

Key words: Long form sample loss, alternate data collection forms

Introduction

The Census 2000 Dress Rehearsal is the Census 2000 testing program that was conducted in Columbia, South Carolina and eleven surrounding counties; Menominee County, Wisconsin; and Sacramento, California. Only two of the Dress Rehearsal sites will be examined in this analysis, Sacramento and South Carolina.

Two forms were used to collect Dress Rehearsal data—the short and long forms. The short form contains basic demographic questions such as sex, race/ethnicity, relationship, length of time lived at their residence, and the number of people living at the residence. The questions on the short form are known as the 100% items. In addition to the short form items, the long form contains questions such as marital status, highest level of education completed, income, citizenship, employment, and housing items. The household's a priori form assignment is determined by a sampling scheme, that assigns on average 1-in-6 households a long form, and all others households a short form.

This paper discusses two factors that contributed to the loss of long form sample data in Dress Rehearsal: 1) the use of alternate data collection forms and 2) total nonresponse to all long form items. For the Dress Rehearsal, respondents were allowed to respond through two alternate data collection methods: the Be Counted Form (BCF) and interviews taken through Telephone Questionnaire Assistance (TQA). There are also cases in which a return was received without any long form data. Each factor will be discussed independently.

Alternate Data Collection Forms

Background

One of the goals of Census 2000 is to maximize response, both through the use of respondent friendly forms and through more accessible alternate data collection forms. There are two alternate data

collection forms--Be Counted Forms (BCF) and interviews taken through Telephone Questionnaire Assistance (TQA).

There are several possible situations which could lead respondents to use Be Counted Forms (BCF) or the TQA data collection instrument. First, the BCF was designed to provide an opportunity for people to be counted who think that they were not included on any other census form. For example, if a boarder in a single family household was not listed as part of that household on the mail form, he/she could pick up a BCF, complete and return it. Though not the original intention of TQA, it can accommodate this situation as well.

Second, the TQA interview instrument was developed to help people respond who for any reason were having difficulty completing their mail form by giving them the opportunity to provide the census information over the phone with the assistance of a trained interviewer. This includes people who have more than one place to live and are not sure at which residence to count themselves, as well as people who have misplaced their mail forms, never received a form, or received a form with an incorrect address. People who have misplaced their form, never received a form, or received a form with the incorrect address can also request another form be mailed to them. For the purpose of this analysis, when the caller cannot provide their census identification number in either an interview or replacement mailing situation, it will be considered an alternate data collection. Without an identification number it is not possible to identify the form type assigned to the caller's address, so a form type is assigned within the TQA instrument (i.e., short to long forms). Thus, it is possible in either the TQA interview or TQA replacement mail situation to collect short form data from housing units that were intended to be long form recipients, and vice versa.

The objective of this evaluation is two-fold, 1) to determine the percentage of households and people who are counted by a short form rather than a long form as intended; 2) to determine whether there are

significant statistical differences between those who respond by long form as intended and those who do not.

Methodology

The first step of the analysis compared percentages of final selected returns¹ that were a priori long form but were counted by a short form at the household level. The second step included a test of differences for age and race/ethnicity.

Results

There were 19,469 households in Sacramento that were a priori long form, of these:

- 19,196 (98.6%) were counted as assigned
- 62 (0.4%) were counted by an alternative data collection method
- 211 (1.0%) were counted by a short form mail or enumerator return

There were 37,223 households in South Carolina that were a priori long form, of these:

- 36,847 (99.0%) were counted as assigned
- 29 (0.1%) were counted by an alternative data collection method
- 347 (0.9%) were counted by a short form mail or enumerator return

The last bullet in each of the sites show that there are instances in which a household can return either a short mail or enumerator form, such as the case when 1) multiple returns are received from a household, and a short form is selected over a long form or vice versa, 2) through the mislabeling of forms during Nonresponse Followup or 3) households added during Update Leave that were given a short form instead of a long form.

Test of Difference

Tables 1 - 2 include a test of difference for age and race/ethnicity of a priori long form persons that were counted as assigned and those that were not. Each column includes all form types- mail returns, update leave (South Carolina only), enumerator returns, and alternate data collection forms. Only significant statistics are shown in the table.

