



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2013 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES ACS13-RER-07

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: James B. Treat(**signed 02/21/2013**)
 Chief, American Community Survey Office

Prepared by: Deborah H. Griffin
 American Community Survey Office

Subject: Feasibility of Implementing Mode Switching Rules to Improve
 Survey Estimates for Language Households

Attached is the final American Community Survey Research and Evaluation report on the Feasibility of Implementing Mode Switching Rules to Improve Survey Estimates for Language Households. This report looks at the potential value of using paradata on the reason for a CATI noninterview to select a different sample for CAPI follow up.

If you have any questions about this report, please contact Deborah Griffin at (301) 763-2855.

Attachment

cc:
ACS Research and Evaluation Working Group
M. Beaghen (DSSD)
S. Fish (ACSO)

Feasibility of Implementing Mode Switching Rules to Improve Survey Estimates for Language Households

FINAL REPORT

Deborah Griffin and Samantha Fish
American Community Survey Office

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND	2
Language Assistance.....	2
Available Paradata and Adaptive Design	2
Household Language and English Proficiency	3
RESEARCH QUESTIONS	3
METHODOLOGY	3
Estimating the Language Characteristics of CATI Noninterviews.....	3
Estimating the CATI Noninterview Reasons by Household Language and English Proficiency	4
Determining Effectiveness and Potential Workloads	4
LIMITATIONS.....	4
RESULTS	5
What are the language characteristics of the ACS sample cases that result in a CATI noninterview?	5
What are the reasons for CATI noninterviews and do the reasons vary by language characteristics?	6
Would using the CATI noninterview reason of language barrier accurately identify cases that are language households?.....	6
CONCLUSIONS.....	7
REFERENCES	8

INTRODUCTION

Shin and Kominski (2010) document the growth in the number of people speaking a language other than English at home from 23.1 million in 1980 to 55.4 million in 2007. They found that in 2007 about 1-in-5 households spoke a language other than English at home. Of these, about 8 percent reported not speaking English at all. In the 2011 American Community Survey (ACS), 60.6 million people (about 20.8 percent of the population 5 years and over) reported speaking a language other than English at home with nearly 9 percent speaking English less than “very well” (U.S. Census Bureau, 2012).

The ACS is a large national sample survey that produces demographic, social, economic, and housing estimates for a broad set of geographic areas. It uses four modes of data collection – Internet, mail, telephone, and personal visit. After allowing time for Internet and mail responses, the call centers conduct telephone follow up activities using Computer Assisted Telephone Interviewing (CATI) methods. All noninterviews after Internet, mail, and telephone attempts and all sample cases that were not eligible for Internet, mail, or telephone are eligible for selection in the subsample for personal visit follow up. We use Computer Assisted Personal Interviewing (CAPI) for this follow up. We currently apply differential sampling rates at the tract level to select the CATI noninterviews sent to CAPI. Specifically, we select a larger sample of nonrespondents in tracts with the lowest expected levels of response by mail and telephone. U.S. Census Bureau (2010), Section 4.3, includes a complete discussion of this subsampling. We currently do not use any other information (e.g., paradata) when selecting the sample for CAPI follow up.

The motivation for this research is concern (articulated by advisory committee members) about the loss of sample for households that speak languages other than English and have limited English skills. Joshipura (2008) found that “linguistically isolated households” have a significantly lower propensity to respond by mail when compared with households that are not “linguistically isolated.”¹ Based on the 2005 ACS she estimated that only about 31 percent of linguistically isolated households (versus 62 percent of households that were not linguistically isolated) responded by mail. This suggests that these households with limited English proficiency are more likely than English-speaking households to be in the universe for CAPI subsampling. It may be possible to improve the quality of survey estimates for the limited English population if the CAPI subsample took advantage of paradata to identify these cases. Specifically, oversampling CATI nonrespondents with limited English proficiency could improve the reliability of survey estimates for these populations.

To assess the potential benefits derived from using paradata on noninterview reason in CAPI subsampling when selecting the CAPI subsample, we must first demonstrate that these paradata successfully identify the cases we wanted to oversample. This project focuses on this aspect of the research. If the results show promise we should consider

¹ Prior to 2010, households were classified as “linguistically isolated” if no one in the household that is 14 or over speaks only English or speaks another language and speaks English “very well.”

additional research to determine if oversampling these cases will yield lower overall variances of survey estimates for limited-English proficient households.

BACKGROUND

Language Assistance

In the United States, we mail English questionnaires to all ACS sample addresses. As part of the advance letter, we provide a multilingual brochure telling respondents how to gain telephone assistance in Spanish, Chinese, Vietnamese, Russian, and Korean. Our call centers employ bilingual staff to provide telephone questionnaire assistance in Spanish and about 19 additional languages (Fish, 2010). During initial contacts, call center staff attempt to identify the language spoken by the household and we capture this information as the required language skill. The call centers reassign cases with a language need to staff with the required language skill. During the final stage of data collection, census field representatives, with nearly complete coverage of language needs, conduct interviews in-person.

