67.6	90.2
Banks et al. 2005	Households	UK	(1) Health records; (2) Benefit/tax records	68.3	(1) 81.5 (2) 77.8
Bates and Pascale 2006	Household	USA	Government records	42.4–59.2	63.4 (no ssn)
Christoph et al. 2008	Households	Germany	Employment records	30.5	79.8
Cleary and Jette 1984	Households	USA	Medical/pharmacy records	88.0	89.0
Dahlhamer and Cox 2007	Household	USA	Health records	68.0 (adult) 77.0 (child)	66.2 (adult) 60.7 (child)
Edwards et al. 1996	Household	USA	HMO records	76.0	94.0
Finkelstein 2001	Households	Canada	Health insurance records	88.0	89.0
Haider and Solon 2009	Households	USA	Social security records	81.7	74.9 (matched)
Harris et al. 2005	Patients	UK	Medical records	75.0	92.0
Huang et al. 2007	Households	Taiwan	Health insurance records	91.4	88.0
Jay et al. 1994	HMO	USA	Medical records	66.6	95.0
Jenkins et al. 2006	Households	UK	(1) Tax credit and benefits (2) Employer records	89.0	(1) 77.4 (2) 42.0
Klassen et al. 2005	Children/caregivers	Canada	Medical records (adult & child)	55.0	71.6 (child) 67.0 (caregiver)
Korkela et al. 2001	Household	Finland	Medical registries	40.0	91.8
McCarthy et al. 1999	HMO	USA	Medical records	52.0	19.0
Murdoch et al. 2010	Veterans	USA	Medical records	(1) 67.5 (2) 71.9 (3) 74.2	(1) 59.1 (2) 52.3 (3) 48.7
Olson 1999	Households	USA	Social Security records	81.6	75.0 (matched)
Rhoades and Fung 2004	Households	Canada	Mental healthcare records	74.4	96.5
Sala et al. 2010	Household	UK	(1) Health records (2) Education records (3) Financial records	87.0	(1) 41.0 (2) 39.0 (3) 32.0

Citation	Target population	Country	Type of administrative data	Response rate (%)	Consent rate (%)
Shah et al. 2001	Patients	UK	Medical records	62.8	86.9
Silva et al. 2002	Employees	Australia	Healthcare utilization records	53.0	60.0
Tate et al. 2006	New births	UK	Birth register and maternity records	68.0	92.0
Woolf et al. 2000	Patients	USA	Medical records	94.0	67.0
Young et al. 2001	Households	Australia	Medicare records	43.9	49.4
Zell et al. 2000	Households	USA	Vaccination records	69.5–84.5	85.3

Table 2

Characteristics of sample respondents in the 2008 Health and Retirement Study (Weighted; N=6,384).

Characteristics	All respondents (Col %; N=6,384)	Consent (Row %; N=4,325)	No Consent (Row %; N=2,059)
Overall	--	67.8 (1.0)	32.2 (1.0)
Birth cohort			
AHEAD (< 1923)	11.7 (0.5)	65.3 (2.0)	34.7 (2.0)
CODA (1923–1930)	6.7 (0.5)	65.7 (2.7)	34.3 (2.7)
HRS (1931–1941)	55.5 (0.8)	69.0 (1.3)	31.0 (1.3)
WB (1942–1947)	12.4 (0.5)	68.5 (1.9)	31.5 (1.9)
EBB (1948–1953)	13.7 (0.7)	65.2 (2.4)	34.8 (2.4)
Gender *			
Male	39.5 (0.5)	69.1 (1.1)	30.9 (1.1)
Female	60.5 (0.5)	66.9 (1.1)	33.1 (1.1)
Race *			
White	82.0 (0.8)	69.1 (1.1)	30.9 (1.1)
Black	13.7 (0.7)	60.4 (1.9)	39.6 (1.9)
Other	4.3 (0.4)	65.8 (3.6)	34.2 (3.6)
Hispanic ethnicity			
Yes	8.8 (1.2)	70.4 (2.9)	29.6 (2.9)
NO			
Education			
0–11 years	23.2 (1.1)	66.9 (1.6)	33.1 (1.6)
12 years	34.3 (0.7)	68.0 (1.6)	32.0 (1.6)
13–15 years	20.9 (0.6)	66.2 (1.6)	33.8 (1.6)
16+ years	21.6 (0.9)	69.7 (1.6)	31.3 (1.6)
Marital status *			
Married	58.8 (0.8)	69.5 (1.1)	30.5 (1.1)
Separated/Divorced	14.6 (0.5)	64.9 (1.6)	35.1 (1.6)
Widowed	23.4 (0.4)	65.3 (1.6)	34.7 (1.6)
Never married	3.2 (0.2)	65.9 (2.7)	34.1 (2.7)

Parenthetical entries are standard errors.

* Comparison is statistically significant ($P < 0.05$).

Table 3

Model fit statistics computed after removing indicators for each mechanism in the 2008 Health and Retirement Study.

Hypotheses	LR statistic	P-value	Rel AIC	Rel SC
Accuracy/Uncertainty	0.86	0.99	0.04	0.21
Preexisting Relationship	9.47	0.49	0.07	0.10
Acquiescence/Attentiveness	5.11	0.88	0.04	0.05
Privacy/Confidentiality	43.48	<0.01	0.56	0.44
Interview Resistance	283.82	<0.01	3.63	2.73
Interviewer Attitudes	3.66	0.96	0.17	0.85
Interviewer Experience	7.76	0.65	0.07	0.09
Interviewer-Respondent Matching	10.50	0.40	0.24	1.43

Table 4

Final logistic regression model of consent to data linkage on respondent background characteristics and indicators of consent mechanisms.

