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

2016 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS16-RER-04

DSSD 2016 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS16-MP-04

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: Patrick J. Cantwell
Division Chief, Decennial Statistical Studies Division

Prepared by: Dorothy Barth
Mary Frances Zelenak
Mark E. Asiala
Edward Castro
Decennial Statistical Studies Division
Andrew Roberts
American Community Survey Office

Subject: 2015 Envelope Mandatory Messaging Test

Attached is the final American Community Survey Research and Evaluation report "2015 Envelope Mandatory Messaging Test." This report provides the results of research into the effect of the removal of the mandatory message "YOUR RESPONSE IS REQUIRED BY LAW" from the envelopes for the initial mailing and paper questionnaire packages on return rates and final response rates. The report also investigates the impact of the message removal on "hard-to-count" populations, cost estimates, and survey reliability estimates.

Stakeholders and some respondents have raised concerns about the prominent references to mandatory participation in the American Community Survey (ACS) on the mail materials. In order to be responsive to these concerns we conducted this test. The results show that eliminating the phrase "YOUR RESPONSE IS REQUIRED BY LAW" from the initial mail package envelope and the paper questionnaire package envelope does lower the self-response return rate by 5.4 percentage points, which is statistically significant. This rate reflects response before the start of computer-assisted telephone interviewing and it pushes additional cases into the more expensive follow-up modes. We estimate the cost impact of eliminating mandatory messages

from the envelopes, in the manner we tested, would increase the annual costs of the survey by roughly $9.4 million, a 4.1 percent increase relative to the ACS budget for the 2015 fiscal year, and will increase the variance of estimates by about 3.5 percent, with no changes to the current methodology. This report provides additional analysis on possible methodological changes to ACS sampling that would need to be done in order to maintain current costs or to maintain current survey reliability, if the changes in this test were to be implemented in full ACS production. The report also provides analysis on how the test impacted "hard-to-count" populations by examining response rates in high and low response areas and response distributions of certain characteristics known to be associated with low response propensity.

It is important to recognize that this test is just the first step in our research to examine the impact of modifying the mandatory messages contained in the ACS mail materials. Additional testing conducted in connection with the September 2015 ACS sample will provide important insights on ways to revise the mandatory messages in a more comprehensive manner throughout the various mail pieces. The September test also included other messaging and design enhancements to continue to be responsive to the feedback from stakeholders and respondents while preserving the quality of the survey.

If you have any questions about this report, please contact Elizabeth Poehler at 301-763-9305 or Dorothy Barth at 301-763-1833.

Attachment

cc:

| | | | |
|---|---|---|---|
| ACS Research and Evaluation Workgroup | | Colleen Hughes Keating | (POP) |
| Ashley Barnes | (ACSO) | Rita Schuler | (NPC) |
| Kathryn Cheza | | Michael Bentley | (DSSD) |
| Sandra Clark | | Broderick Oliver | |
| Lauren Difiglia | | Michael Risley | |
| Gregory Mills | | Mary Frances Zelenak | |
| Dameka Reese | | Jason Lee | (SEHSD) |
| Andrew Roberts | | | |
| Isaac Dompreh | (CSRM) | | |
| Edward Porter | | | |

# 2015 Envelope Mandatory Messaging Test

FINAL REPORT

**Dorothy Barth**
**Mary Frances Zelenak**
**Mark E. Asiala**
**Edward Castro**
**Decennial Statistical Studies Division**
**Andrew Roberts**
**American Community Survey Office**

## Table of Contents

## List of Attachments



## List of Tables

## Executive Summary

The current design of the American Community Survey (ACS) allows the Census Bureau to collect and update demographic, social, economic, and housing data for the United States every year. These data are essentially the same as the "long-form" data that the Census Bureau traditionally collected once a decade from a sample of housing units as part of the decennial census that ended with the 2000 Census. Response to the ACS is mandatory, because it is part of the decennial census.

A previous study indicated that the use of mandatory language on envelopes produces higher response rates than envelopes without the use of such language (Dillman et al.,1996). In addition, a recent study conducted on behalf of the Census Bureau suggested that the "required by law" message in the ACS mailing package may be the single most effective message in attracting attention and motivating response (Reingold, 2014). Prior to August 2015, the ACS sent up to six mailings to sampled addresses, and most of those materials contained language referring to the mandatory nature of the survey. The envelopes that contain the initial mail package and paper questionnaire package both display the following message in bold text, "**YOUR RESPONSE IS REQUIRED BY LAW**."

Stakeholders and some respondents have raised concerns about the prominent references to the mandatory participation in the ACS. In order to be responsive to these concerns, we conducted the 2015 Envelope Mandatory Messaging Test. The purpose of the test was to evaluate the impact on self-response of removing the phrase "**YOUR RESPONSE IS REQUIRED BY LAW**" from the envelopes used to mail the initial mail package as well as the paper questionnaire package. In order to field this test as quickly as possible, materials within these packages, as well as other materials sent to sampled addresses, were not changed. A separate, more extensive test, conducted in September 2015, focused on modifications to the mail materials to soften the mandatory messages while emphasizing the benefits of participation in the survey.

The results of this test show that eliminating the phrase "**YOUR RESPONSE IS REQUIRED BY LAW**" from the two mailing envelopes does significantly lower the self-response return rate, the final response rate among self-respondents, and the overall final response rate.

- Self-response return rates before the start of computer-assisted telephone interviews
  - Control Sort Group: 42.5 percent; Test Group: 37.1 percent
  - Difference: 5.4 percentage points
- Final response rates after telephone and personal interviews
  - Control Sort Group: 96.2 percent; Test Group: 95.6 percent
  - Difference among Self-respondents: 5.8 percentage points
  - Difference in Overall Response: 0.6 percentage points

The cost of implementing the changes from this test in production would be substantial, because of the significant decrease in self-response. The lower rates of response for both Internet and mail would lead to an increase in the workloads of both telephone and personal visit interviews, which are more costly modes of data collection. Therefore, methodological changes that reduce

self-response and increase workloads in follow-up operations have significant cost impacts. Our estimate of the cost impact of eliminating mandatory messages from the envelopes, in the manner we tested, shows

- increased annual survey costs of roughly $9.4 million (4.1 percent increase relative to the ACS 2015 fiscal year budget)
- decreased total interviews by approximately 124,000 annually
- increased variance for survey estimates by approximately 3.5 percent

Although the contact history analysis does not suggest that this test had a significant impact on the number of attempts needed to obtain an interview in the telephone and personal visit modes, methods that reduce self-response may increase the total number of contacts that respondents would receive by including them in the telephone or personal visit operations, therefore increasing respondent burden.

Our research lacks sufficient evidence to prove that this test had a statistically significant impact on response of hard-to-count populations. However, for mail responses and combined self-responses, there were significant differences in the response distributions of *age* and *tenure*. These differences could influence the estimates for these and other related characteristics, if the changes in this test were implemented in full ACS production.

We investigated several different scenarios on how a change in ACS sampling methodology would affect cost and reliability estimates if the changes of this test were implemented in full production.

- To maintain current data collection costs and personal visit subsampling rates by decreasing the initial sample size:
    - decrease initial sample sizes by about 5.5 percent
    - increases variance for survey estimates by about 8.8 percent
- To maintain current data collection costs and initial sample size by decreasing Computer-Assisted Telephone Interviewing and Computer-Assisted Personal Interviewing workloads:
    - subsample telephone interviews; adjust the subsampling of personal interviews
    - increases variance for survey estimates by about 8.6 percent
- To maintain current levels of reliability and personal visit subsampling rates:
    - increase initial sample size by about 3.5 percent
    - increases annual data collection costs by approximately $15.4 million, a 6.7 percent increase over the fiscal year 2015 budget

Notice that in the first two options above, current survey costs are maintained. The first option saves money by decreasing the initial sample size while the second option saves money by additional subsampling. The increase in variance of the two options is about the same (8.8 versus 8.6 percent), so it would be difficult to justify the investment of time and resources necessary to implement the second option and explore improvements of this nature in the overall sample design of the ACS.

This test is just the first step in our research to examine the impact of modifying the mandatory messages contained in the ACS mail materials. Additional testing conducted in connection with the September 2015 ACS sample will provide important insights on ways to revise the mandatory messages in a more comprehensive manner throughout the various mail pieces. The September test also included other messaging and design enhancements to continue to be responsive to the feedback from stakeholders and respondents while preserving the quality of the survey.

## 1. Introduction

The American Community Survey (ACS) data provide a wealth of information. Government officials rely on the data to make informed decisions on matters of public interest such as access to emergency services, public transportation, education, medical needs, and much more. Moreover, businesses in the private sector use the data to determine business risks and opportunities. However, the data are only as good as the information we collect from sampled addresses.

The current design of the ACS, with a sample of roughly 3.5 million housing unit addresses, allows the Census Bureau to collect and update demographic, social, economic, and housing data for the United States every year. These data are essentially the same as the "long-form" data that the Census Bureau traditionally collected once a decade from a sample of housing units as part of the decennial census and ended with the 2000 Census. Response to the ACS is mandatory, because it is part of the decennial census.

A previous study indicated that the use of mandatory language on envelopes produces higher response rates over envelopes without the use of such language (Dillman et al., 1996). In addition, a recent study conducted on behalf of the Census Bureau suggested that the "required by law" message in the ACS mailing package may be the single most effective message in attracting attention and motivating response (Reingold, 2014). Prior to August 2015, the ACS sent up to six mailings to sampled addresses, and most of those materials contained language referring to the mandatory nature of the survey (See Table 1, Section 2.3). The envelopes that contain the initial mail package and paper questionnaire package also include mandatory language. Both envelopes display the following message in bold text, "**YOUR RESPONSE IS REQUIRED BY LAW**" (Attachment A).

Stakeholders and respondents have raised concerns about the prominent references in the mail materials to the mandatory participation in the ACS. In order to be responsive to concerns about the prominence of the mandatory message on the envelopes, we conducted the 2015 Envelope Mandatory Messaging Test.

The purpose of the Envelope Mandatory Messaging Test was to study the impact on self-response of removing the phrase "**YOUR RESPONSE IS REQUIRED BY LAW**" from the envelopes used to mail the initial mail package and the paper questionnaire package (Attachment B). In order to field this test as quickly as possible, materials within these packages, as well as other materials sent to sampled addresses, were not changed for this test. A more extensive test, conducted in September 2015, focused on modifications to the mail materials to soften the mandatory messages while emphasizing the benefits of participation in the survey (Oliver, et al., 2016).

## 2. Methodology

This report addresses the following research questions: What is the impact on response of removing the message, "**YOUR RESPONSE IS REQUIRED BY LAW**" from the two mailing envelopes used for the initial and paper questionnaire packages? What is the relative impact on cost and reliability of survey estimates of removing the mandatory message from the envelopes?

### 2.1. Experimental Design

For the experimental treatment, the mandatory language was removed from the envelopes used to mail the initial mail package and the paper questionnaire package. We will refer to the experimental treatment group as the Test Group. The control treatment received the same mail materials as production (envelopes with the mandatory language still included), but were sorted separately from the Test Group and the rest of production to ensure that the mail delivery time was as close as possible to the Test Group mail delivery time. We will refer to this group as the Control Sort Group.[1]

### 2.2. Sample Design

The ACS sample design consists of randomly assigning each monthly sample panel into 24 groups of approximately 12,000 addresses each. Each group within a monthly sample is representative of the full monthly sample. Each monthly sample is a representative subsample of the entire annual sample and is representative of the sampling frame. We chose two randomly selected groups each for our Control Sort and Test Group treatments because we also wanted to test mail delivery timing issues. The remaining 20 groups received production materials and were sorted as usual. There were approximately 24,000 addresses in the Test Group, approximately 24,000 addresses in the Control Sort Group, and approximately 240,000 addresses in the rest of the production sample.

### 2.3. ACS Operational Schedule for the May 2015 Panel

Each monthly ACS sample panel consists of three main data collection operations: a six-week mailout period during which the majority of Internet and mail responses are received, a one-month Computer-Assisted Telephone Interview (CATI) period, and a one-month Computer-Assisted Personal Interview (CAPI) period. Although the mailout period lasts only six weeks, mail and Internet responses, as well as responses received through telephone questionnaire assistance (TQA), are accepted until the end of the CAPI month.

The May 2015 panel mailout period was from April 23, 2015 to June 4, 2015. Table 1 shows the date of each mailing for the May 2015 panel and briefly describes the contents and purpose of each mailing. [2] Materials marked with a single asterisk (*) include language related to the

[1] We have found in past tests that the United States Postal Service (USPS) treats mailings differently based on the number of mailing pieces, so we created the Control Sort Group, which is the same size as the Test Group, to control for that. A separate report (Heimel, 2016), addresses research done in this area specific to the ACS May 2015 mailings.