The English questionnaire also includes Spanish text explaining that the Census Bureau can provide a Spanish questionnaire if they call the telephone support line. Recently, the ACS added language guides in Chinese and Korean. A respondent can request a guide in Chinese or Korean to help them fill out the English questionnaire if they do not wish to complete an interview over the phone. Despite these efforts, we believe that households with language needs may be more likely than households without language needs to become noninterviews after CATI due to language barriers.

Available Paradata and Adaptive Design

Adaptive design takes advantage of survey paradata to make design and operational decisions at various stages of a survey. These decisions can involve the assignment of sample cases to an initial data collection mode. Paradata or response data are also useful in defining rules to switch modes or to stop data collection. Such mode switching or stopping rules can include selecting samples for follow-ups in other modes.

The CATI operation provides paradata about noninterviews including call attempt outcomes and timing of call attempts. Two paradata sources considered for this research include the reason for the noninterview (refusal, noncontact, language barrier, and other) and the required language skill. Call center staff assign a set of intermediate outcome codes based on each call attempt and the WebCATI system assigns a final outcome code when cases are closed out. Interviewers define or update the required language skill set based on the outcome of each call attempt.

In this research we are interested in understanding if paradata (specifically, the reason for noninterview variable from the CATI operation) can help us identify cases that should be switched from the CATI mode into the CAPI mode or cases that should be oversampled after CATI attempts for CAPI. Future research could explore the feasibility of using the language skill variable.

Household Language and English Proficiency

The ACS collects data on language spoken and English proficiency for every person in the household age 5 and older. A 3-part question asks, “Does this person speak a language other than English at home?” For “yes” responses, two additional questions are asked, “What is this language?” and “How well does this person speak English?” While language responses include over 400 different language groups, for this analysis we recoded these languages into 40 language groups. We use the responses to these questions to assign a household language to every ACS interviewed household. Household language is defined as the first non-English language spoken by either the reference person, husband/wife, father/mother, brother/sister, son/daughter, grandchild, in-law, other relative, unmarried partner, housemate/roommate, roomer/boarder, foster child, or other nonrelative in that order. If no household members report speaking a language other than English at home, that household has a household language of English. Note that the definition used to assign a household language results in a generous estimate of households by language as it would include a household in a language group if only one of the members (even a nonrelative) reported speaking that language.

We define all households speaking a language other than English as “language households” and further classify these households using information on English proficiency. For this study, we define a household as a “language household lacking English proficiency” if no one in the household that is 14 or over speaks only English or speaks another language **and** speaks English “very well.” We can think of these “language households lacking English proficiency” as households in which all adults have some limitations communicating in English. This is the same definition as “linguistically isolated” that is used in the referenced reports (the name for this concept recently changed because some individuals found it to be offensive).

RESEARCH QUESTIONS

1. What are the language characteristics of the ACS sample cases that result in a CATI noninterview?
2. What are the reasons for CATI noninterviews and do the reasons vary by language characteristics?
3. Would using the CATI noninterview reason of language barrier accurately identify cases that are language households?

METHODOLOGY

Estimating the Language Characteristics of CATI Noninterviews

As noted earlier, we select subsamples of CATI noninterviews for CAPI. Since only a portion of all CAPI cases is eligible for CATI (i.e., those with an available phone number), we restricted our analysis to only those CAPI cases that were CATI-eligible. We identified these cases using variables on the ACS control file.

We weighted every CATI-eligible household interviewed in CAPI using the CAPI subsampling weights (differential sample rates used at the tract level to select the CAPI subsample). Generally, 1 in 3 households are subsampled for CAPI, so the weight for each of these households averages about 3. We weighted CATI noninterview cases interviewed in CAPI this way to estimate the language characteristics of other CATI noninterviews.

CATI noninterviews that we interviewed in CAPI provide us with survey results on language spoken and English proficiency, which allow us to estimate the language characteristics of most CATI noninterviews. We look at the CATI-eligible universe by household language and language households lacking English proficiency. For this, we used the edited, unswapped housing-level files (final edited data that are untouched by data-disclosure review) from the 2006, 2007, and 2008 ACS.

Estimating the CATI Noninterview Reasons by Household Language and English Proficiency

During the CATI operation, interviewers assign an outcome code to identify the reason for noninterviews. For purposes of this evaluation, we combined outcome codes to distinguish between four sets of outcomes:

- Refusals (179, 181, 182, 186),
- Language Barriers (24, 191),
- Noncontacts (183, 193, 194, 195), and
- Other Reasons (all other codes).

