# Designing Consent Protocols to Link Sensitive Health and Administrative Records in Social Science Surveys: Phase I*

## Final Report

V. Joseph Hotz
Duke University

Vladislav Slanchev
Duke University & L&M Policy Research

September 1, 2017

*This research was funded by The MURDOCK Study. We wish to thank D. Sunshine Hillygus and Steven Snell from the Duke Initiative on Survey Methodology for their work on developing the survey instrument for this study and Kristen Cooksey, Stephanie Karol and Meagan Madden for their research assistance. All opinions expressed in this report and any errors are the sole responsibility of the authors. In particular, this report does not necessarily represent the position of The MURDOCK Study.

# SUMMARY OF FINDINGS

The main goal of the study was to assess what considerations influence survey respondents' **willingness to consent** (**WTC**) to participate in research studies that include linking administrative data to their survey responses. We conducted this research through online surveys, which were fielded using standing panels of online survey responses. Our main findings are as follows:

- If designed properly, vignettes describing hypothetical study consent requests can be a viable tool for assessing determinants of consent decisions for linking administrative records and providing the **personally identifiable information** (**PII**) necessary to do the matching.
- From all sociodemographic groups explored in this study, likely consenters and non-consenters to administrative linkages differed the most in terms of age and sex, with women and people older than 65 years of age reporting lower willingness to consent.
- Answers to questions related to specific individual practices, behaviors or experiences related to handling personal data are stronger predictors of how likely respondents are to participate in a study involving record linkages than most socio-demographics and measures capturing general attitudes about privacy.
- Engagement in a large scope of online activities and trust in organizations handling personal information are characteristics, which have a strong positive correlation with self-reported likelihood to provide permission for administrative record linkages.
- Both likely consenters and non-consenters reported "Confidence in researchers to keep responses and information private" as the most important determinant for their study participation decisions.
- The type of administrative record for which linkage permission is requested (medical versus earnings) did not have an effect on the willingness to consent to record linkage. On the other hand, requesting the last 4 digits of SSN and date of birth rather than full SSN resulted in a substantially higher self-reported likelihood to provide the requested PII.
- Reported willingness to permit record linkages was considerably higher than willingness to provide the PII necessary for performing that linkage.
- Providing additional data security assurance to respondents did not have an impact on the willingness to consent to record linkages, while emphasizing potential benefits from the described research study led to a small but statistically significant decrease in the willingness to consent for linking medical records. The data security assurance was associated with higher willingness to provide full SSN for both medical and earnings record linkages.
- Two additional features of the request for PII were associated with higher willingness to provide such information – asking first for the set of PII that respondents are likely to find more sensitive and showing the exact identifying information researchers will have access to.

## 1. Background

It is well-documented that social and behavioral factors are key contributors to the onset and course of disease and associated morbidity and mortality.[1] While those factors are regularly measured in existing population-based social science surveys, such surveys either collect nothing on health or rely solely on self-reported measures of health and physical limitations.[2] At the same time, most biomedical studies collect detailed health information and health records, but typically only collect limited socioeconomic information, such as educational attainment.

The possibility of obtaining administrative records, such as electronic health records (EHRs) and/or Social Security records on earnings and benefits represents an important way of facilitating a rich research agenda on the relationships between social and behavioral factors and health outcomes. As suggested by leaders of the Population Association of America (PAA) and the Association of Population Centers (APC)[3], obtaining such information for population-based samples in existing or future studies, such as Health and Retirement Survey (HRS) or Add Health, can help ensure that inferences drawn from such research are less susceptible to selection or referral biases that may result from reliance on data drawn from volunteer samples, disease- or condition-based cohort studies, or health-system-based samples.

A key issue in the feasibility of linking such data to new or existing studies – especially population representative ones[4] – is obtaining the consent of subjects for such record linkages. Of particular interest are population-based surveys in which (1) potentially low consent rates restrict the sample size for analysis, and (2) if consenting respondents are not sufficiently representative of the population sampled that can introduce a form of non-consent bias into analyses based only on the subset of the survey sample with linked records. Understanding who consents to linking and why is crucial in designing strategies to increase respondents' willingness and comfort with providing their consents and personally identifiable information necessary to facilitate the linkage.

More specifically, in this study we address the following four questions:

1. Can hypothetical study consent requests be reliably used as a tool to study issues of consent for linking administrative records?
2. How do subjects who consent to linkage differ from those who do not in terms of socio-demographics, online behavior, political views, privacy attitudes, trust in institutions and specific practices, behavior or experiences associated with handling personal data?
3. How do rates of consenting for linkage vary by the consent request features such as types of records, framing of the consent question and payment incentives?
4. What type of **personally identifiable information** (**PII**) are subjects willing to consent to provide in order to facilitate the linkage of records?

---

[1] For example, there is clear evidence that behaviors, such as smoking, diet & activity, alcohol consumption, are major contributors to actual causes of death (McGinnis & Foege, 1993); that social relationships (or their absence) affect health (House, Landis & Umberson, 1988); and that socioeconomic status (SES) affects health (Marmot et al. 1991; Marmot, 1999). The latter causal claim of SES to health is controversial (see Smith, 1999; Adams et al., 2003; Deaton, 2013).

[2] Exceptions to this pattern include the Health and Retirement Study (HRS) and the National Longitudinal Study of Adolescent to Adult Health (Add Health), each of which has done in-person collection of biomarkers (height, weight, DNA and blood for assays of various conditions) from their respondents.

[3] See May 7, 2015 letter from Drs. Steven Ruggles, president of the PAA, and Lisa Berkman, president of the APC, responding to Request for information on NIH Precision Medicine Cohort.

[4] For a discussion when one needs and may not need population-representative data, see Duncan (2008).

## 2. Methods

The findings in this report are based on analysis of the responses to an online survey developed and administered by the research team specifically for the purposes of this project.

### *2.1. Sample and Administration*

The web survey was first piloted with about 1,000 respondents through Amazon's Mechanical Turk (MTurk) service to identify any problems with the questionnaire and finalize its content. Our main research was then conducted using Qualtrics' Online Sample Service, which recruits subjects from a collection of online opt-in panels and allows obtaining samples with a particular preferred distribution of population characteristics. With our final sample of 2,064 respondents we targeted a distribution of socio-demographic characteristics in terms of sex, age, education, race/ethnicity and location within US that is similar to the general US population. Table 1 compares the distributions of these characteristics for our sample, the quotas used by Qualtrics based on the 2010 Census and the 2015 US population according to the Current Population Survey 2015. Overall, the sample in this study resembles closely US population in 2015. However, by design it is not a probability sample and, therefore, our research strategy to analyze determinants of consent rates for administrative linkages relies on randomizing hypothetical consent requests among responders rather than on the representability of the sample.

Subjects recruited through MTurk received direct compensation of approximately $.50. Panelists recruited on our behalf by Qualtrics were compensated in a proprietary points system. The compensation awarded by Qualtrics and its contracted vendors generally averages the monetary equivalent of roughly $.50 per survey. All subjects were routed to a Duke University Qualtrics survey and consent was obtained through the opening screen of the particular survey. The median duration of the survey was 15 minutes and 24 seconds.

### *2.2. Questionnaire*

The sections of the questionnaire described below were developed in close collaboration with The Duke Initiative for Survey Methodology to address the research questions of interest. In addition to the questions used for the analysis in this report, the survey instrument includes secondary hypothetical study consent request asking for respondents' willingness to share data from meal planning apps with researchers and a survey experiment that measures the relationship between descriptions of voter registration processes and respondents' concerns about the public nature of voter registration information.