[2] Prior to August 2015, the ACS had six mailings associated with each sample panel.

mandatory nature of the survey. Items marked with two asterisks (**) are the mailings where the message "**YOUR RESPONSE IS REQUIRED BY LAW**" was removed from the front of the envelope for the Test Group.

**Table 1. ACS Mailings for the May 2015 Panel**

| Mailing | Description of Materials | Mailout Date |
|---|---|---|
| Pre-Notice Letter | Letter informs housing unit they are in ACS, and tells them to expect a request to participate in a few days. Also contains a Multi-Lingual Informational Brochure*. | April 23, 2015 |
| Initial Mailing Package ** | Introduction Letter*, Frequently Asked Questions (FAQ) Brochure*, and Internet Instruction Card. This mailing urges housing units to respond via the Internet. | April 27, 2015 |
| First Reminder Postcard | A reminder postcard sent to all addresses that received the previous two mailings. | April 30, 2015 |
| Paper Questionnaire Package ** | Sent to addresses that have not responded via the Internet. Introduction Letter*, Paper Questionnaire, Return Envelope, Internet Instruction Card, FAQ Brochure*, and Instruction Guide*. | May 14, 2015 |
| Second Reminder Postcard | A reminder postcard* sent to all addresses that were also sent the Paper Questionnaire Package. | May 18, 2015 |
| Additional Postcard | An additional reminder postcard* sent to addresses that have not yet responded and are ineligible for CATI follow-up. | June 4, 2015 |

Note: The May 2015 panel CATI universe was created on May 25, 2015 and the CATI operation was conducted June 1-30, 2015. The May 2015 panel CAPI universe was created on June 29, 2015 and the CAPI operation was conducted July 1-31, 2015.

## 2.4. Impact on Response Analysis Design

For our analysis we calculated three types of rates: return rates, check-in rates, and response rates. Return rates and check-in rates measure the effects of the treatment on the survey operation and costs by including more cases in the denominator. Response rates measure more directly the effect on respondent behavior by removing from the denominator cases that could not have responded (vacant units, addresses that turn out not to be housing units, etc.)

Return rates were calculated for the self-response mode of the survey (Internet and mail) at selected points in the data collection cycle. Check-in rates are similar to self-response return rates; the differences are described in Section 2.4.1. They were calculated before the start of CATI to help determine data collection workloads for cost analysis. The final response rate was calculated after all modes of data collection were complete. The final response rate is reported as a distribution of the final rate by mode of response. Part of our analysis is to show what would most likely occur if the tested changes were implemented in a full ACS data collection cycle with respect to mail volume and potential workloads after the completion of the mailout phase.

We did not exclude out-of-scope addresses[3] from the self-response return rates, because most out-of-scope addresses were not discovered until the personal-visit stage of data collection. However, out-of-scope addresses were excluded from the final response rate calculations.

We calculated self-response return rates for all mailable addresses in our survey and among only those that were mailed the paper questionnaire package. We calculated check-in rates for all mailable addresses in our survey. After the CAPI operation was complete, we calculated the final response rate that included self-responses, CATI responses, and CAPI responses. For each calculation, we compared rates between the Test Group and the Control Sort Group. For each comparison, we used a two-tailed test so that we could measure the impact in either direction. Self-response return rates and final response rates were used to assess the impact on response of removing the mandatory message from the envelopes.

To assess the impact of the test on "hard-to-count" populations we compared response rates of the Control Sort Group to the Test Group by high and low tract-level response areas. We performed the following analysis on data collected from responses: response distributions for *race*, *Hispanic origin*, *age*, *tenure*, *educational attainment*, and *building type*; point estimates of *average household size*; and percentage of *limited English-Speaking households*. Self-responses included Internet and mail responses and total responses included Internet, mail, TQA, CATI, and CAPI responses. The calculation of rates and differences of rates between treatments used the methodology described in Section 2.4.4. The calculations of standard errors used the methodology described in Section 2.4.5.

### 2.4.1. Self-Response Return Rates and Check-In Rates

*Self-Response Return Rates*
We evaluated the impact of removing the mandatory message from the envelopes by calculating self-response return rates overall and by mode. The first set of self-response return rates included all mailable and deliverable sample addresses that were sent the initial mailing package. These rates were calculated at the following points in the data collection cycle: prior to the First Reminder Postcard mailing, prior to the Paper Questionnaire Package mailing, and prior to the start of the CATI operation. The second set of self-response return rates included all mailable and deliverable sample addresses that had not responded to the survey prior to the mailing of the paper questionnaire package. Only these addresses would have received the paper questionnaire package. These rates were calculated prior to the start of the CATI operation.

From both of these universes, we removed addresses where the initial mail package or the paper questionnaire package was returned by the United States Postal Service (USPS) as Undeliverable as Addressed (UAA) and a response was not received. We compared the rates between the Test Group and the Control Sort Group. For each comparison, we used a two-tailed test so that we could measure the impact in either direction.

[3] Some examples of addresses out-of-scope for the survey are: demolished home, home under construction, house or trailer moved, or the address is a permanent business or storage facility.

All self-response return rate comparisons were also broken out by mode (Internet and mail), and were calculated using the following formulae:

$$\text{Self-Response Return Rate} = \frac{\text{\# of mailable and deliverable sample addresses that provided a non-blank[4] return by mail, TQA, or a complete or sufficient partial response[5] by Internet}}{\text{Total \# of mailable and deliverable sample addresses[6]}} *100$$

$$\text{Internet Return Rate} = \frac{\text{\# of mailable and deliverable sample addresses that provided a complete or sufficient partial response[5] by Internet}}{\text{Total \# of mailable and deliverable sample addresses[6]}} *100$$

$$\text{Mail Return Rate} = \frac{\text{\# of mailable and deliverable sample addresses that provided a non-blank[4] return by mail or TQA}}{\text{Total \# of mailable and deliverable sample addresses[6]}} *100$$

*Check-In Rates*

Check-in rates are used to evaluate workloads for cost analysis purposes. In addition to Internet and TQA responses, these rates consist of all mail cases that are returned and processed, not just mail returns that are considered valid responses and are in-scope for the survey. These rates are very similar to self-response return rates for all households in the initial mailing universe with the following exceptions:

- Addresses determined to be UAA by the USPS are <u>not</u> omitted from the universe, because each mailing is an expense. Thus, every address that is sent a mailing is included.
- A mail return is considered to be checked-in if we received a questionnaire from the respondent at the Census Bureau's National Processing Center. A mail check-in is not verified to be a non-blank questionnaire, whereas a mail return must be non-blank.
- Check-in rates are not weighted because they are used to estimate workloads for analysis of data collection costs.

### 2.4.2. Response Rates, Response Distributions, and Other Analysis

At the end of all data collection, we calculated final response rates by combining the self-responses, CATI responses, and CAPI responses. These rates were calculated for the total number of sample addresses in the final response universe (described in Section 2.4.3.). We

[4] A blank form is a form on which there are no data defined-persons and the telephone number listed on the form by respondents is blank.

[5] A response is deemed a "sufficient partial" when the respondent reaches the first question in the detailed person questions section for the first person in the household.

[6] We removed addresses where the initial mail package or paper questionnaire package was returned by the USPS as UAA and a response was not received. For the second set of rates, the universe only includes sample addresses sent the paper questionnaire package.

compared the rates between the Test Group and the Control Sort Group. For each comparison, we used a two-tailed test so that we could measure the impact in either direction.

We also calculated final self-response and total response rates in high response and low response areas in order to assess the impact of the test methodology on hard-to-count populations. Using the Census Bureau's Planning Database (U.S. Census Bureau, 2015a), we defined these areas at the tract level. The Tract Level Planning Database is a database that uses Census Data to "assemble a range of housing, demographic, and socioeconomic variables that are correlated with mail nonresponse[7]. Tract levels are assigned a hard-to-count (HTC) score based on the concentration of attributes found in that tract that make enumeration difficult" (Bruce et al., 2008). In 2014, a *Low Response Score* replaced the HTC score (U.S. Census Bureau, 2015b). Tracts with the highest *low response scores* from the database were designated as low response areas, set in such a way that twenty-five percent of the addresses in the population were in low response areas. The remaining tracts were designated as high response areas. Each address was assigned to either the high response area or the low response area based on their tract. The addresses in our sample assigned to the low response area were those that would have a lower propensity to respond to the survey, and, as such, are considered hard-to-count. We compared the response rates for the Test Group and the Control Sort Group in each response area and then compared the difference between those differences. We wanted to see if the test impacted the low response areas more than the high response areas.

To further study the effect of this test on hard-to-count populations, we analyzed the responses to various ACS questions. The questions we used were determined by using some of the variables found on the Planning Database (Bruce et al., 2008)[8]. To determine whether the test affected response distributions, we calculated the distributions of all non-blank responses for the following categories: race, Hispanic origin, age, tenure, educational attainment, and building type. Only responses that were determined to be valid in-scope responses were included in this analysis. We chose only four categories for our analysis on *race* response distribution. If the only box checked was "White", the response was coded as "White alone". If the only box checked was "Black or African American", the response was coded as "Black or African American alone". If any other response was received, either by a check box or a write-in without a check box, and it was the only response received, the response was coded as "Other race alone". If two or more answers were received involving any combination of check boxes or write-ins, then we coded the response as "Two or more races".

In our calculations, we separated the distributions by mode: one for mail and one for Internet. We also calculated combined self-response (Internet and mail combined) and combined total response (Internet, mail, TQA, CATI, and CAPI combined). We wanted to see if the CATI and CAPI interviews made up for any difference we may have seen between Control Sort and Test treatments from the self-response modes. We used chi-square tests of independence to determine whether or not the response distributions were statistically different at the $\alpha$=0.1 level. If the

[7] The Census will be conducted in 2020 with an Internet self-response option, but up until that point the only self-response option has been mail.

[8] A list of all variables used for the 2015 Planning Database can be found on the website in the reference section (U.S. Census Bureau, 2015b).

distributions were significantly different, we performed additional testing on the differences between the Control Sort and Test treatments for each subcategory. To control for the overall Type I error rate for a set of hypotheses tested simultaneously, we performed multiple-comparison procedures. The overall Type I error rate is called the familywise error rate and is the probability of making one or more Type I errors among all hypotheses tested simultaneously. We used the Hochberg method to control for the familywise error rate at the $\alpha$=0.1 level. A *family* for our analysis was the list of p-values for the overall characteristic categories (*race, Hispanic origin, building type, tenure, age,* and *educational attainment*) and the list of p-values for a characteristic's subcategories if the response distributions were found to be statistically significantly different.

Because the rates for the subcategories for a given demographic variable derive from the same base, a shift in the number of responses in one subcategory may significantly change response in another. For this reason we also examined the percent changes in the weighted counts of responses for each overall category and the corresponding subcategories in the Test treatment from the corresponding weighted counts in the Control Sort treatment. These calculations can be used to help interpret the results from the response distribution tables. When the difference in response distributions is statistically significant, we can inspect the percent changes in weighted response counts. Note that these counts are not estimates for the categories or subcategories. The weighted counts are merely a means to help gain insight into whether or not a proportional shift in response distribution is related to the amount of response from that category for the Test Group as compared to the Control Sort Group.

We compared the weighted *average household size* among respondents in the Control Sort Group and the Test Group. Between the two groups, we also compared the weighted percent of respondent households that are considered to be *limited English-speaking households*. A *limited English-speaking household* is defined as a household in which all residents living in the household, ages 14 and older, speak a language other than English and report that they speak English less than very well.

A large portion of the relative cost impact of implementing the changes from this test in full production is the increase in CATI and CAPI workloads, caused by a decrease in self-response. CATI and CAPI are also the most burdensome modes of data collection for the respondents. Using paradata from the CATI and CAPI follow-up operations, we compared the average number of visits needed to complete an interview (get a response) in both modes between the Test Group and the Control Sort Group. We also calculated the distribution of attempts in both modes between the two groups. We used chi-square tests of independence to determine whether or not the response distributions were statistically different at the $\alpha$=0.1 level.

### 2.4.3. Universe Eligibility and Response Criteria

*Self-Response Universe Eligibility (Internet and Mail)*
For the self-response modes (Internet and mail), we excluded addresses designated as "undeliverable" (UAA) by the U.S. post office, unless we received a response. We also excluded addresses in remote Alaska and Puerto Rico. All addresses not excluded by one of these criteria were counted as eligible for both the Internet and mail modes.