We refer to these recodes as the “noninterview reasons” in the remainder of this report. Although the individual outcome codes are reasons too, we chose to aggregate reasons into these four broad categories. We used the weighted results (weighted only for the subsampling factors) to estimate the number of sample cases by reason, by household language characteristics. From these estimates, we produced distributions of the reasons for noninterview for all households, English only households, language households, and language households lacking English proficiency.

Determining Effectiveness and Potential Workloads

Any effort to use survey paradata to identify cases for oversampling in CAPI must use data that would be available at the time that sampling needs to take place. The noninterview reasons are a possibility, specifically the language barrier reason. To measure the effectiveness of this specific reason as a proxy for cases with language needs, we need to determine the proportion of all cases coded as having a language barrier that would identify a household that speaks a language other than English. We defined CATI workloads as the estimated proportion of the total noninterview universe that we would identify using the language barrier reason for noninterview code.

LIMITATIONS

Interviewer coding of reasons for noninterviews is subjective. For example, a household that refused to participate may refuse because they do not speak English and interviewers

could correctly code this case as either a language barrier or a refusal. We also code households that do not pick up the phone because they do not speak English as noncontacts, when they might be language barriers.

We did not make any adjustments for noninterviews in CAPI. This means that these estimates potentially underestimate the true number of CATI-eligible noninterviews that we subsample out of CAPI. However, CAPI response rates were about 95 percent for 2006, 2007, and 2008, so we do not expect this to be a major limitation (Cepietz, 2009).

This research began several years ago using the 2006, 2007, and 2008 ACS. If the findings show promise, we should repeat the analysis using more recent datasets reflecting the ACS sample expansion and increased staffing in certain languages in the call centers. In 2009, the call centers expanded their language support, reducing the level of language-related noninterviews and improving the identification of specific language needs. In 2013, we added an Internet response option. We should base the decision to implement changes in production on results that are more recent.

RESULTS

What are the language characteristics of the ACS sample cases that result in a CATI noninterview?

Table 1 summarizes the total CATI noninterviews and the language characteristics of ACS sample cases each year that were noninterviews in CATI. As noted earlier, we chose to classify these households as English only households (households where no one reported speaking a language other than English at home), language households, and language households lacking English proficiency. As described in the background section, language households are a generous estimate of language needs because these households may include one or more adult members that speak English “very well.” Likely, these types of household do not require language assistance to respond to the ACS since one or more members could respond in English.

Table 1. Language Characteristics of ACS CATI Noninterviews

	2006	2007	2008
TOTAL CATI noninterviews	255,156	309,036	298,005
Percent English only households	75.9	76.2	75.8
Percent Language households	24.1	23.8	24.2
- Spanish	16.4	16.2	16.4
- Other Indo-European	4.0	4.0	4.0
- Asian and Pacific Island	2.8	2.6	2.8
- Other	0.9	1.0	1.0
Percent Language households lacking English Proficiency	7.3	7.2	7.3
- Spanish	5.3	5.3	5.3
- Other Indo-European	0.9	0.9	0.9
- Asian and Pacific Island	0.8	0.9	0.9
- Other	0.2	0.2	0.2

Table 1 includes estimates for all languages combined and for four subgroupings of languages: Spanish, Other Indo-European languages, Asian and Pacific Island languages, and other languages. The proportions are consistent over these three years with about 24 percent of the CATI noninterviews being language households and about 7 percent of the CATI noninterviews being language households lacking English proficiency.

What are the reasons for CATI noninterviews and do the reasons vary by language characteristics?

We are interested in understanding why we were unable to interview these cases in CATI and want to determine if additional call center resources in certain languages could be successful in reducing the number of cases that are lost due to subsampling.

Table 2 summarizes the overall distribution of CATI noninterview reasons for the 2008 ACS, although the distributions for 2006 and 2007 are similar. About 77 percent of the total 2008 CATI noninterviews were noncontacts, about 16 percent were refusals, and less than 1 percent were due to language barriers. Table 2 also displays these distributions by household language characteristics. One key observation is that the distribution of noninterview reasons for language households is similar to the distribution for English only households. For both groups most noninterviews are noncontacts with refusal being the second most frequent noninterview reason.

Table 2. CATI Noninterview Reasons by Household Language Characteristics - 2008 ACS

Household Language Characteristics	CATI Non-Interviews	Percent Refusals	Percent Language Barriers	Percent Non Contacts	Percent Other Reasons
English only households	226,062	14.8	0.1	77.9	7.2
Language households	71,943	18.1	1.8	72.4	7.6
Language households lacking English proficiency	21,650	18.6	3.4	71.2	6.8
Total households	298,005	15.6	0.5	76.6	7.3

Noncontacts represent about 71 percent of the 2008 CATI noninterviews for language households that lack English proficiency. Refusals explain another 19 percent with other reasons accounting for about 7 percent. It is interesting to note that while the rate of language barriers is higher, still only a small portion of these noninterviews (about 3 percent) is due to a language barrier. It is possible that interviewers are coding some cases as refusals when they really are language barriers or that language need households with a home telephone may be less likely to pick up the phone and we count them as noncontacts (see Limitations section).

