

## Development and Testing of Informed Consent Questions to Link Survey Data With Administrative Records

Nancy Bates  
U.S. Census Bureau

### ABSTRACT<sup>1</sup>

In this paper we report results from qualitative and quantitative testing of an informed consent question. The question asks permission to perform data linkages between demographic survey data and administrative records. The new consent question was first tested with cognitive interviews followed by a larger field test in a random-digit-dial survey. In the field test, the new question was tested against the traditional means of informed consent to perform data linkage – a request for a social security number (SSN).

Keywords: informed consent, cognitive testing, RDD Surveys, SSN

### 1. Background

Administrative records are routinely matched to survey data as a way to enhance the survey data for methodological, policy, and other statistical research purposes. Historically, the U.S. Census Bureau has used Social Security Number (SSN) as a means of linking demographic survey data with administrative records. If the respondent provides his/her SSN during the interview, this is viewed as implied consent to perform data linkages.

Census Bureau efforts to obtain SSNs have become increasingly difficult over the last few years. For example, in the Survey of Income and Program Participation (SIPP), SSN refusals increased from 12 percent to 35 percent between the 1996 and 2004 panels, respectively (Riccini, 2003; Lewis, 2005). The percent refusing to provide SSN in the Current Population Survey (CPS) increased from approximately 10 percent in 1994 to almost 23 percent by 2003 (Tucker, 1999; Marshall, 2004). Obviously, the public is becoming less willing to provide SSNs. According to a debriefing of Census Bureau field staff, the top three reasons for refusing to provide SSN include: fear of identify theft, personal policy not to disclose SSN, and media messages discouraging practice of giving out SSN (Lewis, 2005).

In response to survey respondents' increasing reluctance to divulge SSNs, agencies have begun to explore alternatives. For example, the National Center for Health Statistics (NCHS) recently performed linkage tests to explore what other criteria could be used in place of a full nine-digit SSN to select records for linkage to the National Death Index (Sayer and Cox, 2003). The theory is that respondents may be more willing to provide a subset of SSN digits (the last four), because people are accustomed to seeing it on credit card receipts, as bank

password codes, and the like. The 4-digit number could then be used in place of the entire SSN to perform record search and selection. However, while the agency has performed research to explore the technical processes of linking, it has not performed empirical tests to explore whether survey respondents are more forthcoming when only a portion of SSNs is requested.

The Census Bureau has also established a research program to investigate alternatives. The Census Bureau developed and tested a new record linkage search methodology that does not require collection of SSNs (Killion, 2002). An evaluation of this method indicates the system yields accurate and satisfactory match rates compared to the traditional SSA validation processes (Roemer and Stinson, 2003). This has fueled interest in permanently eliminating the collection of SSNs in the SIPP and other surveys. At the same time, however, elimination of SSN would effectively eliminate the implied consent to link data. Consequently, in 2002 the Census Bureau began to conduct research to explore new ways to request consent to link data without collecting SSN. In this paper we report on cognitive research to develop a new informed consent question followed by results from a random-digit-dial field test that tests the new question.

The primary objective was to assess whether a new question could replace the request for SSN in current Census Bureau surveys. One critical criterion for success is whether a new question will yield a larger proportion of respondents granting consent to data linkage compared to the traditional request for SSN.

Research in this area consisted of two phases: (1) a qualitative question development and testing phase and (2) a quantitative field-test phase. Phase 1 consisted of cognitive interview research to develop the best wording for a new consent question. (See Landreth, 2002; Bates 2003). Phase 2 consisted of a random-digit-dial (RDD) field test to measure response to the new informed consent question compared to a control (the request for SSN).

### 2. Phase 1 - Cognitive Development of New Consent Question

#### 2.1 Methodology

Four versions of a new informed consent question were tested using the concurrent think-aloud cognitive interview technique (Willis, 2005). The four versions tested are detailed below:

##### *Version A (Opt-in)*

“The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits.

