

THE EFFECT OF SSN REQUESTS AND NOTIFICATION OF ADMINISTRATIVE RECORD USE ON RESPONSE BEHAVIOR IN CENSUS 2000¹

Jennifer A. Guarino, Joan M. Hill, and Henry F. Woltman, U.S. Census Bureau
Jennifer A. Guarino, 4301 Suitland Road, Rm 1807/2, Suitland, MD 20746

Keywords: Social Security Number, administrative records, Census 2000

Abstract

The possibility of using administrative records from other federal government agencies to supplement census data has been investigated for some time at the Census Bureau. The use of administrative records could potentially increase completeness of measurement by reducing respondent burden with shorter questionnaires and improve data quality by eliminating memory/respondent errors. To realize maximum benefits, Social Security Numbers(SSNs) for each individual would be needed to link responses to administrative data. Past research has suggested that public concern about the collection of SSNs could harm response to the census if such information was sought. To investigate the validity of this claim, the Census Bureau imbedded an experiment within Census 2000. Findings suggest that the request for the SSNs slightly decreases mail response in addition to increasing the amount of forms returned with at least some missing data. Notification of administrative record use contained in a cover letter with the form is also associated with small decreases census response. However, notification significantly increases response to the SSN item for Person 1 compared to no notification.

1. Introduction

Decennial censuses beginning in 2010 may rely on the use of administrative records from other Federal agencies to complement or supplement census data. Because the use of administrative records has wide implications on decennial methodology, it is important to collect behavioral and attitudinal data on how the public responds to requests for SSNs on census questionnaires, how the public responds to differently worded notifications about the Census Bureau's use of administrative records, and the public's attitudes on privacy and confidentiality issues pertaining to the use of administrative records in a decennial census.

Past qualitative and quantitative analyses have been

conducted to assess the effects of census SSN requests on public opinion and response behavior. While focus groups revealed an extreme negative reaction to a SSN request, a 1992 mailout/mailback test indicated a smaller-than-anticipated actual decrease (-3.4%) in mail response rates (Dillman *et al.* (1994); Singer *et al.* (1992); Aguirre International (1995)). This decrease occurred in conjunction with a shorter, respondent-friendly questionnaire, a prenotice letter, a reminder post card, and a replacement questionnaire. Also, among respondents listed on the census form, just over 1 in 10 failed to provide a SSN (Bates, 1992). These findings seemingly contradicted the anticipated extent to which respondents would resist providing an identifier with data linking implications.

For further investigation, a question asking respondents' willingness to provide their SSNs on census forms was included in a series of surveys aimed at measuring privacy attitudes of U.S. residents over time. Singer (forthcoming) reports that the percentage of respondents willing to provide their SSN on a census form declined from 68% in 1996 to 55% in 1999 and 56% in 2000. The drop in willingness was significant between 1996 and 1999, with no further significant change between 1999 and 2000.

To date, no empirical research has measured the effects of a SSN request or notification of administrative record use on response in a decennial census. An experiment imbedded in Census 2000 contains a research component designed to address the effects of different notifications and strategies for obtaining SSN information on response behavior.

1.1 Hypotheses

Given the past research, several *a priori* hypotheses were developed. For the notification treatments for which past research is limited, hypotheses were developed based on expectations from privacy research.

1. With regard to the SSN request, we hypothesized that

¹This 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.

mail response rates will drop when a SSN request is present, with a larger observable effect in areas of typically low census coverage, where response is already low, compared to high coverage areas.

2. We anticipated that the request for SSN will increase the amount of incomplete forms returned compared to no SSN request.

There is little guidance from past research driving the hypotheses about the effect of notification on response.

1. We suspect that notification of administrative record use will cause significant drops in mail response and increases in the amount of incomplete forms returned, with specific notification (including agency names) having a stronger effect than general notification.

2. With respect to item nonresponse, we suspect that the SSN item for Person 1 will be missing at a higher rate when general or specific notification is included with the SSN request.

3. Lastly, we believe that notification of administrative record use will increase the amount of incomplete forms returned in a more pronounced way when coupled with the long form compared to the short form, due to privacy issues raised publicly regarding long form questions during Census 2000. (Cohn, 2000)

2. Methodology

2.1 Panel Design

The experimental treatments are implemented within ten panels, seven short and three long form. General panel descriptions and sample sizes are listed below:

Short Form

1. Short form control (5249): Standard Census 2000 short form.
2. SSN request for household (5248): Modified form with SSN request for all household members.
3. SSN request for Person 1 (5247): Modified form with SSN request for Person 1.
4. General notification (5250): Vague notification of administrative record use.
5. Specific notification (5244): Notification of administrative record use including names of specific agencies.
6. SSN request for household, general notification (5255): Combination of treatments listed above.
7. SSN request for household, Specific notification (5248): Combination of treatments listed above.