#### *2.2.1. Hypothetical Study Consent Requests*

At the core of this study is analyzing survey responses from a randomized experiment in which we randomly assigned respondents to see a single description of a hypothetical research study and then gauged respondents' opinions regarding that study. This experiment mimicked our ideal research design, in which we would randomize how researchers would approach potential study participants and request their participation in an in-person study. However, instead of asking for consent to participate in an actual study, we varied how we described a hypothetical study and then evaluated respondents' opinions about it. The **willingness to consent** (**WTC**) to participate was measured on a scale from 0 to 10 by asking respondents immediately after each vignette has been presented to them[5] with the following:

---

[5] We also included an alternative measure of willingness to consent based on how willing the survey respondents think members of their community are to take part in the same study based on the same scale from 0 to 10.

*How likely is it that you would agree to take part in the hypothetical study with the script available to you on the previous screen? Use a scale from zero to ten, where zero means you would definitely not agree to participate and ten means you would definitely agree to participate in this kind of study.*

Some of the elements of that fictional study were the same for each respondent – its topic and mode, the institutions conducting it, and the monetary incentives for participating in it. Three elements varied randomly across respondents – whether or not the study is initially described as being linked to administrative data, whether the proposed link is to medical or earnings records, and the type of benefit or assurance offered in the consent preamble – emphasizing data security, benefits to researchers or no additional assurance or benefit. Each survey respondent received only one of the 7 hypothetical study consent requests. Table 2 shows that between 284 and 343 respondents received each of the hypothetical study requests associated with medical or earnings records linkages.

### *2.2.2. Hypothetical Requests for Personally Identifiable Information (PII)*

After gauging respondents' initial WTP in the hypothetical study consent requests, we additionally asked for their willingness to consent if such a study required the participant to disclose (1) the last 4 digits of the Social Security number (SSN) and date of birth or (2) their full SSN in a signed form. Those two sets of PII were chosen because they both allow linking to the Medicare/Medicaid records of study participants, as well as potentially to Social Security Administration records. All respondents saw both hypothetical requests in a randomized order. Important to this design, we did not actually collect SSNs or birthdates—we only assessed hypothetical willingness to provide such PII. After the hypothetical request for PII, the survey included a debriefing question which asked respondents how important particular factors such as reputation of the institution conducting the research or being paid to participate are for the respondent's decision to participate in such a study and provide necessary PII for linking to administrative records.

### *2.2.3. Randomized Experiment for Requesting Respondents' IP address*

At the end of the survey there was a request that respondents provide relatively non-invasive information that would allow researchers to pair their individual responses with very basic information about the area in which they live. Specifically, the experiment asked respondents if the researchers could use their IP address to identify the area in which they lived and match it to neighborhood characteristics. In randomly assigned versions, respondents were asked to confirm or to provide their IP address. While IP addresses are collected routinely within Qualtrics surveys and IP addresses do not uniquely identify respondents, the goal of this experiment was to draw an analog to surveys that request identifying information such as name, address, SSN, and birth date in order to facilitate matching across data sources.

### *2.2.4. General Questions*

The survey also included some more general questions about the respondents which were not directly linked to the hypothetical study consent requests and the IP experiment, and instead fitted in the categories below:

- socio-demographics such as age, sex, race/ethnicity, location, marital status, place of residence and income;
- party affiliation, political views, religion;
- general trust towards others and trust toward particular institutions which keep information about them;
- perceived value of scientific research to them, their community and the society in general;
- attitude toward the U.S. Census;
- concerns about privacy and confidentiality;

- past experience that can potentially affect willingness to provide PII such as being victim to identity theft;
- individual practices the respondent is undertaking to safeguard sensitive personal data;
- type of information respondents find sensitive;
- medical history; and
- alcohol and recreational drug use.

## 3. Findings

We group our findings based on the major questions which the study aimed to address. All reported results concerning statistical significance are using the 5% level.

### *3.1. Can hypothetical study consent requests be reliably used as a tool to study issues of consent for linking administrative records?*

While vignettes have been extensively used to explore determinants of consent for study participation and record linkages (see Couper et al., 2008), there has been less documented research effort on assessing whether respondents' reactions to hypothetical study and linkage consent requests are an indicator as to what individuals would do if they were presented with the actual consent decision. We utilized several indirect approaches to identify any potential issues for interpreting respondent's WTP for a hypothetical study as a proxy measure for how likely they are to participate in it if they were actually recruited to join it. The results did not point to evidence that WTP measure constructed from our survey questions cannot be reliably used as such a proxy measure for the actual consent decision.

First, as shown in Table 3, we compared the average WTP for respondents, who received the hypothetical consent request for a study that requires authorization for linking medical or earnings administrative records, to the average WTP for respondents, who received a vignette for a study that did involve any linking request. If respondents understand the implications of the hypothetical consent request they are receiving, all else equal, one would expect that they would be less willing to participate in a study that involves them agreeing to administrative records linkage. Table 3 shows that, indeed, respondents who did not receive a linkage consent request had 1.47 points (or 0.46 fraction of the standard deviation of 3.16 for WTP) higher average WTP than those who had a linkage authorization request in the consent preamble for their hypothetical study. The difference in WTP is statistically significant and provides some evidence that respondents are taking into consideration administrative records linkages when assessing the hypothetical study consent request.

Second, in Table 3 we also compared the average WTP of respondents who agreed to provide their IP address to the average WTP of those who did not. The IP address request was not a hypothetical one and asked respondents for information which, on average, they found less sensitive than SSN and information about their finances, but more sensitive than health history, political views, purchasing habits, date of birth and browsing history. The average WTP of respondents who did not consent to providing their IP address is 2.09 points (or 0.66 fraction of the standard deviation for WTP) lower than the average WTP for respondents who provided their IP address. Therefore, people who reported higher likelihood to participate in the hypothetical study were also more likely to actually provide some form of sensitive information, which supports our method of using responses to the vignettes in this study as proxy measures for consent decisions regarding actual studies.

Third, the average WTP for respondents who provided more complete information on their income, health history and alcohol and marijuana use was compared to the WTP of respondents who refused to provide any information on income or provided incomplete information on health and marijuana and alcohol use. It has been established in previous research (Sakshaug et al., 2012) that individuals who refuse to answer questions asking for sensitive information are also more likely to refuse linking administrative records. As the results in Table 3 show, we found similar evidence for our WTP measure – respondents who did not provide any information about their income had on average 2.37 points lower WTP than respondents who at least pointed out the range in which their income falls; respondent who did not answer at least one of the questions on health history had on average 1.11 points lower WTP than those who answered all of those questions; and respondents who did not answer the alcohol or marijuana use questions had on average 1.17 points lower WTP than everyone else. All those differences are statistically significant and show that WTP and refusals to provide consent for administrative records have similar correlates.

Fourth, we addressed the possibility that presented with hypothetical situations, respondents may want to present themselves as more inclined to perform the action that seems more socially acceptable, which in this case is likely to be participating in the hypothetical research study and authorizing administrative record linkage. That is why we developed and compared results for two measures of WTP – one asking respondents to mark how likely it is that they themselves would participate in such a hypothetical study and one asking how likely it is that most people in their community would take part in it. Respondents considered themselves more likely to participate in the presented hypothetical studies than most people in their community since their own WTP measure was on average 0.75 points (or 0.24 fraction of the standard deviation for that WTP measure) higher than the WTP measure based on their community members' likely consent decisions. The correlation between those two WTP measures is 0.61. We chose to perform our analysis using the WTP measure based on respondents' own participation and linkage permission decisions, since it has a stronger conceptual appeal and a higher correlation with the IP address consent – that is, correlation of 0.33 versus 0.17 for the WTP measure for most people in their community.