<sup>1</sup> DISCLAIMER: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are the author's and not necessarily those of the U.S. Census Bureau.

Would that be okay with you?"

- Yes  
 No  
 Don't Know

*Version B (Opt-in with confidentiality sentence)*

Version B was identical to A with one extra sentence at the end:

"All of the additional information collected by the Census Bureau is kept confidential and is protected by the same law that protects your survey information."

*Version C (Opt-out)*

"The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits.

Do you have any objections?"

- Yes, object  
 No, do not object  
 Don't Know

*Version D (Opt-out with confidentiality sentence)*

Version D was identical to C with one extra sentence at the end:

"All of the additional information collected by the Census Bureau is kept confidential and is protected by the same law that protects your survey information."

Census Bureau staff conducted twenty cognitive interviews in July-September 2003. Question version was assigned prior to each interview, so that each version was tested in five interviews. The informed consent questions were tested as part of a larger cognitive interview that contained a household roster with demographics and questions about health insurance and coverage. The informed consent question was placed at the end of the cognitive interview.

Participants were encouraged to "think aloud" and verbalize their thought process as they attempted to answer questions. To facilitate this process, each interviewer used a standardized protocol that contained the experimental questions immediately followed by a list of probes and follow-up questions.

The participant demographics were wide-ranging, with the exception of gender (16 of the 20 interviews were with women). Ages ranged from 18 to 81, and education levels from 12<sup>th</sup> grade to a Master's degree. Eight respondents self-reported race as black, eight as white, three as biracial, and one as something else.

## 2.2 Cognitive Interview Results

To evaluate the different versions, we established two criteria and then assessed how each version performed. The first criterion was: Did the respondent grant permission to the informed consent request? For Versions A–B, this was indicated by an answer of "Yes." For Versions C–D, it was indicated by an answer of "No, do not object."

The vast majority of participants granted permission in response to the informed consent question. Only two respondents declined (one for Version A and one for Version B). Both versions where respondents declined were the opt-in versions, one with the extra confidentiality clause and the other without. Declining consent was rare and did not appear concentrated in a particular version. Thus, using only affirmative answers as criteria, the cognitive interviews suggest little difference between the four versions and, when asked directly, that most people are inclined to allow Census Bureau access to other agencies data.

When asked why they responded the way they did, respondents who granted consent said things like these:

- "I have nothing to hide."*  
*"Information is protected by law."*  
*"It's confidential."*  
*"Would be the correct thing to do."*  
*"Me authorizing to release information to Census."*

In the two cases where respondents declined consent, these were the verbatim reasons:

- "I would want to know what information is being given and for what purpose before I would just say yes."* (White female, age 38)  
*"Somebody may get hold of it but if anything they needed to know they could just ask me...anything personal about me, I'd like to know it for sure."* (Black female, age 72)

The second criterion was whether respondents understood the intent and purpose of the informed consent question. We determined comprehension by listening to verbatim transcripts of the interviews at the point where respondents were asked to paraphrase the informed consent question "in their own words." The evaluation of answers is subjective to a certain degree, but we divided respondents into two categories according to two dimensions. The first dimension was purpose: Did the respondent understand that the Census Bureau was requesting informed consent to obtain and analyze additional information they had previously provided other agencies? The second dimension was process: Did the respondent grasp the direction of the data linkage request; that is, that the Census Bureau was asking permission to receive information from other agencies, not give the current survey data to other agencies? This difference is subtle, and the concept is often lost on the general public when trying to measure attitudes about data linkage and data sharing (Bates 1995).

After deconstructing answers into these more simple dichotomies, the results were somewhat surprising. Based on the transcripts, we

estimate that 9 of the 20 respondents failed to understand the basic purpose of the request and another 11 most likely misunderstood the logistics of the administrative record linking procedure. These misunderstandings were not concentrated in one version or another; quite the contrary, they were evenly spread out among the four versions.