*Response Criteria for Internet*

We counted a case as an Internet mode response if the address was in the self-response universe and one of the following conditions was satisfied:

- There was a complete Internet response.
- There was a sufficient partial Internet response. That is, the respondent viewed all basic demographic questions for all people in the household, all questions about the housing unit, and at least the first detailed question for one person and provided some data.
- There was an insufficient partial Internet response and we did not receive any other response from the address. An insufficient partial response indicates that the respondent started the Internet roster and may have reached the basic demographic portion of the survey, but did not reach the criteria for a "sufficient partial" response. This type of response was only included in the final response rates, not the return rates.
- The unit was suspected to be vacant, based on the Internet response received, but was not confirmed to be vacant during the mailout period of data collection. This was considered a response when calculating the return rates. Final response rates do not include confirmed vacants.
- The unit was suspected to be a business–which would have made it out of scope for the survey–but was not confirmed to be a business during the mailout period of data collection. This was considered a response when calculating the return rates. Final response rates do not include confirmed businesses as they are out-of-scope for the survey.

*Response Criteria for Mail*

We counted a case as a mail mode response if the address was in the self-response universe and we received either a complete mail response or a complete response via TQA. For the self-response modes, if more than one response was received from a single address, the response that was received first was considered the mode of response for this test.

*Self-Response Universe Eligibility and Self-Response Criteria for Those Mailed the Paper Questionnaire Package*

The universe eligibility and self-response criteria for the Paper Questionnaire Package mailing include the same criteria as the Initial Mailing Package, along with the following restrictions:

- Only those addresses that were mailed a paper questionnaire package were included in this universe.
- Housing units that responded via the Internet before they could possibly have received the paper questionnaire were not included in this universe unless they returned to the Internet instrument and provided a more complete response after the questionnaire was mailed to them.
- If we received a response in more than one self-response mode (Internet, TQA, or mail), we used the earliest response in our analysis. If the earliest response was received before the Paper Questionnaire mailing was sent, the response was used in the universe of the Initial Mailing only.

The intent of calculating rates with the above-mentioned responses and universe criteria is to quantify the impact on self-response of the removing the mandatory message from the envelopes specifically for questionnaire package mailing.

*CATI Universe Eligibility and Response Criteria*
Cases eligible for CATI were previous nonresponding addresses and a small subset of unmailable addresses (those with an undeliverable ZIP codes) for which we had telephone numbers. We counted a case as a CATI response if the address was in the CATI universe and, through a CATI interview, we obtained enough information for the response to be considered a complete or sufficient partial response. If we received a self-response after a CATI response, the self-response was considered the mode of response for this test.

*CAPI Universe Eligibility and Response Criteria*
The CAPI universe consisted of a subsample of all remaining nonresponding addresses (including unmailable and undeliverable addresses) after the CATI operation, because the CAPI operation is the most costly mode of data collection. To account for the subsampling of cases, we applied adjusted weights to all cases in the CAPI universe. We counted a case as a CAPI response if the address was in the CAPI universe and, through a personal interview, we obtained enough information for the response to be considered a complete or sufficient partial response. Addresses deemed out-of-scope for the survey during the interview were not considered to be responses. If we received a self-response after a CAPI response, the self-response was considered the mode of response for this test.

*Final Response Universe Eligibility and Response Criteria*
The final response universe included the self-response universe for the initial mailing, the CATI universe and the CAPI universe with the following exceptions:

- All addresses sampled out of the CAPI operation were excluded, unless we received a response via mail or Internet before the close of the May 2015 panel data collection operations. Those self-responses received a base weight, not an adjusted weight, as per standard ACS methodology.
- Unmailable addresses were not excluded, unless they were sampled out of CAPI.
- All cases that were determined to be out-of-scope during the CAPI operation were excluded from the universe. Some examples of out-of-scope addresses are demolished home, home under construction, house or trailer moved, or the address is a permanent business or storage facility.
- All cases confirmed to be businesses in telephone follow-up, telephone interviews, personal interviews, or TQA were excluded from the universe.

A final response, as determined by the criteria for each mode (self-response, CATI, or CAPI), were included in the calculation of the final response rate. If we received multiple responses from the same address, we chose the final response in the following order of importance: self-response, CATI, then CAPI. If we received more than one self-response from the same address, we chose the one that was received first. In addition, some households choose to respond via Internet or mail after they have completed a CATI or CAPI interview. We calculated the final response rate and how the rate was distributed among all response modes at the end of the CAPI month.

### 2.4.4. Calculation of Weighted Rates and Differences between Treatments

*Internet, Mail, and Total Self-Response*
The numerator is the sum of the base weights[9] of the cases determined to have a valid response in the Internet or mail mode. The denominator is the sum of the base weights of the cases determined to be in the self-response universe.

*Final Response Rate*
For the final response rate, we adjusted the weights of cases eligible for CAPI by applying the appropriate CAPI subsampling factors. Due to the high cost of CAPI, ACS sends only a subsample of the cases that have not responded in previous modes to this final mode of data collection. We assign a subsampling factor to account for those that are not included.[10] We used the CAPI subsampling factor to adjust the base weight of each case in the final response universe that was not a self-response or CATI response and was selected in the CAPI subsample. We applied the adjustment to both interviews and non-interviews in CAPI. We used the adjusted weights to calculate the final response rate in the same way we used the unadjusted weights to calculate the earlier self-response rates.

The numerator is the sum of the weights of the cases determined to have a valid self-response, CATI response, or CAPI response in the ways described in the sections above. The denominator for the final response rate, after CAPI, is the sum of the weights of the cases determined to be in the final response universe.

All numerators, denominators, and rates were calculated separately for the Test Group and the Control Sort Group. The difference between treatments was calculated as the Control Sort Group rate minus the Test Group rate. We used two-tailed hypothesis testing to determine whether the differences between the groups were statistically significant at the $\alpha = 0.1$ level.

### 2.4.5. Calculation of Standard Errors

The variances were estimated using the Successive Differences Replication (SDR) method with replicate weights, the standard method used in the ACS.[11] In calculating the return rates and response rates, we used the replicate base weights that account only for sampling probabilities. For each type of rate and treatment, we calculated the rate for the 80 half-sample replicates. Then, for each replicate, we calculated the difference between the Control Sort Group rate and the Test Group rate.

[9] The base weight for a sample unit is the inverse of the probability of selection for that unit.
[10] Chapter 4 of the ACS Design and Methodology document (U.S. Census Bureau, 2014) has details regarding the process of applying CAPI sampling rates.
[11] Chapter 12 of the ACS Design and Methodology document (U.S. Census Bureau, 2014) has details and references regarding the SDR method for variance estimation.

The variance for each rate and group, and each difference, was calculated using the formula:

$$Var(RR_0) = \frac{4}{80}\sum_{r=1}^{80}(RR_r - RR_0)^2$$

Where:

$RR_0$ = the return rate, response rate, or difference estimate calculated using the full sample base weights,

$RR_r$ = the return rate, response rate, or difference estimate calculated for replicate $r$

Finally, the standard error for an estimate is the square root of the variance.

## 2.5. Impact on ACS Estimates Analysis Design

If the test showed that response rates would be reduced by removing the mandatory messaging from the envelopes, we needed to also assess the impact on the estimates produced from the data collected, if the changes in the test were to be implemented in full ACS production. We calculated changes in the variances under a series of cost and sampling scenarios to assess the change in reliability of the ACS estimates. The two main scenarios, reducing initial sample size to maintain current cost levels or increasing cost levels (by increasing initial sample size) to maintain current survey reliability, are described in detail below in Section 2.6. Some additional analysis was also done to explore different ways of subsampling the interviewer-administered modes, CATI and CAPI, in order to maintain current survey costs and determine the effect on survey reliability.

## 2.6. Impact on Relative Cost Analysis Design

A reduction in response rates has the potential to impact costs for ACS data collection. The cost analysis conducted in this report looks broadly at past cost trends by mode and assumes similar costs per case. The only inputs used are costs and workloads. The methodology used to estimate workloads, costs, completed interviews, and sample size follows.

### 2.6.1. Estimating Workloads

For our analysis, we assumed a current annual sample of approximately 3.541 million cases. We used the workloads for the 2014 data collection year to determine the current production baseline workloads.

Due to the effects of postal sorting of treatments with differing sample sizes[12] on the delivery duration, we could not compare the test results for the Test Group directly with the workloads from the 2014 data collection year. Instead, we compared the Test and Control Sort treatments to evaluate the difference in workloads as a percentage of the sample size for each treatment. We applied that percentage difference to the estimated total annual ACS sample (3.541 million[13]),

[12] See Clark, 2015 for a more detailed explanation of this effect and Heimel, 2016 analysis from this test.

[13] Excluding group quarters, Puerto Rico, and Remote Alaska.

and then added the resulting projected differences (which we contend can be attributed solely to the test methodology) to the workloads from the 2014 ACS production data collection year.

### 2.6.2. Estimating Costs

The data collection cost per case for each mode was determined by dividing the workload for each mode by the Fiscal Year (FY) 2015 budget allocation for each mode. We assumed that costs per case for the test methodology would remain static (relative to current production costs) for each mode.

### 2.6.3. Estimating Completed Interviews

To estimate completed interviews, we used 2014 actual completes as the baseline. Similar to the calculations for estimating workloads, we then used check-in rates[14] from the 2015 Envelope Mandatory Messaging Test to determine the projected annual differences in completions by mode between the test and control treatments. We then applied these differences to the 2014 baseline completed interviews to determine projected numbers reflecting the test methodology.

### 2.6.4. Calculating Adjusted Sample Sizes

If the changes in this test were implemented in full ACS annual production, it would increase the annual ACS budget. We created a scenario that would help us maintain costs at the current FY 2015 data collection budget levels. To maintain costs, we had to reduce the initial sample to stay within the FY 2015 data collection budget. To do this, we used data collection cost per case, by mode, together with projected changes to the percent of the sample Test Group workload to solve for an initial sample size.

The formula for calculating this sample size is:

$$\text{Adjusted Initial Sample Size} = \frac{\text{Total Data Collection Budget}}{M_{pct}{}^{*}M_C + T_{pct}{}^{*}T_C + P_{pct}{}^{*}P_C}\text{, where}$$

$M_{pct}$ = adjusted percent of total sample that is eligible for self-response
$M_C$ = cost per case for self-response operation
$T_{pct}$ = adjusted percent of sample in the telephone workload
$T_C$ = cost per case for telephone operation
$P_{pct}$ = adjusted percent of sample in the personal visit workload
$P_C$ = cost per case for personal visit operation.

[14] Check-in rates differ slightly from self-response return rates in that UAA cases are included in the universe and check-in rates are not weighted when they are used for cost analysis.

## 3. Assumptions and Limitations

### 3.1. Assumptions

1) A single ACS monthly sample panel is representative of an entire year (twelve panels) and the entire frame sample, with respect to both return rates and costs, as designed.
2) A single sample group (1/24 of the full monthly sample) is representative of the full monthly sample, as designed.
3) The cost analysis assumes that costs per case remain static for the test methodology. We also assume that total cost is correlated completely with the size of workloads for each mode—that is, we assume that all costs are variable, with no fixed costs. This is a reasonable assumption given the size of the ACS data collection budget—the costs that make up the largest portions of the data collection budget are variable.
4) For the analysis of the reliability of the estimates when a change in CATI subsampling is made, we assumed a uniform CATI subsampling rate across the country and that those records subsampled out of CATI would also not be eligible for CAPI.

### 3.2. Limitations

1) The 2015 Envelope Mandatory Messaging Test did not include group quarters or Puerto Rico addresses, thus these estimates apply only for data collection in U.S. housing units.
2) We can only use the results from the test to make relative conclusions between the control and test treatments, because the control treatment in the 2015 Envelope Mandatory Messaging Test does not directly match to ACS production methodology (due to the differential sorting of packages for postal delivery). We are assuming these differences will hold proportionately in a production environment, but without a full-scale production test, we cannot know for sure.
3) The 2015 Envelope Mandatory Messaging Test only studied the effect of removing the "Your Response is Required by Law" message from the ACS initial and paper questionnaire package envelopes. It did not attempt to evaluate a fully voluntary ACS. All other materials and communications with respondents maintained the messaging that is currently used in ACS production, which emphasizes the mandatory nature of the survey (as we are required by law to disclose). Therefore, this report serves to evaluate only the impact of removing the mandatory message from the ACS envelopes. It is not a proxy for a test of a fully voluntary ACS.
**4)** Results from postal tracking analysis have indicated that there were significant differences in the distribution of delivery times between the Control Sort and Test group for both the initial mailing and the replacement mailing. (Heimel, 2016)

## 4. Results

### 4.1. Return Rates and Response Rates

In this section, we answer the research question: What is the impact on response of removing the message "**YOUR RESPONSE IS REQUIRED BY LAW**" from the two mailing envelopes for the initial and paper questionnaire packages?

