

Item Nonresponse: 1996 American Community Survey

Anthony Tersine

U.S. Census Bureau
Washington, DC 20233

Presented to the American Community Survey Symposium, March 1998.

This paper reports the results of 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.

Introduction

In this paper, we look at item nonresponse rates for the 1996 American Community Survey (ACS) data. Our motivation for doing this is to see what questions had high nonresponse. These questions might need improvement or be in a bad location on the paper questionnaire, and further analysis should be done on them. We also look at overall nonresponse (across all modes) by site and question, and reported characteristics of item nonrespondents. Another question that we have, but will not be answered here is, "Are there any nonresponse rates that are unusual compared to other surveys and the 1990 census?"

There are four sites in the 1996 ACS test consisting of Brevard County, FL; Rockland County, NY; Multnomah County and Portland, OR; and Fulton County, PA. There are also three modes of collection for the ACS: mail, Computer Assisted Telephone Interview (CATI), and Computer Assisted Personal Interview (CAPI). We look at individual questions ([Table 2](#) has the question numbers with their question/topic) to see how often the item was left blank, was refused, was not known or was an illegal value. Due to the nature of the data available, these responses were all combined and reported as nonresponse. For a complete listing of nonresponse rates by state, mode of collection and question see [Table 3](#) (****note that the CAPI universe is not weighted in this table).

The following gives a brief description of the three modes of data collection for the urban sites.

- Mail Phase -- The Mail phase begins with a prenotice letter mailed to each housing unit on the next to last Wednesday of the month preceding the sample month. The ACS questionnaire was mailed one week later, followed by a reminder card one week after that. A replacement questionnaire was mailed three weeks later if the original questionnaire had not yet been checked in at the processing site. Check-in of mail return questionnaires for a sample panel was cut off at the start of the third month following the sample month.
- CATI Phase -- Approximately five weeks after the mailout of the original ACS questionnaire, the CATI staff began contacting non-responding sample households by telephone. Late mail returns were removed from the CATI workload on a daily basis. This phase of nonresponse follow-up lasted for approximately four weeks.
- CAPI Phase -- The CAPI universe consisted of all outstanding non-response cases remaining after the completion of the CATI phase. A 1 in 3 subsample was selected from the outstanding cases and forwarded to the field interviewers. Field interviewers visited each assigned housing unit and attempted to conduct an interview. Late mail returns were removed from the CAPI workload on a daily basis. The CAPI phase of nonresponse follow-up lasted approximately one month.

For Fulton county, a questionnaire is delivered to the housing unit by a field representative and contact is made if there is someone home. If there is no one home, the questionnaire is left at the address. No further contact is made until it is time for CAPI. At which time a 1 in 3 subsample is taken of all addresses that have not returned their forms. CAPI occurs at the same time as in the urban sites.

Evaluation Methodology

For these analyses, we use the unedited data. This is because the editing process had not been completed at the time the analysis was started. The main problem this caused was in determining the current question's universe. This is where we had a couple of options to choose from, since sometimes the answer to a previous question determined which question you answered next. One option was if the previous question for determining the universe was left blank then include this person/household in the universe for the next question in the path plus the people/households that actually met the universe criterion. This is not the option that we used. Instead we only used the people/households that actually met the universe criterion for the current question. This means that if a previous question that determined the universe for the current question was left blank, then that person/household was not included in the universe for the current question. For example, say that question P2 (age) was left blank; then this person was not included in the universe for any subsequent question where age helped determine the universe. Thus for P40a (wages/salary income) this person was determined not to be in the universe, since P40a is asked only of people who are 15 or older. When we look at the overall nonresponse rate by site and question, the CAPI portion of the data is weighted by a factor of 3 due to the fact that CAPI respondents are subsampled at a rate of 1 in 3.

Questions with Overall High Nonresponse

We will first discuss some of our findings when looking at overall nonresponse when averaged overall sites and modes for each individual question. These are the most serious problems, and they should be addressed.