Long Form

1. Long form control (5236): Standard Census 2000 long form.
2. General notification (5230): Vague notification of administrative record use.
3. Specific notification (5229): Notification of administrative record use including names of specific agencies.

There are two notifications, referred to as “general” and “specific.” The notification is 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 mentions the Census Bureau’s possible use of statistical data from other Federal agencies, while the specific notification goes further to name the agencies. The general notification is:

To improve the quality of census statistics, the Census Bureau sometimes uses records from other government agencies. Using other agencies’ records helps make the census more complete. By making better use of government records that already exist, the Census Bureau may be able to ask you fewer questions in the census.

The specific notification is:

To improve the quality of census statistics, the Census Bureau sometimes uses records from other government agencies, such as the Social Security Administration, the Internal Revenue Service, or agencies providing public housing assistance. Using other agencies’ records helps make the census more complete. By making better use of government records that already exist, the Census Bureau may be able to ask you fewer questions in the census.

Each panel receives the full complement of census mailout materials in the same sequence and timing as the official Census 2000 schedule.

2.2 Sample Design

The sample, selected from mailout/mailback areas of the United States, is 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 the Census 2000 mail return rates. Strata are denoted as low and high coverage areas (LCA and HCA). The LCA stratum is expected to contain a very high proportion of the Black and Hispanic populations and renter occupied housing units. The HCA stratum, which comprises approximately 81% of the total

mailout/mailback universe, contains the remaining addresses. All figures in this report are weighted to account for the oversampling of the LCA stratum.

2.4 Measurements

The mail response rate used in this analysis is defined as the number of non-blank questionnaires returned by mail for the treatment group divided by the number of questionnaires mailed out less those returned as undeliverable.

A form completeness indicator is created to determine which households have any missing data on their census forms. Cases with at least some missing data are identified as those households with complete respondent reported household counts that are missing any of the following for the household or household members: tenure, sex, date of birth/age, race, and ethnicity.

2.5 Analytic Procedures

The analysis of the experimental treatments is conducted by modeling response and form completeness rates using logistic regression. Statements about the significance of treatment effects are made while maintaining a 90 percent confidence level.

In order to take into account the stratified sample design in the data analysis, WesVarPC Version 2.12 is used to compute standard errors for all estimates and models using a replication methodology. WesVarPC Version 2.12 requires a two PSU per stratum design in order to use a stratified jackknife variance estimation methodology. Since there is no clustering in the sample and only two strata to which the sample is allocated, a simple jackknife approach is substituted for the stratified jackknife. This methodology requires forming 256 replicates by numbering observations consecutively within strata. Note that replicate samples are combined across the two strata, which is generally avoided in analyses that involve stratified jackknife replication. The proposed design balances the replicates by selecting them from both strata.

Twenty replicates are used in computing the standard errors for analyses involving item nonresponse since the sample size is much smaller when only respondents are considered.

3. Results

3.1 Mail Response Rates

Mail response rates for the various experimental panels are included at the end of this report in Tables A-1 and A-2.

3.1.1 What is the Effect of the SSN Request?

Based on past research, it has been suggested that the request for the SSN of Person 1 or all household members will decrease response, with a more pronounced effect in low coverage areas compared to high coverage areas.

In order to assess the effect of the SSN request in the presence of other treatments, logistic regression analysis is used to model a household's odds of responding to the census. The Simple Model shown in Table 1 investigates the effect of the SSN request for one or all persons and notification on response, while controlling for strata. This model assumes that the effects are constant within the other experimental treatment and control variables. The purpose of the model with the interaction term is to relax this assumption. The Interaction Model determines whether the effect of the SSN request on response differs based on the stratum from which it is requested.

Table 1. Logistic Regression Coefficients
Predicting the Log Odds of Responding to the Census

Variable	Simple Model	Interxn Model
Person 1 SSN = 1	-.099*	
Household SSN = 1	-.113*	
Person 1/household SSN = 1		-.105*
General Notification = 1	-.090*	-.094*
Specific Notification = 1	-.037	-.041
Long Form = 1	-.454*	-.454*
High Coverage Areas = 1	.757*	.761*
Person 1 or Household SSN * Strata		-.006
Intercept	.429	.430

* Indicates statistical significance at $\alpha = .1$.