The tables in this section provide detailed information about the return rates and response rates calculated for this test. For each table the standard errors are in parentheses next to each rate. We performed two-tailed hypothesis testing for each rate difference to determine whether the difference was statistically significant or not. A "Yes" or "No" in the tables denotes whether or not each difference is statistically significant at $\alpha = 0.1$ level.

Table 2 shows the self-response return rate results at various points in the data collection cycle and by mode (Internet and Mail) for all mailable and deliverable sample addresses in the ACS May 2015 panel. As is evidenced by the table, at all selected points in the data collection cycle up until CATI operations, the Test Group rate is significantly lower than the Control Sort Group rate. This is true regardless of response mode. For all mailable and deliverable sample addresses, the total self-response return rate for the Control Sort Group was 42.5 percent before CATI operations began. At this same point in time, the Test Group had a self-response return rate of 37.1 percent. The difference between the Test Group and Control Sort Group was 5.4 percentage points and was statistically significant.

**Table 2. Self-Response Return Rate Results (in Percents) by Mode at Selected Points in the Data Collection Cycle (for All Mailable and Deliverable Sample Addresses)**

| **Total Self-Response (Internet & Mail combined)** | | | | |
|---|---|---|---|---|
| Point in Data Collection Cycle | Control Sort | Test | Difference | Significant? |
| Before First Reminder Postcard | 3.0 (0.2) | 2.6 (0.1) | 0.4 (0.2) | Yes |
| Before Paper Questionnaire Package* | 22.6 (0.4) | 18.7 (0.3) | 3.9 (0.5) | Yes |
| Before CATI | 42.5 (0.5) | 37.1 (0.4) | 5.4 (0.5) | Yes |
| **Internet** | | | | |
| Point in Data Collection Cycle | | | | |
| Before First Reminder Postcard | 3.0 (0.2) | 2.6 (0.1) | 0.4 (0.2) | Yes |
| Before Paper Questionnaire Package | 22.2 (0.4) | 18.5 (0.3) | 3.7 (0.5) | Yes |
| Before CATI | 28.5 (0.4) | 24.4 (0.3) | 4.1 (0.6) | Yes |
| **Mail*** | | | | |
| Point in Data Collection Cycle | | | | |
| Before CATI | 14.0 (0.3) | 12.7 (0.3) | 1.3 (0.3) | Yes |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: *These totals include responses via TQA. Minor additive discrepancies are due to rounding. Significant at $\alpha$=0.1 level. Standard errors are in parentheses.

Table 3 is similar to Table 2 except that the universe includes only those addresses that were sent the paper questionnaire (addresses for which we had not previously received a response via the Internet). Because this was the second envelope that had the mandatory message removed, we wanted to calculate the return rates for only those that were sent the paper questionnaire package. This table shows results similar to Table 2. The Test Group rate is significantly lower than the Control Sort Group rate. For all mailable and deliverable sample addresses that were mailed the paper questionnaire package, the total self-response return rate for the Control Sort Group was 26.2 percent before CATI operations began. At this same point in time, the Test Group had a self-response return rate of 23.0 percent. The difference between the Test Group and Control Sort Group was 3.2 percentage points and was statistically significant.

**Table 3. Self-Response Return Rate Results (in Percents) by Mode before CATI for All Sample Addresses Mailed the Paper Questionnaire**

| | Control Sort | Test | Difference | Significant? |
|---|---|---|---|---|
| **Total Self-Response (Internet & Mail combined)** | 26.2 (0.4) | 23.0 (0.3) | 3.2 (0.4) | Yes |
| **Internet** | 9.4 (0.3) | 8.1 (0.2) | 1.2 (0.3) | Yes |
| **Mail** | 16.8 (0.3) | 14.8 (0.3) | 2.0 (0.4) | Yes |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Minor additive discrepancies are due to rounding. Significant at α=0.1 level. Standard errors are in parentheses.

At the end of all data collection operations, we calculated the final response rate, including all response modes, for all sample addresses that were not rendered out of scope for this survey. Table 4 shows the final response rate and how the rate is distributed among all response modes at the end of the CAPI month. The final overall response rate and each rate by mode are still significantly lower for the Test Group compared to the Control Sort Group. The Control Sort Group had a final response rate of 96.2 percent at the end of CAPI and the Test Group had a final rate of 95.6 percent at the end of CAPI. The Test Group rate was significantly lower by 0.6 percentage points.

**Table 4. Final Response Rate and Response Distributions (in Percents) by Mode after CAPI, for All Sample Addresses in Scope**

| | Control Sort | Test | Difference | Significant? |
|---|---|---|---|---|
| **Final Overall Response** | **96.2 (0.2)** | **95.6 (0.2)** | **0.6 (0.3)** | **Yes** |
| **Internet** | 30.0 (0.4) | 26.5 (0.3) | 3.5 (0.6) | Yes |
| **Mail** | 21.3 (0.3) | 19.0 (0.4) | 2.3 (0.5) | Yes |
| **CATI** | 3.2 (0.1) | 4.1 (0.2) | -0.9 (0.2) | Yes |
| **CAPI** | 41.8 (0.5) | 46.1 (0.5) | -4.3 (0.8) | Yes |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Minor additive discrepancies are due to rounding. Significant at α=0.1 level. Standard errors are in parentheses.

### 4.1.1. Response Rate Analysis By Designated High and Low Response Areas

Table 5 shows calculations for total self-response rates (both Internet and mail responses combined) and final response rates. These rates were calculated for high and low response areas (as defined in Section 2.4.2.) for both the Control Sort Group and the Test Group. As expected, self-response rates were much lower in the low response areas than in the high response areas for both groups. Also as expected, self-response rates were statistically significantly lower for the Test Group than for the Control Sort Group in both high and low response areas for all rates except the total response rate in the low response area. The difference in self-response rates between the Control Sort Group and the Test Group in high response areas was 5.5 percentage points; it was 6.8 percentage points in low response areas. This is a difference of -1.3 percentage points with a standard error of 1.4 and was not statistically significant at the $\alpha$=0.1 level. Thus, when comparing the difference in self-response between the high and low response areas in the Control Sort to that in the Test treatment, there was not sufficient evidence to prove that there is a difference. The same results are shown below for the calculations of total response in both treatment groups and both response areas and the difference was not statistically significant.

**Table 5. Final Self-Response and Total Response Rates (in Percents) by Designated High and Low Response Areas**

| **Self-Response** | Control Sort | Test | Difference | Significant? |
|---|---|---|---|---|
| High Response Area | 56.9 (0.6) | 51.5 (0.6) | 5.5 (0.9) | Yes |
| Low Response Area | 37.1 (0.8) | 30.4 (0.9) | 6.8 (1.2) | Yes |
| Difference | 19.8 (0.9) | 21.1 (1.0) | -1.3 (1.4) | No |
| **Total Response** | | | | |
| High Response Area | 96.7 (0.3) | 95.9 (0.3) | 0.8 (0.4) | Yes |
| Low Response Area | 94.9 (0.5) | 95.0 (0.5) | -0.1 (0.7) | No |
| Difference | 1.8 (0.6) | 0.9 (0.6) | 0.9 (0.9) | No |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Minor additive discrepancies are due to rounding. Significant at $\alpha$=0.1 level. Standard errors are in parentheses.

### 4.1.2. Response Distributions and Other Analysis

*Response Distributions*
To determine whether the test affected response distributions, we calculated the distributions of all non-blank responses for the following categories: *race*, *Hispanic origin, age*, *tenure*, *educational attainment*, and *building type*. After using the Hochberg method to adjust for the familywise error rate of multiple comparisons at the $\alpha$=0.1 level, there was not sufficient evidence to prove a statistical difference between the Control Sort Group and the Test Group for any of the response distributions for the Internet responses that we analyzed. However, response distributions for two of the question items that we analyzed, *age* and *tenure*, showed some statistically significant differences in proportions of response between the Control Sort Group and the Test Group for mail responses, overall self-responses, and final total responses. The tables showing all of the results of the response distribution analysis can be found in Attachment C. This section will only include a discussion of the response categories that showed statistically different response distributions and a subset of tables found in Attachment C.

A study performed on 2005 ACS respondent data (Joshipura, 2008) showed that ACS respondents who own their homes have a higher propensity to respond by mail[15] than respondents who rent their homes. Also, households where all residents are 65 and older are more likely than other households to respond by mail. Furthermore, younger households (all under the age of thirty) or households with a child under the age of five have a higher tendency to respond by telephone interview or personal interview than the rest of the population. This information helps us to determine what to expect in terms of response rates for those demographics. However, we were interested in finding out how the removal of mandatory messaging on the envelopes affected these demographic groups and if there was a statistically significant difference between the Test Group and the Control Sort group in the proportion of responses in these categories.

**Table 6. Mail, Combined Self-Response, and Combined Total Response Distributions (in Percents) for *Age*: Control Sort Versus Test Treatment ($\alpha$=0.1)**

| | **Mail Mode** | | **Self-Response (Internet and Mail)** | | **Total Response (Internet, Mail, TQA, CATI, and CAPI)** | |
|---|---|---|---|---|---|---|
| | Control Sort | Test | Control Sort | Test | Control Sort | Test |
| **AGE** (*p*-value) | - | **0.01** | - | **<0.01** | - | 0.93 |
| Under 5 years old | 4.0 (0.3) | 3.8 (0.3) | 5.1 (0.2) | 4.8 (0.2) | 5.8 (0.2) | 5.7 (0.2) |
| 5 to 17 years old | 11.7 (0.5) | 11.5 (0.5) | 15.1 (0.3) | 14.4 (0.3) | 17.5 (0.3) | 17.1 (0.3) |
| 18 to 24 years old | 6.2 (0.3) | **5.0 (0.3)** | 7.0 (0.2) | **6.3 (0.2)** | 8.3 (0.2) | 8.0 (0.2) |
| 25 to 44 years old | 17.4 (0.4) | 16.2 (0.4) | 24.0 (0.4) | 22.9 (0.3) | 25.9 (0.3) | 25.8 (0.3) |
| 45 to 64 years old | 31.5 (0.6) | 31.0 (0.5) | 29.9 (0.5) | 30.6 (0.4) | 26.9 (0.4) | 27.3 (0.4) |
| 65 years old or older | 29.2 (0.7) | **32.6 (0.7)** | 18.9 (0.3) | **21.0 (0.3)** | 15.6 (0.3) | 16.1 (0.3) |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Rates in boldface indicate a significant difference between the rates of the Test treatment and the Control Sort treatment. Minor additive discrepancies are due to rounding. P-values have been adjusted using the Hochberg method to control for Type I familywise error rate. Standard errors are in parentheses.

**Table 7. Percent Change of Weighted Response Counts on *Age*, Self-Response and Total Response Distributions**

| | **Self-Response (Internet and Mail)** | | **Total Response (Internet, Mail, TQA, CATI, and CAPI)** | |
|---|---|---|---|---|
| | Control Sort (Weighted counts) | Percent Change | Control Sort (Weighted counts) | Percent Change |
| **AGE** | 1,067,568 | -12.3 (1.4) | 1,861,900 | -2.3 (1.4) |
| Under 5 years old | 54,656 | -18.0 (4.0) | 108,642 | -4.8 (4.6) |
| 5 to 17 years old | 160,683 | -16.1 (2.9) | 326,000 | -4.5 (2.9) |
| 18 to 24 years old | 74,594 | -21.0 (3.4) | 154,207 | -5.5 (4.3) |
| 25 to 44 years old | 256,454 | -16.4 (2.5) | 482,892 | -2.9 (2.2) |
| 45 to 64 years old | 318,977 | -10.1 (2.1) | 500,223 | -0.6 (2.2) |
| 65 years old or older | 202,205 | -2.6 (2.6) | 289,936 | 1.0 (2.5) |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: The percent change [(e-c)/c * 100] was derived from weighted counts for the demographic variable and its subcategories for the Control Sort (c) and Test (e) treatments. Standard errors are in parentheses.

[15] At the time of the study there was not an Internet option for self-response.

For the question on *age*, there are significant differences in response distributions for mail and combined self-response between the Control Sort and Test Groups (Table 6). These differences appear to be driven by the subcategories "65 years old or older" and "18 to 24 years old". The percent differences in weighted responses (Table 7) for self-response reinforce these results. The response distribution differences for the *age* category seen in self-response disappear when we look at the data after all modes of data collection, as seen in the Total Response column of Table 6.