We observe that usually the same questions appear in all sites, that is the last column in [Table 1](#) is blank for the majority of the questions. Also, there are certain types of questions that appear frequently:

- income questions (4 of top 5 population questions)
- dollar amounts (9 of top 11 questions)
- questions that ask for the month or year (2 of top 11)

[Table 4](#) has a list by site of questions that had an overall nonresponse rate of at least 7 %.

Differences in Nonresponse Across Mode of Collection

We now will present some of our findings where the percentage of non-respondents differs with respect to the mode of collection (Mail, CATI, CAPI), starting with some of the questions where there are the biggest differences.

We will discuss the housing questions first.

- The question with the biggest differences between modes is H30, Property Insurance, with CATI having approximately a 1/4 higher missing rate than mail, and CAPI being about 1/5 higher than mail.
- Other questions that have mail with a lower nonresponse rate than CATI and CAPI, but with smaller differences are: H2, H19A, H19C, H24, H28, H29, H31, H32, H36, and H37. H16 (Is there a telephone?) is not asked in CATI, but also has a lower rate for mail than CAPI in all sites except Fulton, PA.
- Question H5, has a higher nonresponse rate for mail than CATI or CAPI.

For population questions, we observe the following:

- P40a, Wages/Salary Income, showed the mail having an approximately 3/10 lower nonresponse rate than CATI or CAPI.
- Other questions that follow this pattern between modes are: P2d, P2m, P2y, P12, P40b, P40c, and P41.
- The following questions have a higher nonresponse rate for mail than CATI or CAPI: P5cod, P14b, and P20m.

[Table 5](#) presents all the questions that have at least two combinations of state and question with at least a 5% difference between modes of collection. We can see that most of these questions have a higher nonresponse percentage for CATI and CAPI than for the mail.

Characteristics of Persons / Households Likely to be Item Non-Respondents

We also identified questions that had at least two combinations of mode and state with a greater than or equal to 7% nonresponse rate and examine them by reported demographic characteristics. There are 16 population questions and 15 housing questions that met this criterion. The characteristics used were age, race, educational attainment, Hispanic origin, and marital status. The specific breakdowns of these characteristics used are given in [Table 6](#). For the housing questions, we used the characteristics of the reference person. (Further analysis does not include Fulton County, PA since it appeared that for the most part there was no characteristic that showed up consistently as being related to item nonresponse.) We compared the proportions of item non-respondents vs the proportion of all respondents for a specified characteristic, and looked for differences of at least 5%. We did not look at combinations of characteristics, although this is one item where further analysis needs to be done. The example below shows how the comparisons were carried out.

Example question P41 in Brevard, mail collection, and age >=55:		
33,928	total respondents of all ages	
14,332	respondents age >=55	
2,601	total non-respondents of all ages	
1,276	non-respondents age >=55	
proportion of respondents age >=55	= 14,332 / 33,928	= 42.2%
proportion of non-respondents age >=55	= 1,276 / 2,601	= 49.1%
Difference in proportions (non-respondents minus respondents) = 6.9%		

So, people older than 55 in Brevard County that respond by mail are more likely to be item non-respondents for question P41 (based on the 5 % difference criterion).

We will present results where at least 8 out of the 31 (25 %) questions we looked at met the 5 % difference criterion for a specified characteristic.

We found that the following characteristics were related to higher item nonresponse:

- Mail
 - Missing education
 - Missing Hispanic origin
 - Older than 55
 - Widowed (except in Rockland County, NY)
- CATI
 - Missing education
 - Older than 55
 - Widowed (except in Rockland County, NY)
- CAPI
 - Missing education
 - Older than 55 (except in Rockland County, NY)

We found that the following characteristics were related to lower item nonresponse:

- Mail
 - 30-54
 - Bachelor's Degree or higher
 - Non-Hispanic
 - Married
 - White
- CATI
 - 30-54
 - Bachelor's Degree or higher
 - Less than a Bachelor's Degree (except in Rockland County, NY)
 - Married
- CAPI
 - <=29 (except in Rockland County, NY)
 - Less than a Bachelor's Degree