### *3.2. How do subjects who consent to linkage differ from those who do not?*

Knowing more about the sociodemographic characteristics of consenting respondents such as age, sex, socioeconomic status and ethnicity provides an opportunity to assess the risk for non-consent bias raised in Section 1. It also allows us to identify groups with potentially low consent rates for which additional effort is required to understand their perspectives and reservations about providing linking authorization, as well as to design specific protocols to decrease the elevated risk of refusal to linkage associated with them. Our interest in this issue is also motivated by the inconclusive nature of the findings from biomedical studies as to how sex and age affect the likelihood of consent for linking to medical records (Kho et al., 2009) and the fact that more systematic research on this topic with US population-based samples (Sakshaug et al., 2012) has been primarily based on the HRS sample which contains only older respondents. We also assessed how consenters and non-consenters are likely to differ by privacy and confidentiality attitudes, political views, online activities and experiences and pattern of behavior related to handling PII.

In order to provide a common analytic framework to present the difference between consenters and non-consenters across multiple dimensions of interest, we analyze the sample of 1,855 survey respondents who received a hypothetical study consent request containing an administrative records linkage request by dividing it into three groups corresponding to the terciles based on respondents' expressed likelihood to participate in such a hypothetical study. Therefore, the high willingness-to-consent group contains the third of the respondents with linkage requests who expressed the highest likelihood to participate in the study (593 respondents with WTP of 9 or 10), the low willingness-to-consent group includes the third who expressed the lowest likelihood to participate (647 respondents with WTP in the range between 0 and 5), and the medium willingness-to-consent group consists of the remaining survey respondents in that sample (615 respondents with WTP between 6 and 8). The characteristics of those three *willingness-to-*

*consent groups* are compared to identify the factors in which consenters and non-consenters are likely to differ.

Table 4 captures the different distribution of socio-demographics across the three groups. Relative to the other groups, the high willingness-to-consent group contains a higher percentage of respondents who are male (53.1% versus 47.6% for all respondents with linkage requests), aged between 35 and 54 (43% versus 36.7%) and married (54.6% versus 50.7%). The patterns for education, race/ethnicity and area of residence were less pronounced. The identified populations for which obtaining consent for linkages seems to be most problematic are women, people 65 years of age or older and respondents with no high school degree.

In Table 5 we present the distribution of political views, party affiliation, frequency of attending religious services, income and time spent online and on social media across for all three groups. The low willingness- to-consent has a disproportionate percentage of people who do not define their political views as liberal, do not report being closer to the Democratic party and never attend religious services. Daily hours spent online and on social media appeared to be the characteristics most correlated with likelihood to consent according to the table. The low-willingness-to-consent group contains a much larger proportion of people who use Internet less than 4 hours a day (31.8% versus 22.6% for the high willingness-to-consent group) or do not use social media at all (30.4% versus 18.7% for the high willingness-to-consent group). The distribution of reported income from wages is not characterized by very consistent patterns since for all income ranges the percentage distributions for the medium willingness-to-consent group do not fall between those for the low and high willingness-to-consent groups. Furthermore, non-neighboring income ranges sometimes share more similarities than neighboring ones when it comes to comparing the percentages across the three groups. In this context, it is important to note that the statistics here capture only people who gave some information about their income.

The level of importance that respondents assign to different aspects of privacy does not seem to be a particularly strong determinant of which willingness-to-consent group they belong to. Table 6 shows that the vast majority of respondents in any group find particularly important privacy aspects such as being in control of who can get information about them (87.2% for low willingness-to-consent group and 90.4% for high willingness-to-consent group) or not being disturbed at home (74.6% for low willingness-to-consent group and 74.2% for high willingness-to-consent group). A different picture emerges when privacy attitudes are assessed with a question pertaining to a concrete initiative – the Decennial Census. 38.7% of the respondents in the low willingness-to-consent group considered an invasion of privacy while only 28.4% shared the same feeling in the high willingness-to-consent group.

Table 7 and 8 present further evidence that questions related to specific individual actions, behavior or experiences are stronger correlates of belonging to a particular willingness-to-consent group than some socio-demographics and measures capturing general attitudes about privacy. For example, 47.9% of the respondents in the high willingness-to-consent group report very frequently signing up for frequent buyer discounts, while for the low willingness-to-consent group this is the case for only 27.5% of the respondents. Similar differences between the likely non-consenters and consenters can be identified for all reported online behavior associated with providing potentially sensitive data. For example, 45.9% of the high willingness-to-consent group reported very frequently buying a product online as opposed to some other way compared to 25.5% of the low willingness-to-consent group.

In terms of concrete action related to preserving one's privacy and personal information, Table 7 presents evidence that respondents in the high willingness-to-consent group report more frequently reading privacy notices that came in the mail from banks and insurance companies, changing passwords on financial accounts and checking the security features of websites. However, as Table 8 shows, higher percentage of the respondents on the low willingness-to-consent group report being on the do-not-call list (62.6% versus 54.3% for the high willingness-to-consent group), having no SSN on cards in their wallet or purse (82.1% versus 62.6% for the high willingness-to-consent group). Therefore, there is mixed evidence as to which group of respondents do more, on average, to protect their personal data. Prior experience with fraudulent purchases with one's own information does not seem to have an impact on selection into willingness-to-consent groups. Individual online practices on social media differ across likely consenters and non-consenters where respondents in the high willingness-to-consent group are all more likely to have provided personal information on social media such as date of birth, home address, phone number and names of family members than the respondents in the other willingness-to-consent groups.

### ***3.3. How do rates of consenting for linkage vary by the consent request features?***

Analysis of the characteristics of the linkage request which are likely to be associated with consenting rates allows (1) assessing in which cases (e.g., types of records) non-consent bias can be of concern due to low consent rates, and (2) identifying best practices for designing the consent process (e.g., incentives for respondents, wording of the consent question) which have the potential to increase consent rates. Our findings are based on comparing the willingness to consent across the different versions of the hypothetical study consent request, as well as on survey responses to questions regarding the factors behind respondents' decisions to participate in the hypothetical study, what type of information they find sensitive and how much they trust institutions which handle information about them.

#### *3.3.1. Types of administrative records*

Table 9 shows that in all three willingness-to-consent groups defined in Section 3.2 there were more people defining financial information as particularly sensitive (84.7% in the low willingness-to-consent group and 76.4% in the high willingness-to-consent group) than people who found health information to be that sensitive (64.9% for the low willingness-to-consent group and 59.0% for the high-willingness-to-consent group). This pattern is consistent with the statistics presented in Table 3 that respondents were less likely to answer questions about their income questions relative to questions related to their health history. However, the evidence that respondents found financial information more sensitive than health history did not carry over when the results from the vignette experiment are taken into account.

From all 1,855 respondents who received a hypothetical linkage authorization request as part of the study consent, 1,001 were asked for permission to link to medical records and 854 were asked to allow linking to earnings records. In the survey questions medical records are defined as "*your medical records from the healthcare providers you most frequently visit*, while the earnings ones as "*your earnings from past employment that are contained in governmental records*". Table 10 shows that the average WTP for earning records is higher but the difference is small (0.18 points or 5.6% of the standard deviation for WTP of 3.20) and not statistically significant. There are only three groups for which the difference between the mean of WTP for medical and earnings records is statistically significant – for respondents with high school degree or some years of college (0.46 points higher for earnings records), people aged between 55 and 64 (1.04 points higher for earnings records) and single respondents (0.50 points higher for earnings records). Thus, the combined results from the vignettes and the relevant survey questions do not suggest that people are more likely to consent to link earnings records rather than medical ones or vice versa.