**Table 8. Mail, Combined Self-Response, and Combined Total Response Distributions (in Percents) for *Tenure*: Control Sort Versus Test Treatment ($\alpha$=0.1)**

| | **Mail Mode** | | **Self-Response (Internet and Mail)** | | **Total Response (Internet, Mail, TQA, CATI, and CAPI)** | |
|---|---|---|---|---|---|---|
| | Control Sort | Test | Control Sort | Test | Control Sort | Test |
| **TENURE** (*p*-value) | - | **0.01** | - | **<0.01** | - | **0.07** |
| Owned with a mortgage | 42.5 (0.8) | 40.6 (1.0) | 49.5 (0.5) | 48.5 (0.6) | 43.1 (0.5) | **41.4 (0.7)** |
| Owned free and clear | 30.6 (0.8) | **34.5 (0.8)** | 24.3 (0.4) | **26.7 (0.5)** | 21.5 (0.4) | **23.2 (0.5)** |
| Rented | 24.4 (0.7) | 23.1 (0.8) | 24.3 (0.4) | 23.4 (0.5) | 33.2 (0.5) | 33.5 (0.6) |
| Occupied without payment of rent | 2.5 (0.3) | **1.8 (0.2)** | 2.0 (0.2) | **1.4 (0.1)** | 2.2 (0.2) | 1.9 (0.2) |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Rates in boldface indicate a significant difference between the rates of the Test treatment and the Control Sort treatment. Minor additive discrepancies are due to rounding. P-values have been adjusted using the Hochberg method to control for Type I familywise error rate. Standard errors are in parentheses.

**Table 9: Percent Change of Weighted Response Counts on *Tenure*, Self-Response and Total Response Distributions**

| | **Self-Response (Internet and Mail)** | | **Total Response (Internet, Mail, TQA, CATI, and CAPI)** | |
|---|---|---|---|---|
| | Control Sort (Weighted Counts) | Percent Change | Control Sort (Weighted Counts) | Percent Change |
| **TENURE** | 435,317 | -10.3 (1.3) | 416,998 | -0.6 (1.2) |
| Owned with a mortgage | 215,368 | -12.0 (1.8) | 315,502 | -4.5 (1.9) |
| Owned free and clear | 105,735 | -1.4 (3.0) | 157,812 | 7.2 (3.1) |
| Rented | 105,641 | -13.5 (2.7) | 242,926 | 0.4 (2.6) |
| Occupied without payment of rent | 8,573 | -35.7 (7.5) | 16,260 | -15.5 (9.2) |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: The percent change [(e-c)/c * 100] was derived from weighted counts for the demographic variable and its subcategories for the Control Sort (c) and Test (e) treatments. Standard errors are in parentheses.

For the question on *tenure*, there are significant differences in response distributions for mail and combined self-response between the Control Sort and Test Groups (Table 8). These differences appear to be driven by the subcategories "Owned free and clear" and "Occupied without payment of rent". The percent differences in weighted responses (Table 9) for self-response reinforce these results. The difference in the combined total response distribution on *tenure* is a response shift between "Owned with a mortgage" and "Owned free and clear" (Table 8). Upon inspection of the percent change in response (Table 9) we see an increase in response in the subcategory "Owned free and clear" and a decrease in the subcategory "Owned with a mortgage" as compared to the overall percent increase of response in the *tenure* category. However, ACS published data on *tenure* only distinguishes between owner and renter and therefore this shift is most likely not noteworthy.

The findings from the response distribution analysis may suggest that adults 65 years old or older, who typically own their homes without a mortgage, are more willing than younger adults (especially those ages 18 to 24) to self-respond to the ACS by mail if the mandatory message does not appear on the mailing envelopes. This could influence the estimates for both *age* and *tenure*, and potentially other related characteristics, if the changes in this test were implemented in full ACS production.

*Average Household Size and Limited English-Speaking Households*
We calculated the weighted *average household size* reported by respondents in occupied housing units for the Control Sort Group and the Test Group. The difference between the two groups was statistically significant (see Table 10) at the $\alpha$=0.1 level for self-response (non-blank Internet and mail responses) and for the final response (Internet, mail, TQA, CATI, and CAPI responses combined.) This difference may indicate that larger households are less likely to respond to the survey when the mandatory message is removed from the mailing package envelopes.

**Table 10. Comparisons of Average Household Size and Percentage of Limited English-Speaking Households**

| | Control Sort | Test | Difference | Significant? |
|---|---|---|---|---|
| Average Household Size (Number of People) | | | | |
| Combined Self-Response | 2.4 (0.01) | 2.36 (0.01) | 0.05 (0.02) | Yes |
| Combined Total Response | 2.52 (0.02) | 2.48 (0.01) | 0.03 (0.02) | Yes |
| Percentage of Limited English-Speaking Households (%) | | | | |
| Combined Self-Response | 2.2 (0.2) | 2.0 (0.2) | 0.2 (0.3) | No |
| Combined Total Response | 4.1 (0.3) | 4.6 (0.2) | -0.5 (0.3) | No |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Minor additive discrepancies are due to rounding. Significant at $\alpha$=0.1 level. Standard errors are in parentheses.

Another group we considered to be potentially affected by the removal of the mandatory language was *limited English-speaking households*. These are households in which all residents living in the household, ages 14 and older, speak a language other than English and report that they speak English less than very well. We used data collected from Internet, mail, TQA, CATI, and CAPI responses to determine which addresses were *limited English-speaking households*. We calculated the weighted percentage of *limited English-speaking households* in the Control Sort and Test Groups. The difference between the two groups was not statistically significant (Table 10) at the $\alpha$=0.1 level for either self-response or combined total response, thus there is not sufficient evidence to prove that removing the message from the envelopes had an impact on this category of respondent.

*Average Number of CATI and CAPI Contacts Needed to Obtain an Interview*
Using paradata from the CATI transaction files and CAPI Contact History Instrument (CHI) files, we compared the average number of visits needed to complete an interview in both modes between the Test Group and the Control Sort Group. We also calculated the distribution of attempts in both modes between the two groups.

Hypothesis testing did not show sufficient evidence to prove that there was a difference between the Control Sort Group and the Test Group concerning the average number of visits needed to complete an interview in either CATI or CAPI mode.

**Table 11. Average Number of Attempts to Complete an Interview (Unweighted)**

| | Control Sort | Test | Difference | Significant? |
|---|---|---|---|---|
| CATI Mode | 1.91 (0.05) | 1.96 (0.05) | -0.05 (0.07) | No |
| CAPI Mode | 2.65 (0.04) | 2.68 (0.04) | -0.03 (0.06) | No |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: Minor additive discrepancies are due to rounding. Significant at $\alpha$=0.1 level. Standard errors are in parentheses.

The distributions of the number of contact attempts in the CATI and CAPI modes are shown in Table 12. Although the difference is not statistically significant, nominally, in both modes, there seems to be a slightly higher number of contact attempts needed to complete an interview for the Test treatment compared to the Control Sort treatment. Note that these are unweighted percentages which were calculated using raw data that reflects actions recorded by interviewers and captured in the paradata files.

**Table 12: Distribution of Attempts (in Percents) to Obtain A Complete Interview in CATI and CAPI (Unweighted)**

| | **Number of CATI Attempts** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Treatment** (0.59) | **1** | **2** | **3** | **4** | **5** | **6+** | **Total** |
| Control Sort | 51.4 | 27.5 | 10.8 | 5.0 | 3.0 | 2.3 | **100.0** |
| Test | 48.6 | 28.4 | 12.7 | 5.0 | 2.5 | 2.8 | **100.0** |

| | **Number of CAPI Attempts** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Treatment** (0.93) | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8+** | **Total** |
| Control Sort | 38.1 | 24.8 | 14.6 | 8.5 | 4.8 | 2.9 | 2.1 | 4.2 | |
| Test | 37.8 | 25.3 | 13.8 | 8.2 | 5.3 | 3.0 | 2.1 | 4.4 | **100.0** |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: P-values are reported in parentheses.

## 4.2. Cost Analysis

In this section, we will answer the following research question: What is the relative impact on cost and reliability of survey estimates of removing the mandatory message from the two mailing envelopes?

As previously stated, the self-response and overall response rates for the Test Group were significantly lower than the Control Sort Group for this test. A reduction in self-response results in an increase in costs due to increased CATI and CAPI workloads. Table 13 outlines the cost per case for each mode in the current ACS design. The Personal Visit mode has by far the highest cost per case at $141.92. Telephone cases cost almost twice as much as mail and Internet cases, but still substantially less than the Personal Visit mode. Additionally, lower response rates have an impact on the reliability of the ACS estimates due to an increase in variance. In this section, we explore the impact of a reduction in response rates on expected workloads, costs, completed interviews, and changes in reliability of survey estimates.

**Table 13. Estimated Cost per Case by Mode**

| | 2014 ACS Workload (millions) | FY15 Budget (millions) | Approximate Cost per Case |
|---|---|---|---|
| **Mail/Internet** | 3.448 | $35.628 | $10.33 |
| **Telephone** | 1.158 | $21.919 | $18.93 |
| **Personal Visit** | 0.726 | $103.033 | $141.92 |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

We evaluated the impact of the Test Group treatment under three different scenarios: maintaining current initial sample size, maintaining current reliability, and maintaining current cost.

***Maintain current initial sample size***: This option applies the results from this test to a full year of ACS sample to evaluate the effect on the cost of using the test treatment methodology for an entire ACS data collection year. Maintaining the current sample size will result in higher data collection costs, as the drop in response rates for the paper and Internet modes would result in higher workloads for the more expensive interviewer-administered modes. We also evaluated the effect on the reliability of survey estimates, which would suffer compared to current levels due to a decrease in total responses (as a result of both nonresponse from self-response and an increase in cases eligible for CAPI subsampling).

***Maintain current reliability***: This option applies the results from the 2015 Envelope Mandatory Messaging Test to determine the initial sample size necessary to maintain the reliability achieved using current ACS methodology. We also evaluated data collection costs, which would increase due to the increased initial sample and workloads for all modes.

***Maintain current costs***: This option applies the results from the 2015 Envelope Mandatory Messaging Test to determine how much the initial sample size would need to decrease to collect ACS data using the test strategy within the FY 2015 budget. Stratification and current CAPI subsampling rates would be maintained. We also evaluated the impact on the reliability of the survey estimates.

### 4.2.1. Projected Workloads with Current Sample

Table 14, below, summarizes the projected workloads for a full year of data collection if the test treatment methodology were to be implemented with the same sample size currently used for ACS data collection. Workloads are projected to increase by about 2.9 percent for mail (due to fewer Internet responses before the Paper Questionnaire Package mailing), 9.7 percent for CATI, and 7.0 percent for CAPI, because we project a reduction in self-responses.

**Table 14. Projected Workloads For a Full Year of ACS Data Collection**

| | Current ACS Methodology - 2014 | | | Test Treatment Methodology | | |
|---|---|---|---|---|---|---|
| | Workload (Millions) | Completed Interviews (Thousands) | Percent Completed Interviews | Workload (Millions) | Completed Interviews (Thousands) | Percent Completed Interviews |
| Internet | 3.448 | 917 | 26.6 % | 3.448 | 799 | 23.2% |
| Mail | 2.820 | 665 | 23.6% | 2.902 | 607 | 20.9% |
| Telephone | 1.158 | 143 | 12.3% | 1.270 | 176 | 13.9% |
| Personal Visit | 0.726 | 591 | 81.3% | 0.777 | 639 | 82.2% |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

### 4.2.2. Projected Costs to Maintain Current Sample

Table 15 summarizes projected workloads and data collection costs associated with removing the mandatory messaging from the ACS envelopes. Lower rates of response, for both Internet and mail, increase the workloads of both the telephone and personal visit modes. Because these modes are more expensive than the self-response modes, we project that the overall cost of data collection for the ACS will increase by about $9.4 million (see Test Strategy – Maintain Current Sample columns in Table 15). This estimate reflects a full year of data collection at the projected workload levels and does not take into account initial costs resulting from increased staffing to handle the larger telephone and personal visit workloads.

