

THE EFFECT OF ADMINISTRATIVE RECORD USE NOTIFICATION ON SOCIAL SECURITY NUMBER REPORTS

Courtney N. Stapleton, US Census Bureau

KEY WORDS: item nonresponse, social security number, administrative records

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

1. Background

Administrative record use and data sharing among federal agencies have been proposed as ways to deal with the rising costs of the decennial census. The use of administrative records also has the potential to reduce respondent burden by means of a shorter census questionnaire. Improvements in data quality may also emerge by eliminating or reducing respondent memory errors. Decennial censuses beginning in 2010 may rely on expanded use of administrative records information obtained from other Federal agencies. At the time of this study, it was thought that effective use of these administrative records would require the addition of a question asking for Social Security Number (SSN) (for all household members) to the census form¹.

This paper studies the effects of two strategies for obtaining SSN information and differently worded notifications of administrative record use on response to the SSN item when requested in a census environment. SSN requests for the first person listed on the form (Person 1) and all household members, along with two alternate notifications of administrative record use are tested. This paper also studies whether the effects differ across subpopulations.

1.1 Past Research

Past research suggests that respondent uneasiness about the collection of personally identifying information reflects both privacy and confidentiality concerns. Westin (1979) found that one in two

Americans are worried about how businesses and the Federal government will use the personal information it gathers on individuals (Westin, 1979). Martin (2000) studied public perceptions of privacy and confidentiality during the two most recent censuses (1990 and 2000). She notes that, during Census 2000, people increasingly came to believe that information they provided on the census form could be used against them, through a breach of confidentiality. Thus, often times respondents are simply reluctant to reveal personally identifying information. However, in some cases, respondents may be willing to provide the information, but may find difficulty in remembering or locating the requested information, particularly when responding for other household members.

Many researchers have examined the effects of attitudes related to privacy, confidentiality, and data sharing on survey response. Singer (2002) reports results from a 1992 field experiment in which, 13.9% of households returned a census form, but did not provide a SSN for any household member. SSN was missing for at least one person in 17.7% of the households for which a census form was returned. Dillman, Sinclair and Clark (1993) also found that asking SSN lowers survey completion rates. These drops could be due to the respondent's objections to providing personally identifying information, or to the difficulty in obtaining this information for some household members.

Other studies have measured the relationship of privacy concerns and demographic characteristics. Singer, Mathiowetz, and Couper (1993) found that the effects of confidentiality and privacy concerns varied among black and white respondents and concluded that blacks have higher privacy concerns (over Hispanics and Whites). Singer (forthcoming) also found that women were more concerned about privacy and less willing to say they would provide their SSN to the Census Bureau.

There is some evidence that failure to provide SSN (or other personally identifying information, for that matter) is not always due to respondent resistance. Nonresponse may sometimes be a result of the lack of availability, or inaccessibility of the information

¹ New matching techniques are now enabling data linkages to be made without the need for SSN.

to the respondent. Bates (1992) cites focus group evidence that providing SSNs for children might be more difficult because SSNs are not routinely used before a certain age. Dillman, Reynolds, and Rockwood (1991) report focus group evidence that even though some people had no objection to providing social security numbers, finding this information, especially for children and unrelated household members, might be difficult. In 1992, Bates analyzed response to SSN by age, and found that, compared to other age groups, rates of nonresponse were significantly higher for persons aged 1-4 and 5-17.

Thus far, no empirical research has measured the effects of a particular type of SSN request or notification of administrative record use on item nonresponse to the SSN item in a decennial census environment.

The analysis in this paper focuses on response rates to the SSN item and begins with simple logistic regression models including only the main treatment variables (type of SSN request and notification of administrative record use). Additional variables are added to test whether differences in response for each of the main experimental treatments are influenced by the presence of other variables.

2. Experimental Design

The Census Bureau undertakes a program of experimentation during decennial censuses to measure the effectiveness of new techniques, methodologies, and technologies in the special environment that a decennial census generates. Experimental forms were embedded into mailout/mailback short forms during Census 2000, designed to test the effects of different notifications, two strategies for obtaining SSN information, and notification combined with the SSN request on response behavior (Neugebauer, 2000).

Two types of administrative record use notification were tested: general and specific. The notification was written in the letters accompanying the questionnaires and describes how and why the Census Bureau may use administrative records data from other Federal agencies. The general notification mentioned the Census Bureau's use of statistical data from other Federal agencies, while the specific notification went further to name the Federal agencies (Guarino et al.) Households selected for this experiment were randomly assigned

to one of four groups. Two groups received forms with a SSN request either for all household members or for the first person listed on the form (i.e. 'Person 1'). Apart from a statement that informed respondents that providing SSN was voluntary, notification was not a part of these groups (Guarino, et al., 2001). The remaining two groups received differently worded notifications of administrative record use, combined with a SSN request for all household members.