#### *3.3.2. Incentives to participate*

After respondents reported the likelihood of them participating in a hypothetical study and providing PII for the linkages involved in that study, they were asked how important were particular factors for their decision to take part in that study. Results in Table 11 show that if the factors are ordered by how important they are for each group[6], all three groups reported “Confidence in researchers to keep responses and information private” as the most important one followed by “Understanding how the provided data will be used” and “Reputation of institution conducting the research”. Factors such as payment incentives or potential benefit of the study to the respondent and their family appeared to be secondary for the decision-making process.

Respondents in the high willingness-to-consent group reported much more frequently than the respondents in the low willingness-to-consent group that the factors listed as potential explanations for their consent decision had been particularly important determinants of that decision. The biggest difference in percentage points (37) is found for the “Feeling a duty to participate” factor which 55.1% reported as particularly important in the high willingness-to-consent group compared to 18.1% in the low willingness-to-consent group. The smallest such difference in percentage points is 13.4 for “Confidence in researchers to keep responses and information private”.

### *3.3.3. Trust in the organization initiating the request*

67.3% of the respondents[7] who received a hypothetical study consent request with record linkage reported the reputation of the institution as being “Extremely important” or “Very important” for their decision whether to participate in that study[8], while 79.5% did so regarding their confidence in researchers to keep their information private. Both of those factors are linked to the issue of trust in the organization handling their data which proves to be a very important determinant of the consent decision. That is consistent with the success of models for obtaining consent like the one employed by HRS, which depend on multiple contacts with the respondents to gain their trust (Sakshaug et al., 2012).

In order to identify what type of institutions respondents find more trustworthy when it comes to handling their data, in Table 12 we show the percentage of respondents within each willingness-to-consent group who trust government, survey organizations, financial institutions, companies, political parties, healthcare providers, IRS and research institutions. The institutions which all groups trust the most are healthcare providers (81.8%), universities (77.0%) and financial institutions (71.3%) while government (49.8%) and political parties and candidates (36.8%) are at the bottom of that ranking. Even though the ranking of the organizations is very similar for the low and high willingness-to-consent groups, the level of trust in the low willingness-to-consent group is lower for all of the organizations and the difference is most significant in magnitude for government and political parties.

The low willingness-to-consent group is also characterized by a lower mean of the general trust measure[9] compared to the high willingness-to-consent group. The difference in the means of 0.21 points is 17.8% of the standard deviation for that measure and is statistically significant. However, unlike the measures of trust in the particular institutions, the general trust measure fails to reflect any differences between the medium and high willingness-to-consent groups.

### *3.3.4. Framing of the consent question*

With our vignette experiment we tested three versions of the consent request preamble that were randomly assigned among all survey respondents who were asked for permission to link their

---

[6] Importance is measured by the fraction of the respondents in that group who determine a particular factor to be “Extremely important” or “Very important” for their decision.

[7] See Table 11.

[8] The vignettes described the study as conducted by researchers at Duke University and funded by the National Institute of Health.

[9] The measure is based on a 3-item questionnaire to assesses participants’ general beliefs about honesty and trustworthiness of others.

administrative records. The first (benchmark) version of the preamble (named “No Additional Assurances” in Table 13) contained a brief summary of who conducts the study, what kind of questions are included in it, the general goal of the study, the approximate time it would take to complete and the amount of money participants will receive for completing it. The second version of the preamble (“Data Security Assurance” included the same elements plus an additional sentence on data security provisions:

*“All of the data collected will be stored and analyzed on highly protected computer systems. Only researchers who have been approved by the study team at Duke University can have access to the data and are subject to strict security protocols.”*

The third version of the preamble (“Emphasis on Research Benefits”) also contained the basic information from the “No Additional Assurances” preamble but in addition there are two sentences in it elaborating on how the produced analysis will benefit research and members of the respondents’ communities who have experienced problems:

*“We hope to learn more about the role that family members play in helping each other secure access to needed social and health care services. Insights based on this research will be used to improve the quality of life for members of your community, especially those who face challenging economic and social problems.”*

Table 13 shows that for medical records linkages the “Data Security Assurance” preamble does best in terms of average WTP associated with the responses to it, followed by “No Additional Assurance” and then “Emphasis on Research Benefits”. While the difference in the mean of WTP between “No Additional Assurance” and “Data Security Assurance” is minimal and not statistically significant, the additional sentences in “Emphasis on Research Benefits” led to lower average WTP in comparison to the benchmark case – that is, a decrease of 0.28 or 8.7% of the standard deviation for WTP. This pattern is observed only for medical records, whereas with earnings records, WTP has similar levels across the three versions of the preamble. The results suggest that increasing the consent rates by putting emphasis on particular issues that are likely to affect the motivation of the potential study participants is not straightforward and adding more information to the consent request can potentially even lead to a negative effect on consent rates as reported by studies like Das and Couper (2014).

### ***3.4. What type of personally identifiable information (PII) are subjects willing to provide in order to facilitate the linkage of records?***

Even if people are willing to authorize linkage to their administrative records, the actual process of matching often requires the use of identifiable information such as address, date of birth or a unique identification number for type of records in question. In most studies participants need to provide this information themselves as part of the consent process which leads to refusals for linkages in cases where the respondents are fine with researchers having access to their administrative records but not to the identifiable information necessary for obtaining them. The statistics in Table 9 show that on average respondents reported SSN to be more sensitive information than their medical and financial records, while at the bottom of that ranking IP address and date of birth were more frequently defined as particularly sensitive information than political views and supported candidates, websites visited and basic purchasing habits.

The survey vignettes also addressed the issue of study participants’ willingness to consent to provide PII for research purposes. After respondents reported how likely they are to participate in a hypothetical study with linkages, they were asked also to report on a scale from zero to ten how likely they are to provide full SSN in a signed form and separately the last digits of their SSN plus their date of birth. Those sets of PII were requested from each respondent in a randomized order. The willingness to consent to provide full SSN for linking purposes of 3.19 (as measured by the self-reported likelihood to do so) was on average more than 3 points lower than the WTP of 6.40 for studies with medical and earnings record linkages. As Table 14 shows, the socio-demographic groups which were least likely to provide their full SSN were women and people 55 years of age or older. Asking for the last 4 digits of SSN and date of birth was associated with higher levels of self-reported likelihood to provide them (4.43) which, however, is still well below WTP for a study that involves administrative record linkages. While Dahlhamer et al. (2007) already established that people are more likely to provide the last 4 digits of their SSN than their full SSN, we found that this result holds even when people are asked to give their date of birth in addition to the last 4 digits of SSN.

We explored the potential impact of three features of the request for PII on respondents' self-reported likelihood to provide it. Since that request was associated with the hypothetical studies described prior to it, we looked into whether the preambles "No additional assurance", "Data security assurance" and "Emphasis on Research Benefits" had any effect on the willingness to consent to provide full SSN or the last 4 digits of SSN and date of birth. As Table 13 shows, the difference between the mean of WTP for the "Data security assurance" and "No additional assurance" preambles of 0.47 points (or 14.7% of the standard deviation for WTP) was statistically significant and suggests that data security language has the potential of increasing consent rates. For the last 4 digits of SSN and date of birth there was no statistically significant difference between the mean of WTP for the three preambles.