The Simple Model results in Table 1 suggest that the request for the SSN of Person 1 as well as the SSN request for all household members decreases response to the census, while controlling for notification and type of area (odds of responding to the census decrease by 9.5% for the Person 1 SSN request, 11% for the SSN request for all households members). While this decrease in response is significant and supports our hypothesis, it is also fairly small. The SSN-Strata Interaction model reveals no differential effects of the SSN request on response (SSN*Strata = -.006) between areas which differ with regard to their propensity to respond to the census, while controlling for notification. That is, the data do not

support our initial hypothesis that the SSN effect on response would be larger in low coverage areas than high coverage areas. The slight drop in response due to the SSN request is the same in low and high coverage areas.

3.1.2 What is the effect of notification of administrative record use on response?

In the absence of past literature, it is suspected that notification of administrative record use will cause significant drops in mail response, with specific notification (including agency names) having a stronger effect than general notification.

Testing of the simultaneous significance of general and specific notification in the Simple Model in Table 1 reveals that taken together, notification of administrative record use decreases mail response². Looking at the effect of each notification type separately, the logistic results show that general notification causes a small, yet significant, decrease in response, while specific notification does not. Since this finding disagrees our initial hypothesis, we further compared the parameters associated with general and specific notification. The magnitude of the parameters and therefore the effect on response between general and specific notification is not statistically different ($p=.12$).

3.2 Item Nonresponse Rates

3.2.1 What is the Effect of SSN request on Item Nonresponse?

Given the level of resistance shown in the past for SSN request, we hypothesize that the request for SSN will increase missing data. We anticipate that this increase will be more pronounced when the SSN request is paired with the long form than when coupled with the short form.

As a proxy for data quality, the effect of the treatments on item nonresponse is assessed by looking at the effect of each treatment on the likelihood of a household having any missing data among the 100% person items³ in addition to household tenure on the returned census form. The logistic results in Table 2 are used to address the validity of item nonresponse-related hypotheses.

² The simultaneous significance of the general and specific notification was tested by summing the parameters and comparing the result to zero in a F-test.

³ 100% person items include race, age, birth date, sex, and ethnicity.

Table 2. Logistic Regression Coefficients Predicting the Log Odds of Returning an Incomplete Census Form

Variable	Simple Model	Interxn Model
Person 1 SSN = 1	.103	.107
Household SSN = 1	.201*	.201*
General Notification = 1	-.019	-.015
Specific Notification = 1	.008	.015
Long Form = 1	.189*	.243*
High Coverage Areas = 1	-.820*	-.820*
General Notification * Long Form		-.067
Specific Notification * Long Form		-.097
Intercept	-1.333	-1.337

* Indicates statistical significance at $\alpha = .1$.

In accordance with hypotheses, results from the Simple Model reveal that the SSN request for all household members is associated with having missing data on the returned census form (odds of having missing data increase by a factor of 1.25). It is interesting to note that the request for the SSN of Person 1 is independently not associated with having missing data. However, collectively, any request for SSN seems to increase the odds of having at least some missing data on the form ($H_0: \beta_{SSN \text{ for Person 1}} + \beta_{SSN \text{ for household}} = 0, p=.026$).

3.2.2 What is the Effect of Notification on Item Nonresponse?

We also suspect that notification of administrative record use will cause significant increases in the amount of forms returned with missing data, with specific notification having a stronger effect than general notification. Moreover, we believe that notification of administrative record use will increase forms with missing data in a more pronounced way when coupled with the long form compared to the short form, due to privacy issues raised publicly regarding long form questions. Lastly, we anticipate that the SSN item for Person 1 will be missing at a higher rate when any notification is included with the SSN request.

From the Simple Model in Table 2, it is clear that notification does not appear to adversely affect form completeness. Individually, neither type of notification has an effect. This finding contradicts prior hypotheses

of a suspected correlation between notification and more item missing data.

The Interaction Model allows a test of the hypothesis that notification of administrative record use has a more harmful effect on form completeness with the long form than the short form. The interaction parameters in that model suggest that there are no differential effects of notification on form completeness between the long and short census forms, regardless of the type of notification.