Specifically, the four experimental groups were:

- (1) All Household SSN Request
- (2) One (Person 1) SSN Request
- (3) All SSN Request, General Notification
- (4) All SSN Request, Specific Notification

Each group received full census short form mailout materials in the same sequence and timing as the official Census 2000 schedule. The experimental letters and forms were the official census forms received by the sampled households (Guarino et al., 2001)

The mailout sample size for each experimental group was just over 5,200 addresses. Overall mail response rates by experimental group showed no significant differences.

The sample was equally allocated to two strata that reflect anticipated differences in the race and tenure composition of the population and, based on previous census experience, differences in Census 2000 mail return rates. The low coverage area (LCA) stratum is expected to contain a very high proportion of the Black and Hispanic populations and renter occupied housing units. The high coverage area (HCA) stratum contains the remaining addresses (Guarino et al., 2001). Figures in this paper are weighted to account for oversampling of the LCA stratum.

3. Analytic Plan

3.1 Statistical Analysis

Logistic regression analysis was used to determine the effects of the type of SSN request and administrative record use notification on response to the SSN item. One dependent variable was considered: response to the SSN item. The effect of type of SSN request and administrative record use notification was evaluated by fitting a logistic regression model to include these main effects while controlling for other predictor variables. Control

variables used in the item nonresponse models included strata (coverage area) and various demographic measures.

For this analysis, the definition of what constitutes a response to SSN takes an operational standpoint. That is, a response to SSN was considered valid if that response was 'usable' to the census in terms of data linkage. In light of this view, this analysis considered SSNs missing when SSN was blank OR the SSN entry was less than nine digits long².

To analyze the effect of the experimental treatments in the presence of other variables, two series of logistic regression models were investigated -- a series of models for Person 1 and a series of models for all other household members (Persons 2-6).

3.2 *A priori* Hypotheses

- a. No significant differences will be found in item response rates to SSN for respondents who receive administrative record use notification (general or specific) in addition to a request for SSN as compared to those groups with only a request for SSN (with no notification).
- b. No significant differences are expected in item response rates to SSN for person 1 by type of SSN request.
- c. Models are expected to maintain original findings (see a and b, above) when controlling for certain demographic groups.

3.3 Limits

- This paper does not validate the SSNs collected from the four experimental groups.
- An assumption is made that Person 1 is the respondent. However, DeMaio and Bates (1990) point out that Person 1 is not the respondent about 30% of the time.
- A post-analysis investigation of missing SSNs revealed that just 69% of the SSNs coded as missing were actually blank. The remaining 31% were entries less than 9 digits long. However, this analysis does not separate the types of SSN missingness. Such a separation may have been useful in determining entry error rather than refusal or lack of knowledge.
- Since the universe deals with respondents only, there is some potential for self-selection bias. That is, had nonrespondents to the census

actually responded, it is likely that they would not have responded to the SSN item.

3.4 Variance Estimation

Since the analysis was done at the person level, a clustering effect at the household level was considered on the assumption that each household had one respondent for all household members. To take into account the stratified sample design in the data analysis, standard errors were computed for all estimates using a stratified jackknife approach.

4. Results

Is there a difference in item response to SSN by specific or general notification?

Before proceeding to the bulk of this study, it was interesting to know if the two notifications affected respondents differently in regards to providing their SSN. A chi-square test revealed no significant difference in item response to SSN between general and specific administrative record use notification ($p=0.5683$). That is, the difference between general and specific notifications was indistinguishable with respect to social security number response. Since no difference was found in the type of notification, the two types were combined into one variable, notification, for the remainder of the analysis.

What is the effect of administrative record use notification and type of SSN request on SSN reporting for Person 1?

The *Person 1 Simple Model* shown in Table 1 investigates the effect of the type of SSN request and notification on the odds of SSN being reported for Person 1. Results show that the type of SSN request (Person 1 or all household) is not significant in the model. That is, respondents are just as likely to provide SSN regardless of whether the SSN request is for all household members or just Person 1 (Odds Ratio = 0.991). However, the effect for notification is significant in the model. The odds of a respondent providing their own SSN increases by over 18% when given notification of administrative record use (versus a SSN request with no notification). In short, the *Person 1 Simple Model* logistic results in Table 1 suggest respondents are more likely to give their SSN when given notification about the reasons for asking, than when given no notification at all.

² As a function of the data capture system, there were no SSN entries longer than nine digits.