Table 14 and Table 15 show how the order of asking respondents for the two sets of PII affects their willingness to consent to provide them. Asking for full SSN first and for last 4 digits of SSN and date of birth second lead to higher self-reported likelihoods to provide both sets of PII – 0.9 points higher WTP for full SSN and 0.66 points higher WTP for the last 4 digits of SSN and date of birth. For most socio-demographic groups both differences were statistically significant. Those statistics suggest that if more than one set of PII may be useful to the study, respondents should receive the request for the more sensitive information first.

The IP address actual consent request in the survey had two versions which were randomly assigned to the respondents. The first one asked for the IP address without showing it on the screen, while the second one included it in the request. We were testing the hypothesis that respondents might be more willing to permit researchers to use information which they know exactly how it looks like, is easy to provide and the researchers already have access to it. The consent rate was 56.8% for the versions in which the IP address was provided in the request and 52.3% for the version in which it was not. The difference of 4.5 percentage points was relatively small but statistically significant suggesting that showing the respondents the information for which their consent is required can potentially increase consent rates.

## 4. Concluding Discussion and Next Steps

Our study led to the following findings:

- If designed properly, vignettes describing hypothetical study consent requests can be a viable tool for assessing determinants of consent decisions for linking administrative records and providing the PII necessary to do the matching.
- From all sociodemographic groups explored in this study, likely consenters and non-consenters to administrative linkages differed the most in terms of age and sex, with women and people older than 65 years of age reporting lower willingness to consent.
- Answers to questions related to specific individual practices, behaviors or experiences related to handling personal data are stronger predictors of how likely respondents are to participate in a study involving record linkages than most socio-demographics and measures capturing general attitudes about privacy.

- Engagement in a large scope of online activities and trust in organizations handling personal information are characteristics, which have a strong positive correlation with self-reported likelihood to provide permission for administrative record linkages.
- Both likely consenters and non-consenters reported "Confidence in researchers to keep responses and information private" as the most important determinant for their study participation decisions.
- The type of administrative record for which linkage permission is requested (medical versus earnings) did not have an effect on the willingness to consent to record linkage. On the other hand, requesting the last 4 digits of SSN and date of birth rather than full SSN resulted in a substantially higher self-reported likelihood to provide the requested PII.
- Reported willingness to permit record linkages was considerably higher than willingness to consent to provide the PII necessary for performing that linkage.
- Providing additional data security assurance to respondents did not have an impact on the willingness to consent to record linkages, while emphasizing potential benefits from the described research study led to a small but statistically significant decrease in the willingness to consent for linking medical records. The data security assurance was associated with higher willingness to provide full SSN for both medical and earnings record linkages.
- Two additional features of the request for PII were associated with higher willingness to consent to provide such information – asking first for the set of PII that respondents are likely to find more sensitive and showing the exact identifying information researchers will have access to.

These results outline several directions in which the scope of the research effort initiated here can be broadened to obtain a more comprehensive picture of who consents to administrative records linkages and develop best practices to increase consent rates and address non-consent bias. First, multivariate regression analysis using the data presented in this report can be used to assess whether each of the strong predictors of the willingness to consent to record linkages such as online behavior, trust, individual practices related to safeguarding PII, age and sex remain significant determinants of the consent decision process once all factors are controlled for[10].

Second, administering the survey to a sample that is not restricted only to online opt-in panels would allow testing the validity of the results presented in this report and assessing their generalizability beyond particular types of Internet users. Such a research effort would be especially valuable for confirming whether online behavior has indeed such a strong predictive power when it comes to explaining willingness to consent to records linkages and identify behaviors related to safeguarding PII beyond items related to Internet use which can help predict consent decisions for subjects who are not frequently online. Potential candidate samples for that goal are the participants in the MURDOCK Study and people who after initial contact with the MURDOCK Study decided not to take part in it.

Third, our mixed findings from the vignette experiments with framing of the linkage consent request by putting stronger emphasis on data security or benefits from research coming from the study suggest that

---

[10] Further insights into respondents' privacy and confidentiality concerns in relation to their willingness to consent in research studies involving sharing potentially sensitive information can be obtained from analyzing two other items included in the survey: a secondary experimental vignette asking for respondents' willingness to share data from meal planning apps with researchers and a survey experiment that measures the relationship between descriptions of voter registration processes and respondents' concerns about the public nature of voter registration information.

additional cognitive testing with focus groups may be necessary to identify the next iteration of consent preamble wording for testing. The results presented in this report suggest that it would be worth exploring preambles which attempt to raise respondents' trust in the organization conducting the research and inspire their confidence that their data will be kept secure and confidential.

Fourth, experiments can be designed to incorporate the insights presented in this report for potentially increasing the respondents' willingness to consent to provide access to PII for record matching. In addition to experimenting with different consent language in the preambles, testing can involve putting multiple consent request in an order that would maximize the amount of PII respondents are willing to provide.

Ideally, variants of the described experiments would be conducted with samples from several different populations. These would include (1) participants in a population-representative on-line survey, such as the Understanding America Study (UAS)[11] and (2) individuals included in a Health Care System (for example, in Durham) that has a well-developed system of Electronic Health Records.

## 5. References

Adams, P., M. Hurd, D. McFadden, A. Merrill, and T. Ribeiro. (2003), "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status," *Journal of Econometrics* 112(1):3–56.

Dahlhamer, J., Meyer, P., & Pleis, J. R. (2007), "Questions People Don't Like to Answer: Wealth and Social Security Numbers." *Proceedings of the 2006 Joint Statistical Meetings*.

Das, M. and M. Couper (2014), "Optimizing Opt-Out Consent for Record Linkage," *Journal of Official Statistics*, 30(3): 479–497.

Deaton, A. (2013), *The Great Escape: Health, Wealth, and the Origins of Inequality*, Princeton University Press.

Duncan, G.J. (2008), "When to Promote, and When to Avoid, a Population Perspective." *Demography*. 45(4): 763-784.

House, J. S., Landis, K. R., & Umberson, D. (1988), "Social relationships and health." *Science*, 241(4865), 540-545.

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." *BMJ*, 338, b866.

Marmot, M. (1999), "Multi-Level Approaches to Understanding Social Determinants," in *Social Epidemiology*. Berkman, Lisa and Ichiro Kawachi, eds. Oxford: Oxford University Press.

Marmot, M. G., G. Davey Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, & A. Feeny (1991), "Health Inequalities among British Civil Servants: The Whitehall II Study." *Lancet*. June 8, pp. 1387–93.

McGinnis, J. M., and W. H. Foege (1993), "Actual causes of death in the United States." *Journal of the American Medical* Association. 270(18): 2207-2212.

Sala, E., Knies, G., & Burton, J. (2014), "Propensity to consent to data linkage: experimental evidence on

---

[11] The UAS is a household panel maintained by the Center for Economic and Social Research (CESR) at University of Southern California (USC), consisting of approximately 2,000 respondents ages 18 and older who are regularly interviewed over the Internet.

the role of three survey design features in a UK longitudinal panel." *International Journal of Social Research Methodology*, 17(5): 455-473.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012), "Linking Survey and Administrative Records Mechanisms of Consent." *Sociological Methods & Research*, 41(4): 535-569.

Smith, J.P. (1999), "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status." *Journal of Economic Perspectives*, 13(2): 145-166.