**Table 15. Summary of Data Collection Workloads and Associated Costs for Various Test Strategies (In Millions)**

| | Production ACS | | Test Strategy Maintain Current Sample | | Test Strategy Maintain Current Reliability | | Test Strategy Maintain Current Costs | |
|---|---|---|---|---|---|---|---|---|
| | Workload | Cost | Workload | Cost | Workload | Cost | Workload | Cost |
| Initial Sample | 3.541 | -- | 3.541 | -- | 3.665 | -- | 3.345 | -- |
| Mail/Internet | 3.448 | $35.628 | 3.448 | $35.628 | 3.569 | $36.869 | 3.257 | $33.653 |
| Telephone | 1.158 | $21.919 | 1.270 | $24.039 | 1.311 | $24.804 | 1.200 | $22.707 |
| Personal Visit | 0.726 | $103.033 | 0.777 | $110.336 | 0.805 | $114.312 | 0.734 | $104.220 |
| Subtotal | -- | $160.580 | -- | $170.003 | -- | $175.985 | -- | $160.580 |
| Increase over 2014 | | $0 | | $9.423 | | $15.405 | | $0 |
| Percentage Increase over FY2015 Budget[1] | | 0.0% | | 4.1% | | 6.7% | | 0.0% |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test
Note: [1]The total FY 2015 budget for the ACS was $231,100,808.

### 4.2.3. Projected Costs and Workloads to Maintain Reliability

A reduction in self and telephone response leads to fewer total interviews due to subsampling that occurs during the selection of the workload for personal visits. The "Test Strategy – Maintain Current Reliability" columns of Table 15 summarize the cost impacts that would result from increasing the initial sample to a level that would allow us to maintain our current levels of reliability for survey estimates. To maintain current reliability, total interviews would need to increase slightly to account for the higher sampling weights associated with the interviews completed in the personal visit mode, which accounts for a higher proportion of completed interviews in the test methodology compared to current methods. We estimate that we would

need to increase the sample to 3.7 million addresses (a factor of 1.035) and that the associated costs would increase by about $15.4 million, a 6.7 percent increase over the FY 2015 budget. This cost increase represents only the projected increase in the cost of collecting the data and does not include the additional cost of hiring and training additional staff to support the increased workloads.

### 4.2.4. Projected Workloads and Impacts on Reliability of Survey Estimates to Maintain Current Budget

The "Test Strategy – Maintain Current Costs" columns in Table 15 summarize our projections for the workloads if we reduced the initial sample to maintain the current cost of data collection for the ACS. We calculated the workloads using the same methodology we used to calculate workloads for maintaining the current sample, but used the reduced initial sample size instead of the current sample size. Total cost for data collection will stay the same (although cost for each mode will shift), but there will be a large reduction in the total number of interviews completed. We estimate that to afford an ACS with the test strategy implemented, we would need to reduce the initial sample size to 3,345,000, a decrease of 196,000 cases (5.5 percent). This reduction would result in a loss of reliability with an estimated increase in variance of 8.8 percent (See Table 16 below).

Table 16 outlines the projected total completed interviews. We calculated the change in the number of completed interviews relative to the current ACS data collection strategy.

**Table 16. Completed Interviews and Reliability Measures Associated with the Test Strategy (In Millions)**

| | Production ACS | Test Strategy Maintain Current Sample | Test Strategy Maintain Current Reliability | Test Strategy Maintain Current Costs |
|---|---|---|---|---|
| Initial Sample | 3.541 | 3.541 | 3.665 | 3.345 |
| Expected completed interviews | 2.315 | 2.191 | 2.318 | 2.063 |
| Change in completed interviews | -- | -0.124 | 0.003 | -0.252 |
| Estimated increase in variance | -- | 3.5% | -- | 8.8% |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

### 4.2.5. Subsampling in Both CATI and CAPI Modes

Mentioned above, in detail, is a method of reducing the initial sample size to maintain survey costs. We also calculated the impact of maintaining the initial sample size, but instead subsampling CATI and adjusting CAPI subsampling, in order to maintain costs at the FY 2015 level. To determine the CATI sampling rate in order to maintain costs at the current initial sample size, we assumed a uniform CATI subsampling rate across the country and that those records subsampled out of CATI would also not be eligible for CAPI. By sampling CATI and adjusting CAPI subsampling rates, we maintain the workloads by mode for both CATI and CAPI and thereby maintain overall cost. Table 17 gives a summary of workload estimates and cost estimates for a strategy that maintains current CATI and CAPI workload levels and compares it to the strategy that decreased the initial sample size in order to maintain costs at FY 2015 levels.

**Table 17. Summary of Cost and Reliability Data Comparing Reduction of CATI Sample to Reduction of Initial Sample (In Millions)**

| | Production ACS | | Test Strategy Maintain Current Cost (Maintain CATI and CAPI workloads) | | Test Strategy Maintain Current Costs (Reduce Initial Sample) | |
|---|---|---|---|---|---|---|
| | Workload | Cost | Workload | Cost | Workload | Cost |
| Initial Sample | 3.541 | -- | 3.541 | -- | 3.345 | -- |
| Mail/Internet | 3.448 | $35.628 | 3.448 | $35.628 | 3.257 | $33.653 |
| Telephone | 1.158 | $21.919 | 1.158 | $21.919 | 1.200 | $22.707 |
| Personal Visit | 0.726 | $103.033 | 0.726 | $103.033 | 0.734 | $104.220 |
| Subtotal | -- | $160.580 | -- | $160.580 | -- | $160.580 |
| Increase over 2014 | | $0 | | $0 | | $0 |
| Expected completed interviews | 2.315 | | 2.181 | | 2.063 | |
| Change in completed interviews | - | | -0.134 | | -0.252 | |
| Estimated increase in variance | - | | 8.6% | | 8.8% | |

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

The test strategy to maintain all current workload levels would require subsampling CATI at 91.4 percent. This strategy would maintain the same CATI workload as 2014 production while also maintaining the current initial sample size. This method would also reduce the CAPI workload (that is already subsampled with the current ACS methodology) because addresses subsampled out of CATI would not be eligible for CAPI. The CATI subsampling alone would lower the CAPI workload below the 2014 levels and thus we would select a slightly higher percentage of nonrespondents, in this design, in order to maintain the CAPI workload. The bottom of Table 17 shows the impact on reliability of using subsampling in CATI and altering subsampling in CAPI to preserve the CATI and CAPI workloads as compared to reducing the overall sample size to maintain costs.

Maintaining the CATI and CAPI workloads, by altering ACS subsampling methodology in the manner described above, results in about an 8.6 percent increase in the variances (See Table 17). This is marginally better than the scenario that reduces the initial sample in order to maintain current survey costs, which increases the variances by about 8.8 percent. Although our analysis suggests that subsampling CATI and adjusting CAPI subsampling yields marginally better results than applying an overall sample size decrease, the 0.2 percentage point difference could be just within the error of the cost and workloads models used.

While the analyses for this report are somewhat high-level approximations, they suggest that further subsampling for CATI could mitigate approximately 0.5 to 0.8 percent of the impact of the removal of mandatory messaging from the mailing envelopes, as compared to the strategy that maintains survey costs by decreasing the initial sample size.

It is important to note that this is a relative impact. Accordingly, an estimate of poverty that was 10 percent +/- 3 percent would have its margin of error improve to be 10 percent +/- 2.985 percent with a 0.5 percent relative decrease in the margin of error. These two margins of error, as displayed on American Fact Finder to one decimal place, would be indistinguishable. Thus, to make further research in this area profitable, more meaningful reductions in the variance impact would be required to warrant the investment of time and resources to explore improvements in the overall sample design of the ACS.

## 5. Conclusion

The results of this test show that eliminating the phrase "**YOUR RESPONSE IS REQUIRED BY LAW**" from the initial mail package envelope and the paper questionnaire package envelope does significantly lower the self-response return rate by 5.4 percentage points, before the start of the CATI operation. After telephone and personal interviews, the proportion of the final response rate attributed to self-response in the Test Group was significantly lower than the Control Sort Group rate by 5.8 percentage points; the overall final response rate was significantly lower for the Test Group by 0.6 percentage points.

Decreases in self-response rates, especially during the first month of data collection, lead to large increases in workloads for the interviewer-administered modes, both of which cost substantially more compared to self-response. In particular, increases in the personal visit workload lead to extremely higher costs compared to the current ACS data collection strategy. Assuming no changes to the ACS sample size, we estimate that implementing the test strategy would lead to cost increases of $9.4 million. This represents a 4.1 percent increase relative to the ACS budget for FY 2015. In addition, the test strategy would decrease the reliability of survey estimates — total interviews would decrease by approximately 124,000 annually, and variances would increase by 3.5 percent.

We investigated several different scenarios on how a change in ACS sampling methodology would affect cost and reliability estimates if this test methodology was implemented in full production. In order to maintain current FY 2015 budget costs of data collection, initial sample sizes would need to be reduced by 5.5 percent, about 196,000 households, which would result in an increase of variance for survey estimates of about 8.8 percent. In an effort to maintain current levels of reliability, the initial sample size would need to increase by 3.5 percent, about 124,000 households, which would result in increased annual data collection costs of approximately $15.4 million, a 6.7 percent increase relative to the FY 2015 ACS budget. Our analysis suggests that subsampling for CATI and altering subsampling for CAPI yields marginally better results than applying an overall sample size decrease. More meaningful reductions in the variance impact would be required to warrant the investment of time and resources to explore improvements in the overall sample design of the ACS.

Our research lacks sufficient evidence to prove that this test had a statistically significant impact on response of hard-to-count populations. However, for mail responses and overall self-responses, there were significant differences in the response distributions of *age* and *tenure*. These differences could influence the estimates for these and other related characteristics, if the changes in this test were implemented in full ACS production.

While removing the mandatory messaging from the envelopes does soften the tone of the ACS mailing materials, we project that, ultimately, it would result in more households being in the CATI and CAPI modes of data collection. This would result in an increase in respondent burden, because telephone and personal visits are the most intrusive modes. Although, for this test, more households were sent to CATI and CAPI, once they were in those modes the contact history analysis we performed does not suggest that this test had a significant impact on the number of attempts needed to obtain a telephone or personal interview response.

This test is just the first step in our research to examine the impact of modifying the mandatory messages contained in the ACS mail materials. Additional testing conducted in connection with the September 2015 ACS sample will provide important insights on ways to revise the mandatory messages in a more comprehensive manner throughout the various mail pieces. The September test also included other messaging and design enhancements to continue to be responsive to the feedback from stakeholders and respondents while preserving the quality of the survey.

## 6. References


Bruce, A., Robinson, J.G. (2008) *Tract Level Planning Database with Census 2000 Data.* Washington, DC: U.S. Government Printing Office. Retrieved from http://www.census.gov/research/data/planning_database/2000/

Clark, S., Roberts, A., Tancreto, J., & Raglin, D. (2015). 2015 Replacement Mail Questionnaire Package Test. Washington, DC: U.S. Census Bureau. Retrieved from http://www.census.gov/library/working-papers/2015/acs/2015_Clark_02.html

Dillman, Don A., Eleanor Singer, Jon R. Clark, and James B. Treat (1996). "Effects of Benefits (Appeals, Mandatory Appeals, and Variations in Statements of Confidentiality) on Completion Rates for Census Questionnaires," Public Opinion Quarterly, Volume 60: 376-389.

Heimel, S. (2016). "Postal Tracking Research on the ACS May 2015 Panel," Washington, DC: U.S. Census Bureau.

Joshipura, Megha (2008). "2005 ACS Respondent Characteristics Evaluation," Washington, DC: U.S. Census Bureau.

Oliver, Broderick, Michael Risley, Edward Castro, and Andrew Roberts (2016). "2015 Summer Mandatory Messaging Test Preliminary Report," Washington, DC: U.S. Census Bureau.