Table 1. Coefficients of Weighted Logistic Regression Model predicting SSN Reports for Person 1

Predictors	Person 1 Simple Model		Person 1 Complex Model		Person 1 Interactions Model	
	Coefficient	Odds Ratio	Coefficient	Odds Ratio	Coefficient	Odds Ratio
Intercept	1.535***	--	1.527***	--	1.307***	--
SSN Request						
All Household	-0.009	0.991	0.004	1.004	0.004	1.004
Person 1	--	--	--	--	--	--
Notification						
No Notification	--	--	--	--	--	--
General/Specific	0.185***	1.204	0.184***	1.202	0.317**	1.373
Race						
White	--	--	-0.083	0.921	-0.083	0.920
Non-white	--	--	--	--	--	--
Sex						
Male	--	--	-0.085	0.919	-0.086	0.918
Female	--	--	--	--	--	--
Strata						
HCA	--	--	0.138*	1.148	0.334*	1.397
LCA	--	--	--	--	--	--
Interactions						
HCA*Notification	--	--	--	--	-0.118	0.887

*p<.10, **p<.01, ***p<.001

To what extent do differences in SSN response for Person 1 hold when controlling for other variables?

The *Person 1 Complex Model* in Table 1 investigates the effects on response to SSN when demographic variables are introduced to the model. Although not significant in the first model, type of SSN request was left in for this second model to see if the added covariates would improve the ability to measure the effect of the type of SSN request. Race of respondent, sex of respondent and strata are added to the model. Table 1 shows that strata is the only significant addition to the model. The coefficient suggests that, for respondents in the HCA, the odds of responding to SSN increase by 15%. Race and sex of respondent are not significant predictors of response to SSN in the model. The initial findings on notification hold after the addition of these demographic variables. The added covariates do not impact the effect of SSN request on the model.

Lastly, we determine the effects of adding an interaction term to the model. Interaction terms help determine if a treatment effect differs among subpopulations and other treatments. That is, does the impact of X depend on the level of Y? This

interaction term tests whether notification of administrative record use differs by strata. The *Person 1 Interactions Model* shows the result of the addition of the interaction term, HCA*notification. No significant effect on the model is found for this interaction. That is, while stratum in itself is a significant predictor of response to SSN, the effect of notification affects both strata similarly.

What is the effect of administrative record use notification on SSN reporting for Persons 2-6?

The *Persons 2-6 Simple Model* (shown in Table 2) investigates the effect of administrative record use notification on the odds of SSN being reported for Persons 2-6. As was found for the effects on Person 1, administrative record use notification is a significant predictor of response to SSN. Specifically, the odds of a respondent providing SSN for Persons 2-6 is 12% greater when given notification of administrative record use (versus a SSN request with no notification). In short, the *Persons 2-6 Simple Model* logistic results in Table 2 suggest that notification of administrative record use significantly increases the odds of a respondent providing an SSN for Persons 2-6.

Table 2. Coefficients of Weighted Logistic Regression Model Predicting SSN Reports for Persons 2-6

Predictors	Persons 2-6 Simple Model		Persons 2-6 Complex Model		Persons 2-6 Interactions Model	
	Coefficient	Odds Ratio	Coefficient	Odds Ratio	Coefficient	Odds Ratio
Intercept	0.976***	--	-0.144	--	-0.348	--
Notification						
No Notification	--	--	--	--	--	--
General/Specific	0.117**	1.124	0.121**	1.129	0.226	1.254
Relationship						
Spouse	--	--	0.487***	1.628	0.257	1.293
Child	--	--	0.437***	1.548	0.430*	1.537
Other	--	--	--	--	--	--
Age						
0-4	--	--	--	--	--	--
5-17	--	--	0.380***	1.462	0.381***	1.463
18-24	--	--	0.683***	1.980	0.684***	1.981
25-34	--	--	0.820***	2.270	0.820***	2.271
35-44	--	--	0.914***	2.495	0.913***	2.492
45-54	--	--	0.961***	2.613	0.961***	2.614
55-64	--	--	1.035***	2.816	1.036***	2.819
65-74	--	--	0.887***	2.429	0.888***	2.431
75+	--	--	1.265***	3.544	1.271***	3.564
Strata						
HCA	--	--	0.110*	1.116	0.344	1.411
LCA	--	--	--	--	--	--
Interactions						
HCA*Notification	--	--	--	--	0.120	1.127
Spouse*Notification	--	--	--	--	0.004	1.004
Child*Notification	--	--	--	--	-0.120	0.887

*p<.10, **p<.01, ***p<.001

To what extent do differences in SSN response for Persons 2-6 hold when controlling for other variables?