TABLE 1 -DEMOGRAPHICS OF PERSONS WHO WERE COUNTED AS ASSIGNED VERSUS NOT AS ASSIGNED
SACRAMENTO

Demographic Characteristics	Counted as Assigned	Not Counted as Assigned
	N= 43,796	N=848
Age		
< 6	8.32%	11.79%
6 - 17	17.37	20.52
18 - 29	15.33	19.22
30 - 45	25.13	25.47
46 - 65	20.50	14.62
> 65	13.35	8.37
Race/Ethnicity		
White, alone	54.94%	39.29%
Black, alone	18.09	19.75
American Indian/ Alaska Native alone	2.78	3.07
Asian	14.41	21.46
Two or more races	4.96	6.49
Hispanic or Latino (of any race)	19.03	26.18

p <= .001

- Those in the younger categories, below the age of 30, were more likely to not be counted as assigned.
- Races other than White and Hispanics appear less likely to be counted as assigned compared to Whites.

¹ More than one return may be returned and checked-in for an address. However, the Census Bureau employs an algorithm based on several factors that selects the final return(s) for an address. This analysis is based only on the selected returns.

TABLE 2 - DEMOGRAPHICS OF PERSONS WHO WERE COUNTED AS ASSIGNED VERSUS NOT AS ASSIGNED

SOUTH CAROLINA

Demographic Characteristics	Counted as Assigned N= 86,582	Not Counted as Assigned N=963
Age		
< 6	7.88%	9.87%
6 - 17	17.01	20.52
18 - 29	15.46	19.22
30 - 45	24.74	25.03
46 - 65	22.76	21.50
> 65	12.14	7.48
Race/Ethnicity		
White, alone	57.92%	51.09%
Black, alone	37.79	45.59
Hispanic or Latino (of any race)	1.99	1.14

$p < .001$

- Those below the age of 30 appear more likely to not be counted as assigned.
- In several categories it appears that races other than White were less likely to be counted as assigned compared to Whites.

Conclusions

- Alternative data collection methods had virtually no effect on long form sample loss, about 0.4% in Sacramento and 0.0% in South Carolina.
- The majority of long form sample loss is due to form assignment problems with added units during update/leave or persons being counted by an enumerator short form.
- The distribution of long form persons not counted as assigned was greater for races other than White.
- The TQA operation for Census 2000 is currently being redesigned to collect only short form data. Although this new design has the potential to further increase long form sample loss due to alternate data collection forms, we do not anticipate this as a major concern since the Dress Rehearsal results indicate that a very few households were counted through a TQA long form interview.

Nonresponse to Long Form Items

Background

In 1990 the Census Bureau experienced substantial loss of long form data due to total nonresponse to the long form items. These forms were essentially converted to short forms. In 1990, there were two forms received during a Census-- mailout/mailback forms and enumerator forms. Enumerator forms were used during Nonresponse Followup (NRFU), an operation in which data was collected from households that did not originally respond to the Census.

In 1990 over 22% of the enumerator long forms had to be converted to short forms because they were missing all of the long form data. Only 1% of mail long forms were converted. The design of the 1990 enumerator long form may have been a contributing factor to the loss of long form sample data. The design was as follows: Short form person items, long form housing items, and finally long form person items. The data from 1990 indicates that enumerators were completing the short form person and housing items, but in many cases were not able to get any long form person data. Based on this, the analysis in 1990 examined long form sample loss with respect to long form person items only.

Methodology

In the Dress Rehearsal (DR) the long form was redesigned to begin with the short form person items followed by the long form person items, and long form housing items. Total nonresponse for the long form person items were calculated for all data defined persons within a household. If the entire household completed at most one long form item this was considered nonresponse to the long form. The methodology also includes a test of difference for age and race/ethnicity for those who responded to two or more person items versus those who did not. The test of differences was restricted to enumerator forms only.

Results

The column categories indicate the number of converted mail and enumerator returns from 1990, as well as the percentage of forms missing all long form person items for each of the Dress Rehearsal sites. The mail and enumerator categories for 1990 include

the percentage of converted forms for the entire state (i.e. California and South Carolina). The Dress Rehearsal data are only comprised of the city of Sacramento and 11 counties surrounding and including Columbia, South Carolina.