In order to examine our success at obtaining SSN information⁴ for Person 1, a logistic regression model was formed to determine if the experimental treatments affected response to this item, while controlling for the demographics of Person 1 on the census form in order to make a sharper comparison.

Table 3. Logistic Regression Coefficients Predicting the Log Odds of Person 1 SSN Missing by Experimental Treatments

Variable	Simple Model with Controls
Household SSN = 1	.009
General Notification = 1	-.275*
Specific Notification = 1	-.357*
High Coverage Areas = 1	.061
Person 1 Black = 1	-.221
Person 1 Hispanic = 1	-.072
Age of Person 1	-.001
Number of Persons in Household	.053*
Renter-occupied Household = 1	-.122
Intercept	-1.838

* Indicates statistical significance at $\alpha = .1$.

The model in Table 3 shows that there is no difference in item nonresponse to Person 1 SSN when the panel requesting only one SSN is compared to the panel requesting all SSNs. From the perspective of Person 1, these forms do not differ in their request for SSN, and therefore no difference in response to this item is expected.

⁴ Cases with a reported SSN which is less than 9 digits, contains all nines or zeros, or is missing are counted as missing values.

Interestingly, either type of notification of administrative record use yields higher completion rates for the Person 1 SSN item compared to no notification, while controlling for demographic factors. This finding contradicts our initial hypothesis. In retrospect, we believe that respondents may view this notification as justification for the SSN request.

4. Recommendations

In general, the effects of the treatments (i.e. notification, SSN request) on response are not as substantial as originally anticipated. However, some effects are noticeable. For instance, it is clear that requesting SSN, for all household members or Person 1, causes small, yet significant, drops in response. Moreover, the request for SSN causes higher amounts of incomplete forms returned.

Notification of administrative data use is collectively associated with lower response as well, with general notification showing an individual effect when separated from specific notification. Once again, the drop is significant, but not very large. Notification does not appear to have any negative effects on form completeness. In fact, notification of administrative record use actually increases response to the Person 1 SSN item compared to the case when no notification is given.

Given these findings, the Census Bureau should be aware that any request for SSN in future censuses may decrease mail response. Yet, if better SSN data are desired above higher mail response, the results of this analysis suggest that notification should be included with the request.

5. Limitations

The sampling frame does not entirely represent the Census 2000 universe. Housing units in areas which tend to be more rural than the areas in the mailout/mailback universe, are excluded in this experiment. Additionally, addresses added through coverage improvement programs between the printing of address labels in September 1999 and the questionnaire mailout in early March 2000 are not included in the sampling frame which may result in a slight undercoverage of the target population.

Furthermore, non-English speaking households are excluded from this experiment since the SPAN questionnaires and forms have only been printed in English. This language restriction is in contrast to the rest of Census 2000, where respondents can request questionnaires in a variety of languages.

References

Aguirre International (1995) "Public Concerns about the Use of Administrative Records," contractor's report, July 1995.

Bates, N. (1992) "Revised Item Nonresponse Results for Social Security Number from the Simplified Questionnaire Test," Internal U.S. Bureau of the Census Memorandum, June 18.

Cohn, D. (2000) "Census Complaints Hit Home," Washington Post, May 4, 2000.

Dillman, D.A., Treat, J.B., and Clark, J.R. (1994) "Influence of 13 Design Factors on Completion Rates to Decennial Census Questionnaires," 1994 Annual Research Conference and CASIC technologies Interchange Proceedings, Rosslyn, Virginia, March, 1994.

Singer, E. (Forthcoming) "Public Perceptions of Confidentiality and Attitudes toward Data Sharing by Federal Agencies", Chapter in upcoming book entitled *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*.

Singer, E., and Miller, E. (1992) "Reactions to the Use of Administrative Records: Results of Focus Groups Discussions," Census Bureau report, Center for Survey Methods Research, August 24, 1992.

Household SSN, Specific Notification 70.0%

Long Form Panels:

Long Form Control 63.5%

General Notification 62.2%

Specific Notification 63.9%

Table A-1. Response Rates for SSN Request and Control Panel

Experimental Panel	Response Rate
Short Form Control	73.1%
Household SSN	72.0%
Person 1 SSN	71.7%

Table A-2. Response Rates for Notification and Control Panels

Experimental Panel	Response Rate
<u>Short Form Panels:</u>	
Short Form Control	73.1%
General Notification	71.8%
Specific Notification	73.5%
Household SSN, General Notification	69.5%