Those who correctly understood the purpose paraphrased the question along these lines:

*“You can go and find, if I’ve ever reported anything—like you can go to my Social Security records and find out how much I’ve been working.”*

*“See what type of Medicaid I have, using my Social Security Number.”*

*“You are going to get my Social Security Number and you’re going to access some government documents that might show any kind of benefit I’ve had.”*

It is noteworthy that several respondents mentioned Social Security Number, even though the question itself does not. Apparently, some respondents automatically assumed that SSNs play a part in data linkage, even when it is not explicitly requested as part of the informed consent process. When asked why they mentioned SSN, several respondents noted that the Social Security Administration was mentioned as a government agency example in the question and this prompted their thinking about SSNs.

Respondents who misunderstood the purpose of the consent seemed confused for different reasons. A few fixated on the confidentiality clause while others seemed to confuse the purpose of the specific informed consent question with the more general purpose of consent to use cognitive interview results as a tool to improve survey questions. Still others misunderstood because they thought the Census Bureau was requesting permission to share their answers with other data agencies.

Below are selected examples of paraphrases where respondents expressed misunderstanding of purpose:

*“...if I would have any objection to you asking me some questions that may be beneficial to your surveys but you made me aware of the fact that it will be kept confidential.”*

*“...giving the authorization to release information to Census and understanding that it won’t be provided to other agencies or be used for purposes other than survey data...”*

*“...it’s asking me if... it’s okay to give the information on our policy to other government offices...”*

*“...So you’re asking if you can give this information to ...you’re gonna take my answers and give it to somebody else...”*

Obviously, some of the misunderstanding involves the subtle difference of the direction of data sharing. The informed consent questions tested here were written explicitly with one-way data

sharing in mind; that is, linking information from the current study to information previously provided to other government agencies, in order to enhance and supplement the current collection. It does not seek permission to share the current data collection information with other agencies.

The difficulty of relaying such a subtle and abstract concept is obvious. On one hand, the consent must fulfill the ethical and legal requirements of making respondents aware of plans to use administrative records. On the other hand, it must be careful not to raise unwarranted suspicion or fear about the requested activity. This is a tall order. The public at large is generally not familiar with administrative record research and tends to be suspicious and wary of large personal databases sometimes thought of as “the Big Computer” (Gerber, 2001).

We found no evidence that a particular version was more inclined to yield a negative response to the consent request or misunderstanding of the question meaning. However, we were curious whether the versions with the added confidentiality statement yielded any positive (or negative) feedback. In the ten interviews that tested versions with this clause (Versions B and D), four respondents made some mention of the confidentiality protection in a positive way, either during the paraphrase exercise or as part of the follow-up probes. However, as also noted above, several respondents who subsequently misunderstood the request may have done so because they focused on the latter part of the question (the confidentiality part) and not on the request itself. Additionally, three respondents made a comment that the versions with the confidentiality clause struck them as particularly long questions.

Based on the 20 interviews conducted, there was no clear winner or loser to the four versions tested. With rare exceptions, all versions yielded the desired outcome, that is, acceptance to the informed consent request for Census to obtain additional information. While misunderstanding of the request was somewhat common, neither version appeared more or less likely to cause misunderstanding of the purpose or the process of the request.

Because the versions performed so similarly, we decided not to test both an opt-in and opt-out version during the field test phase. If the cognitive interviews were any indication, both would perform equally well and the test panels could be more efficiently allocated by testing another design. Ultimately, the decision was made to select the opt-out version for the next phase of testing. Because the versions without the confidentiality statement appeared to perform as well as those with it, we decided to use the version without it. This kept the question shorter, which was deemed important in a telephone survey. However, we did keep the additional sentence—along with other verbiage restating the purpose of the request and the protection authority (Title 13)—in an interviewer help screen to be read immediately after the initial request as needed.

As a result of the interviews, we did not recommend specific changes to the basic wording of the opt-out version. While it’s true that a nontrivial number of respondents failed to paraphrase the exact intent of the request when asked, the causes of the misunderstandings were

not made clear in the interviews, so as to suggest alternative wording.