**Table 1. Socio-demographic Characteristics of the Study Sample and the U.S. Population (%)**

| | **Study Sample** | **Qualtrics Quota*** | **2015 US Population**** |
|---|---:|---:|---:|
| ***Sex*** | | | |
| Male | 47.7 | 48.3 | 48.3 |
| Female | 52.2 | 51.7 | 51.7 |
| *Total* | *100* | *100* | *100* |
| ***Age*** | | | |
| 18-34 | 28.9 | 29.0 | 30.2 |
| 35-54 | 36.5 | 37.0 | 35.9 |
| 55-64 | 16.3 | 16.0 | 14.9 |
| 65+ | 18.3 | 18.0 | 19.0 |
| *Total* | *100* | *100* | *100* |
| ***Race*** | | | |
| Non-Hispanic White | 62.9 | 63.0 | 64.9 |
| Non-Hispanic Black | 12.4 | 13.0 | 11.7 |
| Hispanic | 17.4 | 17.0 | 15.5 |
| Other | 7.3 | 7.0 | 7.9 |
| *Total* | *100* | *100* | *100* |
| **Education** | | | |
| Less than High School Graduate | 12.1 | 12.0 | 12.2 |
| High School Graduate or Some College | 56.2 | 56.0 | 58.0 |
| Bachelor or Graduate degree | 31.7 | 32.0 | 29.8 |
| *Total* | *100* | *100* | *100* |
| **Census Region** | | | |
| Northeast | 17.8 | 18.0 | 18.1 |
| Midwest | 22.2 | 22.0 | 21.3 |
| South | 36.8 | 37.0 | 37.2 |
| West | 23.2 | 23.0 | 23.5 |
| *Total* | *100* | *100* | *100* |

** Based on Census 2010*

***Based on Current Population Survey 2015*

**Table 2. Number of Survey Respondents Receiving Each Type of Hypothetical Linking Consent Request**

| | Type of Data Linkage Request: | | | |
|---|---|---|---|---|
| **Type of Security Assurance in Preamble:** | **None** | **Earnings Records** | **Health Records (EHRs)** | **Total** |
| None | 208 | 284 | 341 | **833** |
| Data Security | 0 | 285 | 317 | **602** |
| Research Benefits | 0 | 285 | 343 | **628** |
| Total | **208** | **854** | **1,001** | **2,063** |

**Table 3. Comparisons of Average Willingness to Consent (WTC) to Participate**

| | # Respondents | % of all Respondents | Avg. WTP |
|---|---:|---:|---:|
| **Vignette/Hypothetical Study** | | | |
| Respondents with No Record Linkage Request | 208 | 10.1 | 7.87 |
| Respondents with Medical or Earnings Records Linkage Request | 1,855 | 89.9 | 6.40 |
| *Total* | *2,063* | *100* | *6.55* |
| **IP Experiment** | | | |
| Respondents Who Consented | 936 | 45.4 | 7.50 |
| Respondents Who Did Not Consent | 1,127 | 54.6 | 5.41 |
| *Total* | *2,063* | *100* | *6.55* |
| **Income** | | | |
| Provided Some Information | 1,926 | 93.4 | 6.71 |
| Did Not Provide Any Information | 137 | 6.6 | 4.34 |
| *Total* | *2,063* | *100* | *6.55* |
| **Health History** | | | |
| Answered All Questions | 1,969 | 95.4 | 6.60 |
| Did Not Answer All Questions | 94 | 4.6 | 5.49 |
| *Total* | *2,063* | *100* | *6.55* |
| **Alcohol and Marijuana Use** | | | |
| Answered Both Questions | 1,905 | 92.3 | 6.64 |
| Did Not Answer Both Questions | 158 | 7.7 | 5.47 |
| *Total* | *2,063* | *100* | *6.55* |

**Table 4. Sociodemographic Characteristics of Respondents***

| | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | **Low** | **Medium** | **High** |
| *% Total* | | *34.9* | *33.1* | *32.0* |
| **Sex** | | | | |
| % Male | 47.6 | 40.3 | 49.9 | 53.1 |
| % Female | 52.4 | 59.7 | 50.1 | 46.9 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Education** | | | | |
| % No high school degree | 11.8 | 14.2 | 10.1 | 11.0 |
| % High school degree | 56.2 | 57.2 | 56.1 | 55.1 |
| % College degree or more | 32.0 | 28.6 | 33.8 | 33.9 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Age** | | | | |
| % 18 to 34 | 29.2 | 25.3 | 33.8 | 28.7 |
| % 35 to 54 | 36.7 | 35.2 | 32.2 | 43.0 |
| % 55 to 64 | 16.0 | 17.6 | 16.1 | 14.0 |
| % 65+ | 18.1 | 21.8 | 17.9 | 14.3 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Race/Ethnicity** | | | | |
| % Non-Hispanic White | 62.4 | 61.8 | 64.6 | 60.7 |
| % Non-Hispanic Black | 12.8 | 12.7 | 11.9 | 14.0 |
| % Hispanic | 17.8 | 17.6 | 18.0 | 17.9 |
| % Other | 7.0 | 7.9 | 5.5 | 7.4 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Marital Status** | | | | |
| % Married or cohabiting | 50.7 | 49.5 | 48.1 | 54.6 |
| % Single | 49.3 | 50.5 | 51.9 | 45.4 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Area of residence** | | | | |
| % Rural area | 16.0 | 14.1 | 16.0 | 18.1 |
| % Small or mid-size town | 35.0 | 36.8 | 36.4 | 31.7 |
| % City or a suburb | 48.9 | 49.1 | 47.6 | 50.2 |
| *Total**** | *100* | *100* | *100* | *100* |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular characteristic*

****May be subject to round errors*

**Table 5. Additional Characteristics of Respondents***

| | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| *% Total* | | *34.9* | *33.1* | *32.0* |
| **Political views** | | | | |
| % Liberal | 34.6 | 29.7 | 37.0 | 37.6 |
| % Conservative | 28.7 | 30.3 | 28.3 | 27.3 |
| % Moderate | 36.6 | 39.9 | 34.7 | 35.1 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Party affiliation** | | | | |
| % Democrat | 51.4 | 46.4 | 50.7 | 57.5 |
| % Republican | 28.8 | 30.4 | 28.5 | 27.3 |
| % Other | 19.8 | 23.2 | 20.8 | 15.2 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Attendance of religious services** | | | | |
| % Never | 27.4 | 31.4 | 26.4 | 24.0 |
| % Seldom or a few times a year | 37.7 | 38.2 | 38.8 | 36.2 |
| % Once a month or more frequenty | 34.9 | 30.4 | 34.9 | 39.8 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Reported annual income from wages** | | | | |
| % Less than $25,000 | 41.8 | 41.5 | 38.5 | 45.3 |
| % Between $25,000 and $49,999 | 25.4 | 26.4 | 28.1 | 21.7 |
| % Between $50,000 and $74,999 | 17.3 | 17.5 | 17.0 | 17.4 |
| % Between $75,000 and $99,999 | 7.8 | 6.2 | 9.0 | 8.1 |
| % $100,000+ | 7.8 | 8.4 | 7.3 | 7.6 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Hours online per day** | | | | |
| % Less than 4 hours | 27.3 | 31.8 | 27.2 | 22.6 |
| % More than 4 and less than 9 hours | 51.9 | 48.7 | 53.3 | 53.8 |
| % 9+ hours | 20.8 | 19.5 | 19.5 | 23.6 |
| *Total**** | *100* | *100* | *100* | *100* |
| **Hours on social media per day** | | | | |
| % 0 hours | 23.3 | 30.4 | 20.2 | 18.7 |
| % More than 0 and up to 2 hours | 49.9 | 48.7 | 52.7 | 48.2 |
| % More than 2 hours | 26.8 | 20.9 | 27.2 | 33.1 |
| *Total**** | *100* | *100* | *100* | *100* |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular characteristic*