Indicators of Mechanisms	Odds Ratio (95% CI)
Respondent Background	
Gender	
Male	1.03 (0.92–1.16)
Female	Referent
Age cohort	
AHEAD (< 1923)	Referent
CODA (1923–1930)	0.90 (0.64–1.26)
HRS (1931–1941)	1.64 (0.94–2.88)
WB (1942–1947)	1.13 (0.77–1.66)
EBB (1948–1953)	0.87 (0.57–1.33)
Hispanic	
Yes	1.27 (0.82–1.97)
No	Referent
Race	
White	Referent
Black	0.80 (0.63–1.01) [†]
Other	0.78 (0.53–1.14)
Marital status	
Married	1.25 (0.84–1.86)
Separated/Divorced	1.09 (0.75–1.58)
Widowed	1.15 (0.79–1.67)
Never married	Referent
Education	
0–11 years	Referent
12 years	1.11 (0.90–1.38)
13–15 years	1.01 (0.80–1.27)
16+ years	1.32 (1.02–1.70) [*]
Foreign born	
Yes	1.17 (0.86–1.60)
No	Referent
Perceived health	
Excellent	Referent
Very good	0.99 (0.75–1.29)
Good	0.89 (0.67–1.17)
Fair/poor	0.93 (0.70–1.23)
Numeracy skills	
Good	Referent

Indicators of Mechanisms	Odds Ratio (95% CI)
Poor	0.85 (0.65–1.12)
Employment status	
In labor force	Referent
Disabled	1.16 (0.85–1.57)
Retired	0.99 (0.80–1.22)
Homemaker	0.88 (0.66–1.16)
Household income	
Q1	Referent
Q2	0.87 (0.69–1.11)
Q3	1.14 (0.87–1.47)
Q4	1.19 (0.91–1.56)
Net worth	
Q1	Referent
Q2	0.75 (0.62–0.92) **
Q3	0.63 (0.48–0.82) ***
Q4	0.54 (0.38–0.76) ***
Accuracy/Uncertainty	
Financial “don’t knows”	
2008 wave	0.83 (0.45–1.52)
2006 wave	1.21 (0.52–2.83)
Preexisting Relationship with Agency	
SS income or SSI	1.05 (0.87–1.28)
Yes	Referent
No	
Other government income	
Yes	1.34 (1.10–1.65) **
No	Referent
Acquiescence/Attentiveness	
Rate attentiveness (2006)	
Not very attentive	1.22 (0.96–1.56) †
Very attentive	Referent
Privacy/Confidentiality Concerns	
Confidentiality concern index (2006)	0.62 (0.51–0.74) ***
Financial refusals	
2008 wave	0.68 (0.47–0.99) *
2006 wave	0.88 (0.79–0.97) *
Interview Resistance	
Uncooperative index (2006)	0.76 (0.70–0.82) ***
Initial refusal/resistance	

Indicators of Mechanisms	Odds Ratio (95% CI)
Yes	0.57 (0.38–0.85) **
No	Referent
Total call attempts	
2008 wave	0.86 (0.70–1.06)
2006 wave	0.89 (0.78–1.02) †
Ever wave nonrespondent	
Yes	0.64 (0.50,0.82) ***
No	Referent
Interview mode (2008)	
Face-to-face	Referent
Telephone	0.34 (0.23–0.48) ***
Interviewer Demographics	
Age	1.01 (0.98–1.03)
Gender	
Male	0.86 (0.39–1.89)
Female	Referent
Race	
White/other	Referent
Black	0.84 (0.25–2.83)
Interviewer Attitude Proxy	
Education	Referent
12 years	0.89 (0.04–18.11)
13–15 years	0.91 (0.05–18.16)
16+ years	
Interviewer Experience Measures	
New hire	1.20 (0.58–2.46)
Yes	Referent
No	
Number of prior interviews	0.86 (0.70–1.06)
Number of prior consents	1.03 (1.00–1.06) †
Interviewer variance	1.52 (0.97–2.38) †

† P < 0.10;

* P < 0.05;

** P < 0.01;

*** P < 0.001

Table 5

Percentage of item “don’t know” and refusals to financial questions by consent propensity stratum in the 2008 Health and Retirement Study.

	N	Consent (%)		Item refusals (%)		Item “don’t knows” (%)	
		2006	2008	2006	2008	2006	2008
Consent propensity stratum							
1 (Low)	1276	40.3 (1.4)	8.7 (0.8)	11.4 (0.9)	1.8 (0.4)	1.1 (0.3)	
2	1277	61.9 (1.4)	2.4 (0.4)	2.4 (0.4)	0.3 (0.2)	0.9 (0.3)	
3	1277	72.0 (1.3)	1.0 (0.3)	1.1 (0.3)	0.3 (0.2)	0.4 (0.2)	
4	1277	80.0 (1.1)	0.9 (0.3)	0.4 (0.2)	0.6 (0.2)	0.2 (0.1)	
5 (High)	1277	84.6 (1.0)	0.6 (0.2)	0.6 (0.2)	0.2 (0.1)	0.3 (0.2)	
P-value (trend test)		0.017	0.017	0.083	0.333	0.083	

Parenthetical entries are standard errors.