The *Persons 2-6 Complex Model* in Table 2 investigates the effects of demographic subgroups on item response to SSN. Relationship to the respondent, age of Persons 2-6 and strata are added to the model. All three additional variables are significant. The relationship coefficients suggest that the odds of an SSN being reported for Persons 2-6 increases for spouses and children of the respondent, by approximately 60% and 50%, respectively. The age coefficients show that, for the most part, the odds of providing SSN for Persons 2-6 increases with age. Lastly, for respondents in the HCA stratum, the odds of responding to SSN increase by about 10%. In addition, we see that the initial findings on notification hold.

The *Persons 2-6 Interactions Model* shows the results of the addition of three interaction terms, HCA*notification, spouse*notification and child*notification to the model. The interaction terms test whether notification of administrative record use differs by type of relationship to Person 1 and by strata. We expect the effects to differ by type of relationship to Person 1. Based on previous research, SSN missingness for children is more likely due to lack of knowledge, thus notification should have less of an effect. No specific effect is expected for HCA*notification. Table 2 shows that the impact of notification does not differ by relationship to the respondent (spouse/child vs. others). The terms which interact relationship (spouse and child) and notification are not significant. Since there is no effect of notification on spouse/child vs. other household members, this gives support to the theory that level of knowledge

is what explains the relationship effects. Additionally, no significant effect on the model is found for the term which interacts strata and notification.

5. Conclusions

The analysis found that administrative record use notification does affect the reporting of SSNs. Individual models of Person 1 and Persons 2-6 revealed that administrative record use notification significantly increased the odds of a respondent providing SSN. As some of the privacy literature suggests, these results may indicate that providing a reasonable explanation for a sensitive data request may help respondents' overcome their reluctance to provide personal information.

The type of social security number request did not affect response to the SSN item for the first person listed on the form. That is, whether the request was for all household members or only for Person 1, no effect was found on response to SSN.

The main findings of both model series (Person 1 and Persons 2-6) held when additional significant predictors were introduced. Stratum was found to be a significant predictor for both the Person 1 and the Persons 2-6 models. Relationship to the respondent and age of Persons 2-6 were additional predictors of response to SSN for the Persons 2-6 model. In addition, support was found for the hypothesis that the causes of SSN missingness may vary across subgroups. Specifically, SSN missingness for children is likely to be due to lack of knowledge or inaccessibility, while SSN missingness for others is more likely to reflect unwillingness to provide the information.

As the future of administrative record use unfolds, these are important findings in terms of the collection of personal information. However, new matching techniques are enabling data linkages to be made without the need for SSN. The increasing availability of small area data makes it easier to match people in different datasets because of a reduction in occurrences of possible matches. The Census Bureau has committed to continued research in this area as it embarks on a major program of privacy research.

References

- Bates, N. 'Revised Item Nonresponse Results for Social Security Number from the Simplified Questionnaire Test (SQT),' Memorandum to Robert D. Tortura and Susan M. Miskura, dated June 18, 1992.
- DeMaio, T.J., and Bates, N.A. (1990) 'Who Fills Out the Census Form?,' *Proceedings of the Survey Research Methods Section of the American Statistics Association*.
- Dillman, Don A., Reynolds, R.W., and Rockwood, T.H. (1991) Focus Group Tests of Two Simplified Decennial Census Forms. Pullman, WA: Washington State University, Social & Economic Sciences Research Center.
- Dillman, Don A., Sinclair, Michael D., and Clark, John R. (1993) 'Effects of Questionnaire Length, Respondent Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys.' *Public Opinion Quarterly*, Vol. 57, 289-304.
- Guarino, Jennifer A. , Hill, J. M., and Woltman, H. 'Analysis of the Social Security Number (SSN) Notification Component of the Social Security Number (SSN), Privacy Attitudes, and Notification Experiment (SPAN). US Census Bureau; Planning, Research, and Evaluation Division. July 3, 2001.
- Martin, Elizabeth. 'Changes in Public Opinion during the Census' Paper presented at the Census Advisory Committee of Professional Associations, October 19,2000.
- Neugebauer, Randall, 'Census 2000 Experimentation Program Master Plan: The Social Security Number (SSN), Privacy Attitudes, and Notification Experiment.' US Census Bureau; Planning, Research, and Evaluation Division. November 5, 1999.
- Singer, E. (2002) 'Public Perceptions of Confidentiality and Attitudes toward Data Sharing by Federal Agencies', Chapter in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*.
- Singer, E., Mathiowetz, N.A., and Couper, M. (1993) 'The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 U.S. Census,' *Public Opinion Quarterly*, Vol 57, 465-482.
- Westin, Alan F. "The Dimensions of Privacy: A National Opinion Research Survey of Attitudes Toward Privacy" Conducted for Sentry Insurance by Louis Harris & Associates, Inc. 1979.