The proposed wording for the new informed consent question in Phase 2 field test was:

“The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits.

Do you have any objections?”

- Yes, object
- No, do not object
- Don't Know

**3. Phase 2 – Field Test of New Consent Question**

**3.1 Methodology**

The field test was carried out during the 2004 Questionnaire Design and Evaluation Research Survey (QDERS)—a small, internally-sponsored, Census Bureau survey (random-digit-dial survey). The QDERS was conducted at two Census Bureau telephone centers (Jeffersonville, Indiana and Hagerstown, Maryland). The interview was a household RDD survey conducted between April 16 and June 14, 2004. The instrument was a computer-assisted telephone interview (CATI) consisting of four treatments (i.e., questionnaire versions). Each survey consisted of questions about the basic demographic composition of household members, the co-habitation status of members, detailed questions about health insurance coverage, and a consent question about linking survey answers to other government records.

Telephone numbers screened as businesses or group quarters were ineligible and dropped from the survey. Cases determined to be private residences were screened for a household member 18 or older who could answer basic demographic and health insurance questions about the household members. Each survey treatment took approximately 10 minutes to administer.

Interviewers were divided into four groups – 10 per treatment with 40 interviewers in total. The survey was divided into four interview periods such that all four treatments were administered during each interview period and each interviewer administered all four questionnaire versions.

**3.2 Questionnaire Treatments**

The QDERS experiment was comprised of four different questionnaire treatments, one of which was the new informed consent question developed during the cognitive interview test phase.

*Treatment A – SSN request with mention in introduction*

The informed consent manipulations were placed at the end of each survey. Treatment A served as the control and was designed to try

and emulate the procedures used to obtain SSN in current demographic surveys conducted at the Census Bureau (and in particular the Survey of Income and Program Participation - SIPP). In the case of most Census Bureau demographic surveys, advance letters are mailed to households, making them aware they have been selected, and a Census representative will visit. The letters are also used to inform households that SSN will be collected, and that doing so allows the Census Bureau to gather additional data from other government agencies.

Since the QDERS was an RDD survey, addresses were not readily available and an advance letter was not part of the implementation procedures. To try and emulate an advance letter, the survey introduction in Treatment A mentioned informed consent up front, then asked for SSN at the end of the survey without any explanation as part of the question text. (If needed, interviewers had a help screen they could access at the SSN question that further explained the rationale for the request).

*Introduction:*

“The survey will take about 10 minutes and is voluntary. The Office of Management and Budget has approved this survey under project 0607-0725. The Privacy Act of 1974 requires us to keep your information confidential and use it for statistical purposes only. At the end of this survey, we will ask for your consent to obtain additional information you have provided other government agencies.”

*Question text:*

“To finish up, I have one last question. What is your Social Security Number?”

\_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_

- Don't Know
- Refused

*Treatment B – SSN request with explanation in request*

Treatment B also requested SSN at the end of the survey but attempted to “soften the blow” a bit by providing some rationale for the request. Because it was already part of the question wording, Treatment B did not mention informed consent as part of the survey introduction.

*Question text:*

“To finish up, I have one last question. The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits. To do this, we need your Social Security Number. May I please have your Social Security Number?”

- Yes, \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_
- No (refused)
- Don't Know

*Treatment C– Request for last four digits of SSN*

Treatment C requested the last 4 digits of the SSN. This panel was designed to measure willingness to provide a truncated SSN.

*Question text:*

“To finish up, I have one last question. The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits. To do this, we need your Social Security Number. To protect the privacy of your Social Security Number, we are only asking for the last 4 digits. May I please have your last four digits [of your Social Security Number]?”

- Yes, \_\_\_ \_\_\_ \_\_\_ \_\_\_  
 No (refused)  
 Don't Know

*Treatment D - New informed consent question*

Treatment D tested the new informed consent question produced from the qualitative testing in Phase I. This version asked the respondent if they objected to the Census Bureau getting information from other agencies for research purposes. The question did not ask for SSN.