****May be subject to round errors*

**Table 6. Privacy Attitudes of Respondents***

| Attitudes | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| **% Respondents defining as particularly important***:** | | | | |
| Being in control of who can get information about them | 88.0 | 87.2 | 86.5 | 90.4 |
| Controlling exactly what information is collected about them | 85.0 | 85.0 | 81.3 | 88.7 |
| Having individuals in social and work situations not ask them things that are highly personal | 72.0 | 71.9 | 68.0 | 76.2 |
| Not being disturbed at home | 71.9 | 74.6 | 66.8 | 74.2 |
| **% Respondents considering the Decennial Census an invasion of privacy** | 31.5 | 38.7 | 27.5 | 28.4 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular attitude*

****Options "Extremely Important" and "Very Important" in the survey question*

**Table 7. Respondents' Effort on Activities Related to Safeguarding Personal Data***

| Activities | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| **% Respondents who very frequently***:** | | | | |
| Read privacy notices from banks, insurance companies, etc. | 43.4 | 39.3 | 40.5 | 50.9 |
| Change passwords on financial accounts | 46.4 | 44.8 | 41.8 | 52.8 |
| Check security features of a web site | 60.7 | 59.7 | 56.9 | 65.8 |
| Sign up for frequent buyer discounts | 37.1 | 27.5 | 36.7 | 47.9 |
| Shred financial documents | 69.4 | 70.9 | 65.9 | 71.3 |
| Buy a product online as opposed to some other way | 35.0 | 25.5 | 33.5 | 46.9 |
| Pay bills or managing financial accounts online as opposed to some other way | 55.5 | 46.7 | 57.1 | 63.4 |
| Managing calendar and appointments online as opposed to some other way | 31.0 | 19.8 | 32.1 | 42.2 |
| Handle matter with government agencies online as opposed to some other way | 24.7 | 16.4 | 24.7 | 33.7 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular level of effort*

****Options "Always" and "Most of the time" in the survey question*

**Table 8. Respondents' Behavior and Experiences Related to Personal Data***

| Behavior/Experience | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| % on do-not-call list | 57.9 | 62.6 | 56.5 | 54.3 |
| % with SSN on cards in wallet or purse | 32.2 | 27.9 | 31.5 | 37.4 |
| % with fraudulent purchase with one's information | 28.1 | 28.5 | 29.8 | 26.0 |
| % with stolen identity | 8.1 | 6.3 | 7.8 | 10.5 |
| % ever audited by IRS | 5.4 | 5.6 | 5.5 | 5.1 |
| % provided date of birth on social media account | 71.6 | 65.3 | 72.2 | 77.0 |
| % provided home address on social media account | 24.9 | 17.1 | 24.5 | 32.6 |
| % provided phone number on social media account | 39.8 | 32.2 | 39.7 | 46.9 |
| % provided names of family members on social media account | 31.9 | 27.3 | 31.1 | 36.9 |
| % feeling they do enough to protect privacy of personal information online | 36.4 | 32.7 | 38.1 | 38.7 |
| % filled and returned received U.S. Census form | 95.7 | 96.3 | 94.8 | 96.0 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular behavior/experience*

**Table 9. Respondents' Assessment of Information Type Sensitivity***

| Type of Information | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| **% Respondents defining as particularly sensitive information***:** | | | | |
| State of health and medicines taken | 59.6 | 64.9 | 54.6 | 59.0 |
| Information about finances | 79.8 | 84.7 | 78.0 | 76.4 |
| Political views and supported candidates | 27.9 | 29.1 | 22.8 | 31.9 |
| Social security number | 90.3 | 91.0 | 88.9 | 91.1 |
| Basic purchasing habits | 29.4 | 31.2 | 24.1 | 33.1 |
| Date of birth | 48.0 | 53.2 | 42.6 | 48.0 |
| Websites visited | 43.3 | 45.5 | 38.5 | 45.7 |
| IP address | 71.4 | 73.4 | 68.0 | 72.5 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within a given willing-to-consent group share a particular definition of sensitivity for that type of data*

****Options "Extremely sensitive" and "Very sensitive" in the survey question*

**Table 10. Willingness to Consent to Provide Medical and Earnings Record Linkages Across Socio-demographic Characteristics***

| Characteristic | Medical Records | | Earnings Records | | WTP diff. b/n Records?** |
|---|---|---|---|---|---|
| | N | Avg. WTP | N | Avg. WTP | |
| *All* | *1,001* | *6.32* | *854* | *6.50* | *No* |
| **Sex** | | | | | |
| Male | 477 | 6.56 | 406 | 6.75 | No |
| Female | 524 | 6.09 | 448 | 6.29 | No |
| **Education** | | | | | |
| No high school degree | 117 | 6.21 | 102 | 5.83 | No |
| High school degree | 549 | 6.23 | 493 | 6.69 | Yes |
| College degree or more | 335 | 6.50 | 259 | 6.42 | No |
| **Age** | | | | | |
| 18 to 34 | 285 | 6.56 | 257 | 6.86 | No |
| 35 to 54 | 379 | 6.66 | 302 | 6.59 | No |
| 55 to 64 | 154 | 5.56 | 142 | 6.60 | Yes |
| 65+ | 183 | 5.86 | 153 | 5.67 | No |
| **Race/Ethnicity** | | | | | |
| Non-Hispanic White | 650 | 6.32 | 507 | 6.42 | No |
| Non-Hispanic Black | 117 | 6.27 | 121 | 6.93 | No |
| Hispanic | 168 | 6.40 | 163 | 6.55 | No |
| Other | 66 | 6.14 | 63 | 6.27 | No |
| **Marital Status** | | | | | |
| Married or cohabiting | 512 | 6.54 | 428 | 6.42 | No |
| Single | 489 | 6.09 | 426 | 6.59 | Yes |
| **Area of residence** | | | | | |
| Rural area | 171 | 6.70 | 125 | 6.68 | No |
| Small or mid-size town | 350 | 6.29 | 297 | 6.34 | No |
| City or a suburb | 475 | 6.18 | 429 | 6.57 | No |

**Sample of respondents who received hypothetical study consent with linkage request*

***Difference between WTPs for respondents with that particular characteristic statistically significant at 5%*

**Table 11. Respondents' Reasons for Making A Consent Decision***

| Factor | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| **% Respondents defining as particularly important***:** | | | | |
| Value of study to scientific research | 54.8 | 37.9 | 53.8 | 74.4 |
| Understanding how the provided data will be used | 73.8 | 63.4 | 73.1 | 85.8 |
| Potential benefit of the study to the respondent and their family | 63.5 | 49.1 | 63.7 | 79.1 |
| Feeling a duty to participate | 34.3 | 18.1 | 31.3 | 55.1 |
| Reputation of institution conducting the research | 67.3 | 56.5 | 66.3 | 79.9 |
| Importance of being paid to participate | 64.6 | 51.0 | 65.6 | 78.4 |
| Having to talk about sensitive topics to researchers | 50.4 | 46.4 | 43.6 | 61.9 |
| Confidence in researchers to keep responses and information private | 79.5 | 74.5 | 76.7 | 87.9 |
| Time devoted to participating in the study | 63.4 | 52.2 | 61.4 | 77.7 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular reason for their consent decision*