Would using the CATI noninterview reason of language barrier accurately identify cases that are language households?

If we wanted to try to use survey paradata to identify language households for oversampling in CAPI we would want to use a reason for noninterview that was effective in identifying language households, especially those that lack English proficiency. It is

reasonable to expect that the language barrier code would identify these cases and it does, but it only identifies a small fraction of the total language households.

Table 3 displays the distribution of household language characteristics for 2008 CATI noninterviews by noninterview reason. The 2006 and 2007 results were similar. As noted in Table 1, about 24 percent of all CATI noninterviews were language households and about 7 percent were language households lacking English proficiency. The language barrier outcome code is effective in predicting a language household - almost 90 percent of the CATI noninterviews coded as language barriers identify a language household and about 49 percent identify a language household that lacks English proficiency. However, referring back to Table 2, using this outcome code would only identify about 3.4 percent of the CATI noninterview households that lack English proficiency and only about 1.8 percent of the CATI noninterviews that are language households.

Table 3. Effectiveness of Identifying Language Households Based on CATI Noninterview Reason - 2008 ACS

CATI Noninterview Reason	Total Households	English only Households	Language Households	Language Households Lacking English Proficiency
Percent Refusals	46,489	71.9	28.1	8.6
Percent Language Barriers ²	1,490	10.4	89.6	48.9
Percent Non Contacts	228,272	77.2	22.8	6.8
Percent Other Reasons	21,754	74.9	25.1	6.8
CATI Noninterviews	298,005	75.8	24.2	7.3

If we were to use the reason for noninterview outcome to select all language barrier cases for CAPI, the increase in the monthly workload would be small, about 125 additional cases.

CONCLUSIONS

While the language barrier outcome code from CATI is effective in identifying households that speak languages other than English, especially those lacking English proficiency, it only identifies a small fraction of these households within the CATI nonresponse universe. This raises the issue of whether the investment in using this variable for oversampling is worth the payoff.

The vast majority of CATI noninterviews are due to noncontacts, regardless of language need, reminding us of the increasing problems with completing interviews by telephone. The ACS currently does not contact cell phones and, like other surveys, we find the drop in landline phones to impact contact rates. This problem exists for both non-English and English only households. The low rates of noninterviews due to language barriers

² In Tables 2 and 3, a small number of English only households have an outcome code of language barrier. This could be due to an interviewer assigning an incorrect outcome code. It also is possible that the individual that answered the phone with a language barrier turned out to be a non-household member.

indicate that for many languages, the reason that we did not obtain the interview is not necessarily a consequence of call center language resources, and that is good news.

Given these findings, it does not appear that we should invest additional resources in research to assess if the language skill set variable, either alone or in combination with the language barrier variable, might improve the efficiency of identifying language households for CAPI. Repeating the analysis with current data should not be a priority as recent design changes are unlikely to change the compelling findings of low rates of language barrier noninterviews. Future research on adaptive design applications in the ACS should nonetheless consider this variable in conjunction with other variables and possibly, administrative records.

REFERENCES

Cepietz, Erica. 2009. "American Community Survey and Puerto Rico Community Survey: 2008 Housing Unit Response Rates and Margins of Error by Mode." Memorandum to Susan Schechter, Chief of the American Community Survey Office. U.S. Census Bureau. Sept 14, 2009.

Fish, Samantha. 2010. "Assessment of Language Needs and Language Assistance Resources in the 2006 – 2008 Computer Assisted Telephone Interviewing Operation." ACSO Research and Evaluation Memorandum Series. U.S. Census Bureau. February 17, 2010.

Joshiyura, M. (2008). 2005 American Community Survey Respondent Characteristics Evaluation.
http://www.census.gov/acs/www/Downloads/library/2008/2008_Joshiyura_01.pdf

Shin, Hyon B. and Robert A. Kominski. 2010. Language Use in the United States: 2007, American Community Survey Reports, ACS-12. U.S. Census Bureau, Washington, DC.

U.S. Census Bureau. 2010. "Design and Methodology Report - American Community Survey." http://www.census.gov/acs/www/methodology/methodology_main/

U.S. Census Bureau. 2012. Household Language by Linguistic Isolation: 2006-2008 American Community Survey 3-Year Estimates. Table B16002.
<http://factfinder2.census.gov>

U.S. Census Bureau. 2012. Selected Social Characteristics in the United States: 2011 American Community Survey 1-Year Estimates. <http://factfinder2.census.gov>