*Question text:*

“The Census Bureau would like to conduct additional research without taking up your time with more questions. We would like your permission to obtain the information that you have given to other government agencies on topics such as Social Security and Medicare benefits.

“Do you have any objections?”

- Yes, object  
 No, do not object  
 Don't Know

A random quarter of the RDD sample was delegated to each treatment. The major focus of analysis was a comparison between the percent who refused to supply their 9-digit SSN, refused the last four digits of their SSN, or who opted-out at the new consent question. The null hypothesis was that there would be no detectable differences between the four treatments.

**3.3 Field Test Results**

At the survey closeout, a total of 4,317 completed interviews were conducted (over 1,000 per treatment) with a final response rate between 42.4 – 59.2 percent, depending upon the calculation.<sup>2</sup> The

<sup>2</sup> 42.4% represents a minimum response rate calculated in accordance with AAPOR's RR2 definition. The maximum response rate according to AAPOR's RR6 is 59.2% (AAPOR,

SSN and informed consent questions were asked only of the household respondent.

Descriptive statistics of the household respondents indicate they were likely to be female, between the ages of 41–65, to have attended some college, and be from a household above a “low income” threshold<sup>3</sup>. Additionally, the overwhelming majority of household respondents were white and non-Hispanic. Obviously, these characteristics differ in many respects to the overall U.S. population and therefore limit the generalizability of the data. However, we found no significant differences in demographic characteristics between the four treatments. Thus, while the absolute level of consent granted in QDERS may be different from the overall population, the survey is still a valid assessment of the relative difference in refusal levels between treatments. The QDERS data does not contain any sample, post data-collection, or nonresponse weights, therefore the analyses that follow are based upon raw, unweighted data.

Table 1 displays the outcomes to the requests for SSN and the new informed consent question across treatments. In Treatment D, the “refused” row includes those who objected to the informed consent request (N=394) and those who refused to answer the question (N=7).

As Table 1 indicates, the difference in outcomes between treatments is striking. The two treatments that requested the entire SSN fared the worst in terms of obtaining consent to perform data linkage. Approximately sixty percent refused to provide SSNs in Treatment A while close to three-quarters (74 percent) refused in Treatment B. Recall that the difference between treatments was twofold: in Treatment A the survey introduction mentioned that consent would be asked at the end of the survey and then requested SSN without further explanation to conclude the interview. Treatment B did not mention consent in the survey introduction but did explain the rationale for collecting SSN in some detail as part of the question itself.

The intent of Treatment B's added explanation was to encourage SSN reporting by “softening the blow”. We hypothesized that respondents would be more inclined to supply SSNs in this context compared to Treatment A, where the request is very straightforward with no real connection to the rest of the survey. However, Treatment B's approach appears to have worked in quite the opposite direction with more respondents refusing to supply SSNs. This is somewhat in line with previous research on confidentiality assurances (Singer, Hippler and Schwarz, 1992; Singer, Von Thurn and Miller, 1995). This research suggests that a potential backlash may result if too much attention is drawn to the act of gaining consent – respondents become unintentionally sensitized to the request and become less willing to comply.

Asking for a truncated SSN in Treatment C met with greater success. Roughly one half of household respondents agreed to provide a four-digit number while slightly less than half (around 48 percent) refused. This refusal rate was significantly lower than either treatment

2004).

<sup>3</sup> The threshold was designed to act as a general indicator of households eligible for government assistance programs.

requesting the entire nine-digit SSN (A versus B  $X^2=30.8$ , d.f.=1,  $p<.001$ ; B versus C  $X^2=158.7$ , d.f.=1,  $p<.001$ ).

By far, the survey treatment that avoided asking for any part of an SSN (Treatment D) had the lowest refusal rate (approximately 37 percent). This was comprised primarily of respondents who explicitly objected to the idea of record linkage (N=394) with a small number (N=7) who refused to answer the question altogether. Over three-fifths in Treatment D granted permission for the Census Bureau to obtain additional information from other government agencies.