TABLE 3- PERCENTAGE OF MAIL AND ENUMERATOR FORMS MISSING ALL LONG FORM ITEMS FROM 1990 VS DRESS REHEARSAL

Site	Mail		Enumerator	
	1990 (state level)	DR	1990 (state level)	DR
Sacramento	1.1 (16,145)	.3 (36)	25.1 (368,399)	22.2 (1,194)
South Carolina	1.2 (2,401)	.4 (88)	22.8 (45,610)	13.2 (1,669)

The table above indicates that in Sacramento, there was about a 3% decrease in long form sample loss from enumerator forms in the DR when compared to 1990, while South Carolina observed approximately a 10% decrease in long form sample loss from 1990.

Test of Difference

Only significant statistics are shown in Tables 4-5.

TABLE 4 - DEMOGRAPHICS OF HOUSEHOLDERS MISSING ALL LONG FORM ITEMS VERSUS THOSE THAT DID NOT SACRAMENTO

Demographic Characteristics	All Long Form Items Missing	Did Not Have all Long Form Items Missing
	N= 1,194	N=4,162
Race/Ethnicity		
White, alone	54.94%	39.29%
Black, alone	18.09	19.75
Native Hawaiian/ Other Pacific Islander alone	9.05	11.27
Hispanic or Latino (of any race)	14.74	22.13

$p < .001$

- It appears that Whites are more likely to complete a most one long form item.

TABLE 5-DEMOGRAPHIC COMPARISONS OF HOUSEHOLDERS MISSING ALL LONG FORM ITEMS VERSUS THOSE THAT DID NOT SOUTH CAROLINA

Demographic Characteristics	All Long Form Items Missing	Did Not Have All Long Form Items Missing
	N= 1,669	N=10,983
Age		
< 6	0.12	0.02
6 - 17	0.48	0.15
18 - 29	20.91	18.02
30 - 45	42.24	39.83
46 - 65	26.36	29.41
> 65	9.89	12.56
Race/Ethnicity		
White, alone	52.61%	48.65%
Black, alone	42.72	47.42
Hispanic or Latino (of any race)	2.70	2.050

$p < .001$

- Those older than 45 are more likely to respond to two or more long form person items.
- It appears that Whites are more likely to complete at most one long form items.

Conclusions

Mail back forms had little or no effect on long form sample loss. We expect nonresponse to the mailed back forms to be low because respondents who take time to return the questionnaire have probably completed it. Enumerator forms, on the other hand, significantly contributed to long form sample loss. Compared to 1990, there was a decrease in total long form nonresponse for enumerator forms in South Carolina; this may be a result of the redesign of the long form. After the 100% items which are asked for each person in the household, the redesign allowed respondents to answer person items for each person

within the household, instead of moving directly to the housing items as in 1990. This smooth transition from the 100% items to the person items may be the reason for the decrease in total long form nonresponse for enumerator forms in South Carolina. However, in Sacramento the nonresponse rate remained relatively stable; this may be a result of the urbanicity of Sacramento. Asians and Hispanics account for about 36% (144,721) of the population in Sacramento, but only about 3% (18,617) of the population in South Carolina. English may be a second language to many residents of urban areas such as Sacramento. These individuals may not respond to the Census because they may have some level of distrust for the government or may simply be unaware of the Census, and its benefits to their communities. The Census Bureau is currently examining partnership and advertising programs specific to these communities, to increase the awareness and importance of the Census.

During NRFU, enumerators are to collect data from respondents that did not respond to the Census initially. The Census Bureau must equip enumerators with facts about the importance of the Census, along with the notion that overall only 1 in 6 households receive the long form, therefore it is imperative that quality data be collected from the selected households. Enumerators should also be given extensive persuasion tactics to obtain sufficient data from persons that are not eager to respond.

In South Carolina, significant statistical differences were found for those in older age groups that completed two or more long form person items and those who did not. Specifically, the data indicates that persons over 45 are likely to complete more long form person items than younger individuals. The data also suggest that statistical differences were found for the race/ethnicity distributions for those who completed two or more long form person items and those who did not. The data show that Whites are more likely to complete at most one item, where minority groups complete two or more items.

There are several reasons for nonresponse to the Census, from its seemingly intrusive nature to the design of the questionnaire. All avenues that contribute to nonresponse should be explored extensively for the success of the Census.

References

Woltman, H. (1990) STSD 1990 Decennial Census Memorandum

NOTE: This paper reports the research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.