****Options "Extremely Important" and "Very Important" in the survey question*

**Table 12. Trust Attitudes of Respondents***

| Attitudes | All Respondents | Willingness-to-consent Groups** | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| **% Respondents trusting*** the following institutions to keep information private and secure:** | | | | |
| Government | 49.8 | 39.9 | 51.4 | 59.0 |
| Polling and survey organizations | 63.1 | 55.7 | 62.1 | 72.2 |
| Financial institutions | 71.3 | 67.5 | 71.5 | 75.2 |
| Companies they buy things from | 64.3 | 55.8 | 64.7 | 73.2 |
| Political parties and candidates | 36.8 | 26.6 | 36.9 | 48.0 |
| Healthcare providers | 81.8 | 77.9 | 82.1 | 85.7 |
| IRS | 62.1 | 56.0 | 62.8 | 68.1 |
| Universities and research institutions | 77.0 | 70.7 | 77.6 | 83.1 |
| **Trust index** | | | | |
| Mean | 1.54 | 1.39 | 1.63 | 1.60 |
| Standard deviation | 1.18 | 1.17 | 1.17 | 1.18 |

**Sample of respondents who received hypothetical study consent with linkage request*

***Statistics in each column reflect what percentage of the respondents within the particular willing-to-consent group share a particular attitude, or the mean or the standard deviation for trust within that consent group*

****Options "Great deal" (of trust) and "Some" in the survey question*

**Table 13. Comparison of Average Willingness to Consent for Different Linkage Request Preambles***

| **Type of Linkage Permission** | **Linkage Request Preamble** | | | | |
|---|---|---|---|---|---|
| | **No Additional assurance (A)** | **Data Security Assurance (B)** | | **Emphasis on Research Benefits (C)** | |
| | **Avg. WTP** | **Avg. WTP** | **WTP diff. from (A)?**** | **Avg. WTP** | **WTP diff. from (A)?**** |
| **Administrative Records:** | | | | | |
| Both records | 6.47 | 6.54 | No | 6.19 | No |
| Medical records | 6.48 | 6.51 | No | 5.93 | Yes |
| Earnings records | 6.45 | 6.59 | No | 6.48 | No |
| **Full SSN for:** | | | | | |
| Both records | 2.98 | 3.45 | Yes | 3.14 | No |
| Medical records | 3.00 | 3.48 | No | 3.05 | No |
| Earnings records | 2.95 | 3.41 | No | 3.25 | No |
| **Last 4 digits of SSN plus date of birth for:** | | | | | |
| Both records | 4.38 | 4.58 | No | 4.32 | No |
| Medical records | 4.50 | 4.64 | No | 4.26 | No |
| Earnings records | 4.24 | 4.51 | No | 4.39 | No |

**Sample of respondents who received hypothetical study consent with linkage request*

***Difference between WTPs for the particular type of linkage permission statistically significant at 5%*

Assurances for data don’t seem to work with respect to the reporting of identifying information.

**Table 14. Willingness to Consent to Provide Full SSN in a Signed Form for Record Linkages Across Socio-demographic Characteristics***

| Characteristic | Asked First | | Asked After Last 4 Digits of SSN and Date of Birth | | WTP diff. b/n asked first or second?** |
|---|---|---|---|---|---|
| | N | Avg. WTP | N | Avg. WTP | |
| *All* | *906* | *3.65* | *949* | *2.75* | *Yes* |
| **Sex** | | | | | |
| Male | 429 | 4.16 | 454 | 3.10 | Yes |
| Female | 477 | 3.19 | 495 | 2.44 | Yes |
| **Education** | | | | | |
| No high school degree | 103 | 3.93 | 116 | 2.85 | Yes |
| High school degree | 514 | 3.46 | 528 | 2.67 | Yes |
| College degree or more | 289 | 3.89 | 305 | 2.86 | Yes |
| **Age** | | | | | |
| 18 to 34 | 255 | 4.45 | 287 | 3.49 | Yes |
| 35 to 54 | 342 | 4.18 | 339 | 2.98 | Yes |
| 55 to 64 | 144 | 2.59 | 152 | 2.01 | No |
| 65+ | 165 | 2.22 | 171 | 1.74 | No |
| **Race/Ethnicity** | | | | | |
| Non-Hispanic White | 563 | 3.60 | 594 | 2.55 | Yes |
| Non-Hispanic Black | 119 | 3.86 | 119 | 3.54 | No |
| Hispanic | 159 | 3.57 | 172 | 2.73 | Yes |
| Other | 65 | 3.91 | 64 | 3.25 | No |
| **Marital Status** | | | | | |
| Married or cohabiting | 454 | 3.78 | 486 | 2.93 | Yes |
| Single | 452 | 3.52 | 463 | 2.57 | Yes |
| **Area of residence** | | | | | |
| Rural area | 143 | 3.48 | 153 | 3.02 | No |
| Small or mid-size town | 305 | 3.50 | 342 | 2.77 | Yes |
| City or a suburb | 454 | 3.79 | 450 | 2.67 | Yes |

**Sample of respondents who received hypothetical study consent with linkage request*

***Difference between WTPs for respondents with that particular characteristic statistically significant at 5%*

**Table 15. Willingness to Consent to Provide Last 4 Digits of SSN and Date of Birth for Record Linkages Across Socio-demographic Characteristics***

| Characteristic | Asked First | | Asked After Full SSN | | WTP diff. b/n asked first or second?** |
|---|---|---|---|---|---|
| | N | Avg. WTP | N | Avg. WTP | |
| *All* | *949* | *4.11* | *906* | *4.77* | *Yes* |
| **Sex** | | | | | |
| Male | 454 | 4.63 | 430 | 5.17 | Yes |
| Female | 495 | 3.62 | 476 | 4.41 | Yes |
| **Education** | | | | | |
| No high school degree | 116 | 3.70 | 103 | 5.09 | Yes |
| High school degree | 528 | 4.13 | 514 | 4.61 | Yes |
| College degree or more | 305 | 4.21 | 289 | 4.93 | Yes |
| **Age** | | | | | |
| 18 to 34 | 287 | 4.22 | 255 | 5.00 | Yes |
| 35 to 54 | 339 | 4.48 | 343 | 5.27 | Yes |
| 55 to 64 | 152 | 3.67 | 143 | 4.06 | No |
| 65+ | 171 | 3.56 | 165 | 3.98 | No |
| **Race/Ethnicity** | | | | | |
| Non-Hispanic White | 594 | 4.12 | 563 | 4.87 | Yes |
| Non-Hispanic Black | 119 | 4.42 | 119 | 4.58 | No |
| Hispanic | 172 | 3.72 | 159 | 4.64 | Yes |
| Other | 64 | 4.44 | 65 | 4.57 | No |
| **Marital Status** | | | | | |
| Married or cohabiting | 486 | 4.27 | 455 | 4.96 | Yes |
| Single | 463 | 3.93 | 451 | 4.57 | Yes |
| **Area of residence** | | | | | |
| Rural area | 153 | 4.35 | 142 | 4.86 | No |
| Small or mid-size town | 342 | 4.15 | 305 | 4.78 | Yes |
| City or a suburb | 450 | 3.99 | 455 | 4.74 | Yes |

**Sample of respondents who received hypothetical study consent with linkage request*

***Difference between WTPs for respondents with that particular characteristic statistically significant at 5%*