### 3.4 Treatment differences among subgroups

To further examine differences among the treatment groups, we broke out the percent refusing to provide SSN, last digit of SSN, or consent to record linkage, by various demographic subgroups. Similar subgroupings were used to analyze the Phase 1 cognitive testing of the new informed consent question.

Across different treatment groups, females were more likely to refuse the information needed to perform data record linkages (see table 2). The male/female differential is most obvious in the survey version that requested SSNs with explanation (treatment B) or the last four digits of SSN (treatment C). In both of these cases, females were especially reluctant to provide the information request compared to males.

Education appears to be positively correlated with refusal to supply information about SSNs or grant informed consent. Those with less than a high school education were more willing to provide SSNs or consent to link while respondents with a high school diploma and above were less willing. The differential between education groups appears less obvious when last four digits or informed consent to link is asked in place of full SSN (treatments C and D). A multivariate analysis would have to be conducted, however, to confirm whether a significant interaction is truly present.

Respondents in the youngest category (age 18-30) were generally less likely to refuse the request. Interestingly, this was true even in treatment B which had the highest refusal rate over all groups. Young respondents had the lowest refusal rates of any subgroup to treatments C and D. Perhaps this group is more accustomed to using the last four digits of SSN as a personal identifier and/or is less concerned with granting permission to perform data linkage in the absence of a formal request for SSN. Only 21 percent in this age group refused the informed consent request to get additional information from other government agencies.

The household income measure was divided into two categories: households with income below a proxy indicator for government assistance eligibility, and those with household incomes above such a threshold. Additionally, a category indicating those who refused to answer the income question was also examined. Previous research indicates an association between reluctance to answer income-related questions and willingness to participate in surveys. That is, persons who are reluctant to share income information are also more likely to refuse surveys altogether and are also more likely to attrit surveys if

they participate initially (Bates and Creighton, 2001). With this in mind, we hypothesized that persons refusing to answer income-related questions might also be more likely to refuse the request for SSN/informed consent.

Table 2 confirms this hypothesis – the overwhelming majority of those who refused the income question also refused the request for SSN (100 percent and 96 percent refused in treatments A and B, respectively), 93 percent refused to provide last four digits of SSN, and 87 percent refused to provide permission to match records. These refusal rates are far above any other subgroup examined. In regards to the “below” or “above” income threshold, there is a general trend for households above the threshold to refuse slightly more often across all treatments but there does not appear to be an interaction between income level, refusal rate, and survey treatment. Again, to confirm absence of an interaction, a multivariate analysis must be performed.

The distributions of refusal rates by treatment among Hispanic/non-Hispanic origin indicate no significant differences. That is, these two groups followed a similar trend across treatments, with treatment B having the highest refusal rate followed by treatment A, C and then D. The race breakouts also indicated no significant differences to suggest a race by treatment interaction – this is partly driven by the small sample size of certain groups, e.g., American Indians. With the exception of Blacks, the refusal rate for every subgroup examined was lowest in Treatment D.

## 4. Summary and Conclusions

In this paper, we present research results from qualitative and quantitative testing of new informed consent questions. The consent questions seek respondent permission to link survey data with administrative records.

Cognitive interviewing was used to test and select wording of a new opt-out informed consent question that did not include a request for SSN. The qualitative testing revealed that most respondents consented to the request, but that a significant portion did not completely understand some of the subtleties of the request.

The new opt-out question was then field tested in an RDD survey against different versions of questions designed to obtain information necessary to perform record linkages. The four treatments yielded very different results. In this report, we concentrated on refusal rates because those who refuse to provide SSN (or other informed consent) must be removed from the pool that undergoes a search and verification operation necessary to match survey data to administrative records. Thus, refusals represent that part of the survey population for whom data linkage cannot even be attempted. In the last decade, this population has increased dramatically as refusal to provide SSN has also increased.

Close to 60 percent of household respondents refused to provide an SSN when asked in the traditional manner – the refusal rates increased to almost three-quarters (74 percent) when an explanation for the request was added as part of the question itself. This seems to

confirm previous confidentiality assurance research that suggests a fine line between providing too much information and assuring respondents their data will be safeguarded. If the assurance is too elaborate, it may raise suspicions and unintentionally heighten sensitivity, which ultimately lowers consent rates.

The alternative approaches – asking only last four digits of SSN or asking directly about objections to data linkage (new opt-out question) had overall lower refusal rates of 48 percent and 37 percent, respectively. Compared to the traditional procedure, the new informed consent question suggests great potential for reversing what has been a negative trend in administrative record research.

After examining subgroup differences across the four treatments, several conclusions are made. First, women are more likely to refuse information needed for data linkage compared to men. Second, young people (age 18-30) were less likely to refuse, especially when asked for last four digits of SSN or the informed consent question. Third, those with a high school education or higher were more likely to refuse than those without, but this difference becomes less evident in the treatments that asked last four digits or informed consent without a request for SSN.

Results in this paper begin to fill a research gap surrounding the question of the how best to gain informed consent for data linkage and the public's willingness to grant that consent. While we have learned much from this experiment, the results must be taken in the proper context. The results come from an RDD telephone survey of one respondent per household. As a result, we must be careful not to generalize our findings too far. On the other hand, the experiment in some ways can be viewed as a "worst case" scenario. Very few Census Bureau demographic surveys are conducted by RDD, without advance notice, and without at least one personal contact. Consequently, the refusal rates to record linkage reported here are probably on the high end compared to what might result in a production survey.

Of course, whether or not the Census Bureau decides to move toward a new informed consent question and do away with collection of SSN is ultimately a policy decision that will have to be vetted and formalized by the agency. If the decision is made to eliminate SSN, the agency must decide on a new approach to obtain consent for data linkage or, at a minimum, notify respondents of intent to link. The research in this paper should be considered when making an informed decision in that regard.

*Acknowledgements:* The author thanks Pat Doyle, Carolyn Puckett, Jennifer Rothgeb, Joanne Pascale and Jennifer Hunter for financial support and logistical assistance in implementing the 2004 QDERS. Finally, thanks to Theresa DeMaio for helpful comments on an earlier version of this paper.

## 5. References

- Bates, N. (2003). "Cognitive Testing of Informed Consent Questions: Summary Report from Phase 1 2004 Questionnaire Design and Evaluation Research Survey (QDERS)." Internal Census Bureau Report, October 9.
- Bates, N. and Creighton, K. (2001). "The Last Five Percent: What Can We Learn from Difficult/Late Interviews?" Proceedings of the Government Statistics Section of the American Statistical Association, pp. 120-125.
- Killion, R.A. (2002). "Development of a Production Capability for Social Security Number (SSN) Verification at the Census Bureau." PRED Memorandum Series #58. U.S. Census Bureau, 28 May.
- Landreth, A. (2002). "Report of Cognitive Interview Results from the Alternative Administrative Record Use Paragraph of the SIPP Advance Letter." Internal memorandum, U.S. Census Bureau, 1 March.
- Lewis, D. (2005). "Final Project Report for Improving Social Security Number (SSN) Response Study." Report from U.S. Census Bureau for Social Security Administration funded under Interagency Agreement 2003-00002531. January 27, 2005.
- Marshall, T. (2004). "Recent Current Population Surveys (CPS) Social Security Number (SSN) Reporting Experience." Internal Census Bureau table prepared by T. Marshall, February 2, 2004.
- Ricinni, D. (2003). Personal email correspondence to the author from Donna Ricinni, Chief, Income Surveys Programming Branch, Demographic Surveys Division, U.S. Census Bureau. January 23, 2003.
- Roemer, M. and Stinson, M. (2003). "Review of the Social Security Number Verification and Search Process (PVS) of the Planning, Research, and Evaluation Division". A report conducted in part by NSF Grant SES-9978093 to Cornell University, the National Institute on Aging and the Alfred P. Sloan Foundation, 21 March.
- Sayer, B. and Cox, C. (2003). "How Many Digits in a Handshake? The Role of SSN in NDI Record Linkage." A paper presented at the American Statistical Association Joint Statistical Meetings, 3-7 August, San Francisco, California.
- Singer, E., Hippler, H., and Schwarz, N. (1992). Confidentiality assurances in surveys: Reassurance or threat? *Public Opinion Research*, 4, 256-268.
- Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly*, 59, 66-77.
- Tucker, R. (1999). "Recent Current Population Surveys (CPS) and Social Security Number (SSN) Reporting Experience." Memorandum for the record. U.S. Census Bureau, 8 December.
- Willis, G. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, California: Sage Publications.

**Table 1. Outcome of requests by treatment**

	<i>Question Treatment</i>			
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
	<i>SSN</i>	<i>SSN+expl. Last 4 SSN</i>	<i>Last 4 SSN</i>	<i>Opt-out</i>
<i>Provided SSN, last 4 digits, or granted permission</i>	36.8%	24.0%	50.6%	63.4%
<i>'Don't Know'</i>	3.7%	2.0%	1.9%	0.1%
<i>'Refused' or denied permission</i>	59.5%	74.0%	47.5%	36.6%
<b>(N)</b>	<b>1049</b>	<b>1066</b>	<b>1098</b>	<b>1097</b>

$X^2 = 405.5$ , d.f.=6,  $p < .0001$

**Table 2. Percent refusing request by selected subgroups and treatment**

	<i>% Refusing Request</i>				<b>(N)</b>
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	
	<i>Question Treatment</i>				
<b>SEX:</b>					
<i>Males</i>	58.3	68.7	42.7	32.9	<b>(1619)</b>
<i>Females</i>	60.0	77.1	50.4	38.7	<b>(2683)</b>
<b>EDUCATION:</b>					
<i>Less than High School</i>	37.7	66.3	44.9	34.7	<b>(344)</b>
<i>High School</i>	58.2	73.3	50.0	36.8	<b>(1146)</b>
<i>Some College/Assoc. Degree</i>	60.6	75.3	46.4	32.6	<b>(1216)</b>
<i>B.A. and above</i>	63.0	74.7	46.4	39.0	<b>(1559)</b>
<b>AGE:</b>					
<i>18-30</i>	56.8	63.1	36.3	20.6	<b>(500)</b>
<i>31-40</i>	57.9	73.7	47.3	35.5	<b>(829)</b>
<i>41-50</i>	65.5	71.6	46.0	36.7	<b>(958)</b>
<i>51-65</i>	54.8	81.0	47.2	37.5	<b>(1118)</b>
<i>65+</i>	57.7	73.7	53.2	39.5	<b>(819)</b>
<b>HOUSEHOLD INCOME<sup>1</sup>:</b>					
<i>Below 'low' threshold</i>	53.9	70.8	44.4	31.4	<b>(1295)</b>
<i>Above 'low' threshold</i>	60.8	74.4	47.9	36.1	<b>(2862)</b>
<i>Refused income question</i>	100.0	95.8	93.3	86.7	<b>(92)</b>
<b>ORIGIN:</b>					
<i>Hispanic Origin</i>	57.1	71.2	55.9	32.1	<b>(220)</b>
<i>Not Hispanic</i>	59.0	74.1	46.8	36.5	<b>(4057)</b>
<b>RACE:</b>					
<i>White</i>	58.1	73.4	47.5	35.4	<b>(3591)</b>
<i>Black</i>	69.3	77.6	36.8	40.0	<b>(382)</b>
<i>American Indian/Alaska Native</i>	54.2	65.0	64.0	39.1	<b>(92)</b>
<i>Hawaiian/Asian/Pacific Islander</i>	63.6	79.4	42.4	38.5	<b>(139)</b>

<sup>1</sup>Household income divided into those above and below a proxy indicator for government assistance eligibility.