Reingold, Penn Schoen Berland, Decision Partners, (2014). "American Community Survey Messaging and Mail Package Assessment Research: Cumulative Findings," Washington, DC: U.S. Census Bureau. Retrieved from http://www.census.gov/library/working-papers/2014/acs/2014_Walker_02.html

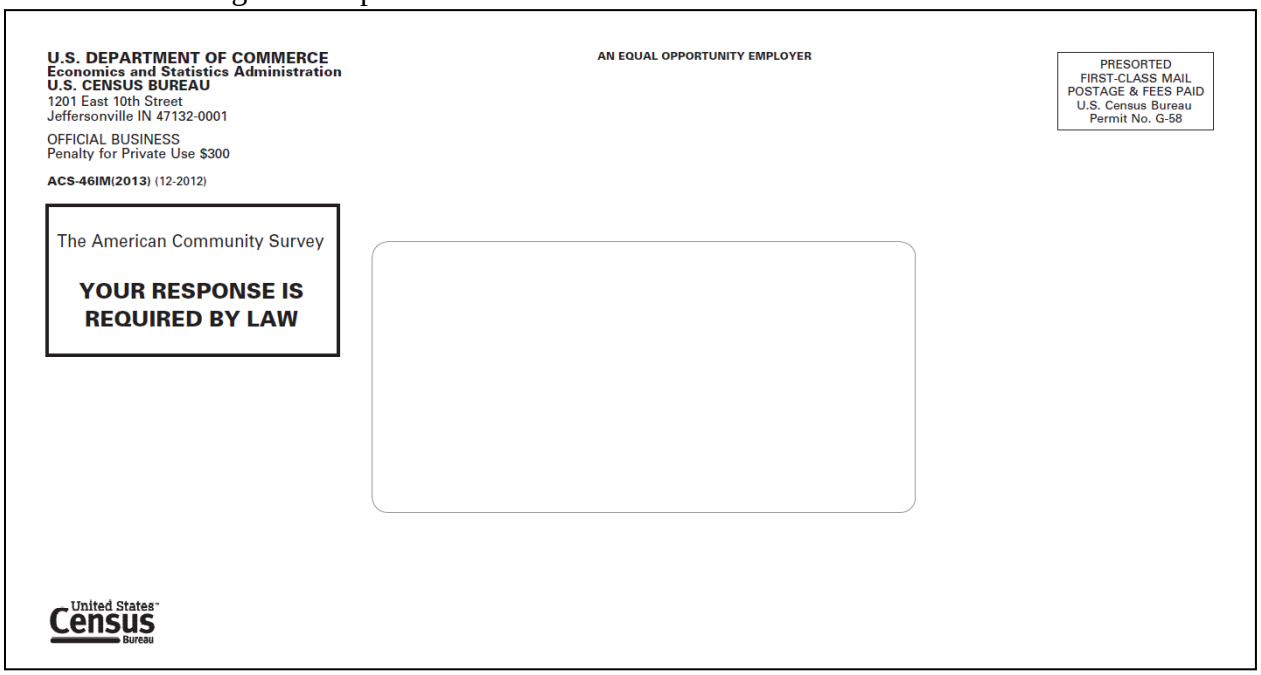
U.S. Census Bureau (2014), “American Community Survey Design and Methodology (January 2014)”, Retrieved from http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

U.S. Census Bureau (2015a). *2015 Planning Database Tract Data* [Data file]. Retrieved from http://www.census.gov/research/data/planning_database/2014/

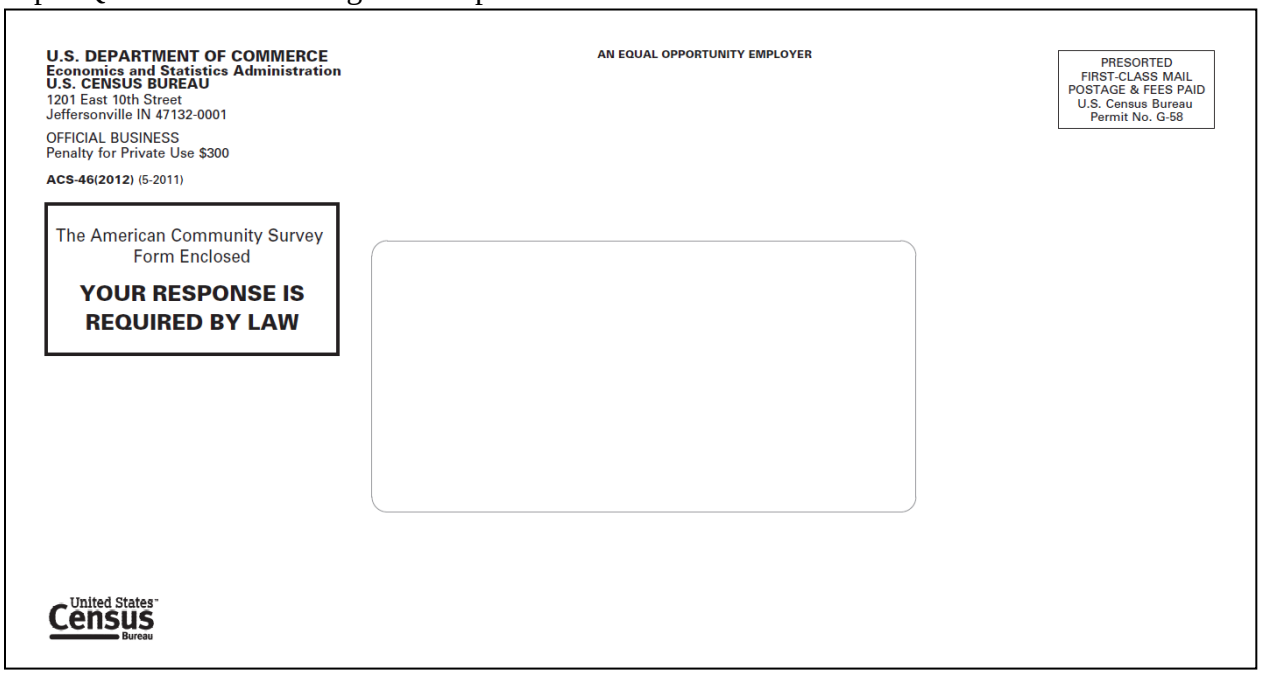
U.S. Census Bureau (2015b). Documentation on the 2015 Planning Database. Retrieved from http://www.census.gov/research/data/planning_database/2015/docs/PDB_Tract_2015-07-28a.pdf

**Attachment A. ACS Production Envelopes**
**May 2015 ACS Production Versions**

Initial Mail Package Envelope:

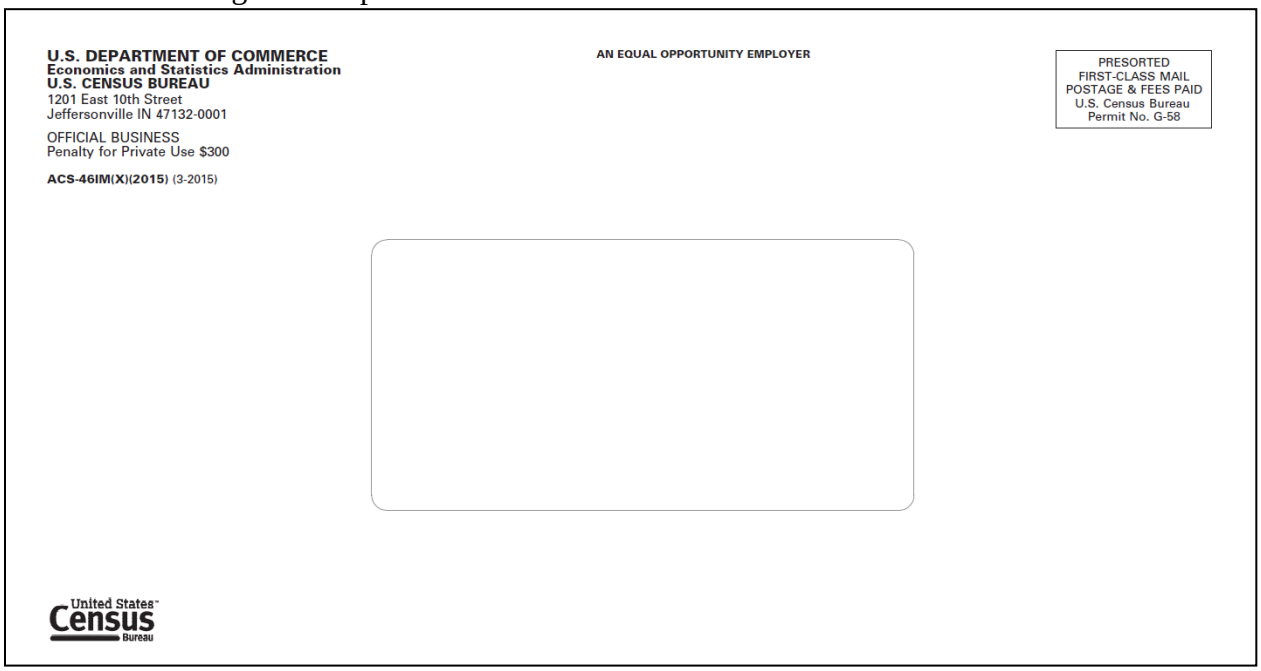



Paper Questionnaire Package Envelope:

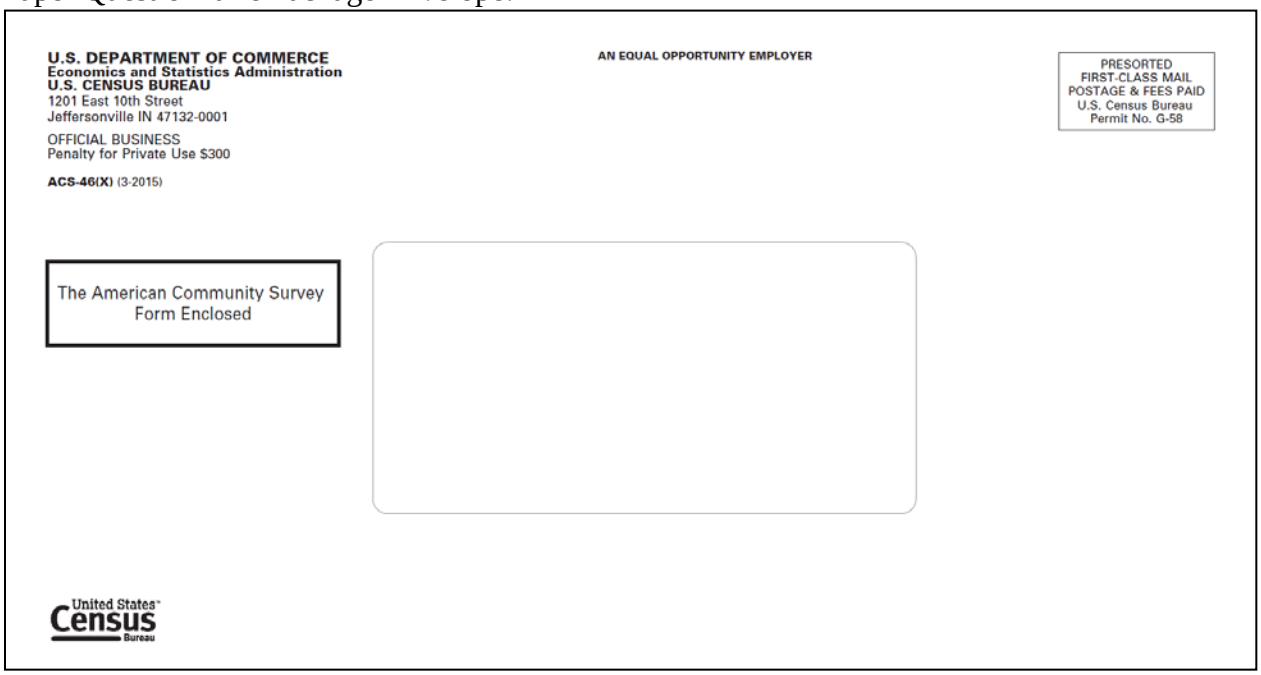


**Attachment B. ACS Test Treatment Envelopes**
**May 2015 ACS Test Versions**

Initial Mail Package Envelope:




Paper Questionnaire Package Envelope:



## Attachment C. Response Distribution Tables

### Table C-1: Self-Response Distributions (in Percents): Control Sort versus Test Treatment (Internet and Mail Modes) (α = 0.1)

| | Internet Mode | | Mail Mode | |
|---|---|---|---|---|
| **Item** | Control Sort | Test | Control Sort | Test |
| **RACE** (*p*-value) | - | 0.99 | - | 0.98 |
| White alone | 80.4 (0.6) | 80.6 (0.7) | 80.9 (0.8) | 80.9 (0.8) |
| Black or African American alone | 5.9 (0.3) | 5.8 (0.4) | 8.2 (0.5) | 8.0 (0.6) |
| Other race alone | 10.9 (0.5) | 10.8 (0.6) | 6.8 (0.5) | 7.0 (0.5) |
| Two or more races | 2.8 (0.2) | 2.7 (0.2) | 4.2 (0.4) | 4.1 (0.4) |
| **HISPANIC ORIGIN** (*p*-value) | - | 0.99 | - | 0.68 |
| Hispanic or Latino | 10.9 (0.5) | 10.7 (0.5) | 11.3(0.7) | 10.0 (0.6) |
| Not Hispanic or Latino | 89.1 (0.5) | 89.3 (0.5) | 88.7(0.7) | 90.0 (0.6) |
| **EDUC. ATTAINMENT** (*p*-value) | - | 0.99 | - | 0.98 |
| No schooling completed | 2.9 (0.2) | 2.8 (0.2) | 3.8 (0.3) | 3.6 (0.3) |
| Nursery to 11th grade | 18.7 (0.4) | 17.9 (0.5) | 14.9 (0.6) | 13.9 (0.6) |
| 12th grade (no diploma) | 1.1 (0.1) | 1.1 (0.1) | 2.3 (0.2) | 2.2 (0.2) |
| High school diploma | 13.3 (0.4) | 13.4 (0.4) | 25.0 (0.6) | 25.0 (0.6) |
| GED† or alternative credential | 2.1 (0.1) | 2.4 (0.2) | 4.4 (0.3) | 4.6 (0.3) |
| Some college | 16.7 (0.4) | 17.4 (0.4) | 19.8 (0.6) | 19.3 (0.5) |
| Associate's degree | 7.4 (0.2) | 7.6 (0.3) | 6.9 (0.3) | 7.8 (0.3) |
| Bachelor's degree | 22.3 (0.5) | 22.1 (0.5) | 14.2 (0.4) | 14.7 (0.6) |
| Advanced degree | 15.4 (0.4) | 15.2 (0.4) | 8.6 (0.4) | 8.9 (0.5) |
| **AGE** (*p*-value) | - | 0.10 | - | **0.01** |
| Under 5 years old | 5.7 (0.2) | 5.4 (0.2) | 4.0 (0.3) | 3.8 (0.3) |
| 5 to 17 years old | 16.9 (0.3) | 16.0 (0.4) | 11.7 (0.5) | 11.5 (0.5) |
| 18 to 24 years old | 7.5 (0.3) | 7.1 (0.3) | 6.2 (0.3) | **5.0 (0.3)** |
| 25 to 44 years old | 27.8 (0.5) | 26.8 (0.4) | 17.4 (0.4) | 16.2 (0.4) |
| 45 to 64 years old | 29.0 (0.6) | 30.4 (0.5) | 31.5 (0.6) | 31.0 (0.5) |
| 65 years old or older | 13.1 (0.4) | 14.4 (0.4) | 29.2 (0.7) | **32.6 (0.7)** |
| **BUILDING TYPE** (*p*-value) | - | 0.99 | - | 0.98 |
| One-family, detached | 69.4 (0.6) | 69.5 (0.8) | 67.5 (0.8) | 68.8 (0.8) |
| One-family, attached | 6.8 (0.4) | 6.8 (0.3) | 6.3 (0.4) | 5.9 (0.5) |
| 2 apartments | 2.5 (0.2) | 2.4 (0.2) | 2.7 (0.2) | 2.9 (0.3) |
| 3 or 4 apartments | 3.3 (0.3) | 3.1 (0.3) | 3.6 (0.3) | 3.1 (0.3) |
| 5 to 9 apartments | 3.8 (0.3) | 3.2 (0.3) | 3.3 (0.3) | 2.9 (0.3) |
| 10 to 19 apartments | 3.4 (0.3) | 4.0 (0.3) | 2.8 (0.3) | 3.0 (0.3) |
| 20 to 49 apartments | 2.8 (0.2) | 2.8 (0.3) | 2.7 (0.3) | 2.8 (0.3) |
| 50 or more apartments | 4.9 (0.3) | 5.1 (0.4) | 4.9 (0.4) | 4.2 (0.3) |
| Other (boat, van, etc.) | 3.0 (0.3) | 3.0 (0.3) | 6.3 (0.4) | 6.4 (0.5) |
| **TENURE** (*p*-value) | - | 0.38 | - | **0.01** |
| Owned with a mortgage | 54.0 (0.7) | 53.8 (0.8) | 42.5 (0.8) | 40.6 (1.0) |
| Owned free and clear | 20.2 (0.5) | 21.5 (0.6) | 30.6 (0.8) | **34.5 (0.8)** |
| Rented | 24.2 (0.5) | 23.5 (0.6) | 24.4 (0.7) | 23.1 (0.8) |
| Occupied without payment of rent | 1.6 (0.2) | 1.1 (0.1) | 2.5 (0.3) | **1.8 (0.2)** |

†General Educational Development

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

Note: Rates in boldface indicate a significant difference between the rates of the Test treatment and the Control Sort treatment. Minor additive discrepancies are due to rounding. P-values have been adjusted using the Hochberg method to control for Type I familywise error rate. Standard errors are in parentheses.

## Attachment C. Response Distribution Tables

### Table C-2: Self-Response and Total Response Distributions (in Percents): Control Sort versus Test Treatment (α = 0.1)

| | Self-Response (Internet and Mail) | | Total Response (Internet, Mail, TQA, CATI, and CAPI) | |
|---|---|---|---|---|
| **Item** | Control Sort | Test | Control Sort | Test |
| **RACE** (*p*-value) | - | 0.99 | - | 0.93 |
| White alone | 80.6 (0.5) | 80.7 (0.5) | 74.2 (0.5) | 73.5 (0.6) |
| Black or African American alone | 6.7 (0.3) | 6.6 (0.3) | 10.8 (0.3) | 11.1 (0.4) |
| Other race alone | 9.4 (0.4) | 9.5 (0.4) | 12.0 (0.4) | 12.4 (0.4) |
| Two or more races | 3.3 (0.2) | 3.2 (0.2) | 3.1 (0.2) | 3.1 (0.2) |
| **HISPANIC ORIGIN** (*p*-value) | - | 0.94 | - | 0.32 |
| Hispanic or Latino | 11.0 (0.4) | 10.5 (0.4) | 17.6 (0.5) | 19.0 (0.6) |
| Not Hispanic or Latino | 89.0 (0.4) | 89.5 (0.4) | 82.4 (0.5) | 81.0 (0.6) |
| **EDUC. ATTAINMENT** (*p*-value) | - | 0.94 | - | 0.93 |
| No schooling completed | 3.3 (0.2) | 3.1 (0.1) | 3.5 (0.2) | 3.4 (0.1) |
| Nursery to 11th grade | 17.4 (0.3) | 16.4 (0.3) | 24.0 (0.4) | 24.4 (0.4) |
| 12th grade (no diploma) | 1.5 (0.1) | 1.5 (0.1) | 1.2 (0.1) | 1.2 (0.1) |
| High school diploma | 17.5 (0.3) | 17.6 (0.3) | 19.5 (0.3) | 19.4 (0.3) |
| GED† or alternative credential | 2.9 (0.1) | 3.2 (0.2) | 2.9 (0.1) | 2.8 (0.1) |
| Some college | 17.8 (0.3) | 18.1 (0.3) | 17.8 (0.3) | 18.0 (0.3) |
| Associate's degree | 7.2 (0.2) | 7.7 (0.2) | 6.6 (0.2) | 6.6 (0.2) |
| Bachelor's degree | 19.4 (0.3) | 19.4 (0.3) | 15.6 (0.3) | 15.1 (0.3) |
| Advanced degree | 13.0 (0.3) | 12.9 (0.3) | 9.0 (0.2) | 9.1 (0.2) |
| **AGE** (*p*-value) | - | **<0.01** | - | 0.93 |
| Under 5 years old | 5.1 (0.2) | 4.8 (0.2) | 5.8 (0.2) | 5.7 (0.2) |
| 5 to 17 years old | 15.1 (0.3) | 14.4 (0.3) | 17.5 (0.3) | 17.1 (0.3) |
| 18 to 24 years old | 7.0 (0.2) | **6.3 (0.2)** | 8.3 (0.2) | 8.0 (0.2) |
| 25 to 44 years old | 24.0 (0.4) | 22.9 (0.3) | 25.9 (0.3) | 25.8 (0.3) |
| 45 to 64 years old | 29.9 (0.5) | 30.6 (0.4) | 26.9 (0.4) | 27.3 (0.4) |
| 65 years old or older | 18.9 (0.3) | **21.0 (0.3)** | 15.6 (0.3) | 16.1 (0.3) |
| **BUILDING TYPE** (*p*-value) | - | 0.94 | - | 0.93 |
| One-family, detached | 68.6 (0.5) | 69.2 (0.6) | 62.6 (0.4) | 62.0 (0.6) |
| One-family, attached | 6.6 (0.3) | 6.5 (0.3) | 6.0 (0.3) | 5.5 (0.2) |
| 2 apartments | 2.6 (0.2) | 2.6 (0.2) | 3.7 (0.2) | 3.9 (0.2) |
| 3 or 4 apartments | 3.4 (0.2) | 3.1 (0.2) | 4.6 (0.2) | 4.3 (0.2) |
| 5 to 9 apartments | 3.6 (0.2) | 3.1 (0.2) | 4.4 (0.2) | 4.8 (0.2) |
| 10 to 19 apartments | 3.2 (0.2) | 3.6 (0.2) | 4.2 (0.2) | 4.4 (0.2) |
| 20 to 49 apartments | 2.8 (0.2) | 2.8 (0.2) | 3.3 (0.2) | 3.4 (0.2) |
| 50 or more apartments | 4.9 (0.3) | 4.7 (0.3) | 4.5 (0.2) | 4.8 (0.2) |
| Other (boat, van, etc.) | 4.3 (0.2) | 4.4 (0.2) | 6.7 (0.3) | 7.0 (0.3) |
| **TENURE** (*p*-value) | - | **<0.01** | - | **0.07** |
| Owned with a mortgage | 49.5 (0.5) | 48.5 (0.6) | 43.1 (0.5) | **41.4 (0.7)** |
| Owned free and clear | 24.3 (0.4) | **26.7 (0.5)** | 21.5 (0.4) | **23.2 (0.5)** |
| Rented | 24.3 (0.4) | 23.4 (0.5) | 33.2 (0.5) | 33.5 (0.6) |
| Occupied without payment of rent | 2.0 (0.2) | **1.4 (0.1)** | 2.2 (0.2) | 1.9 (0.2) |

†General Educational Development

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

Note: Rates in boldface indicate a significant difference between the rates of the Test treatment and the Control Sort treatment. Minor additive discrepancies are due to rounding. P-values have been adjusted using the Hochberg method to control for Type I familywise error rate. Standard errors are in parentheses.

## Attachment C. Response Distribution Tables

### Table C-3: Percent Change of Weighted Response Counts, Self-Response and Total Response Distributions

| Item | Self-Response (Internet and Mail) | | Total Response (Internet, Mail, TQA, CATI, and CAPI) | |
|---|---|---|---|---|
| | Control Sort (Weighted Counts) | Percent Change | Control Sort (Weighted Counts) | Percent Change |
| **RACE** | 1,058,636 | -12.1 (1.4) | 1,849,779 | -2.1 (1.3) |
| White alone | 853,170 | -12.0 (1.6) | 1,374,241 | -3.0 (1.6) |
| Black or African American alone | 71,283 | -13.8 (5.5) | 199,805 | 0.5 (5.6) |
| Other race alone | 99,512 | -11.6 (5.4) | 222,694 | 1.1 (5.1) |
| Two or more races | 34,671 | -13.4 (8.3) | 56,547 | -1.5 (8.8) |
| **HISPANIC ORIGIN** | 1,045,959 | -12.1 (1.4) | 1,841,629 | -1.9 (1.3) |
| Hispanic or Latino | 115,265 | -16.2 (5.0) | 324,780 | 5.7 (4.7) |
| Not Hispanic or Latino | 930,695 | -11.5 (1.5) | 1,516,849 | -3.5 (1.5) |
| **EDUC. ATTAINMENT** | 942,975 | -12.6 (1.5) | 1,660,520 | -2.0 (1.5) |
| No schooling completed | 30,687 | -16.9 (5.9) | 57,382 | -4.9 (5.6) |
| Nursery to 11th grade | 163,874 | -17.4 (2.9) | 397,703 | -0.1 (2.9) |
| 12th grade (no diploma) | 14,559 | -14.6 (6.7) | 20,404 | -4.4 (8.9) |
| High school diploma | 164,763 | -11.8 (2.4) | 323,363 | -2.4 (2.4) |
| GED† or alternative credential | 27,702 | -4.6 (6.8) | 47,587 | -4.9 (6.0) |
| Some college | 168,041 | -11.2 (2.6) | 296,188 | -0.9 (2.9) |
| Associate's degree | 68,054 | -7.0 (3.7) | 108,793 | -1.3 (3.7) |
| Bachelor's degree | 183,114 | -12.7 (2.8) | 259,193 | -5.0 (2.6) |
| Advanced degree | 122,181 | -12.9 (3.2) | 149,907 | -1.1 (3.1) |
| **AGE** | 1,067,568 | -12.3 (1.4) | 1,861,900 | -2.3 (1.4) |
| Under 5 years old | 54,656 | -18.0 (4.0) | 108,642 | -4.8 (4.6) |
| 5 to 17 years old | 160,683 | -16.1 (2.9) | 326,000 | -4.5 (2.9) |
| 18 to 24 years old | 74,594 | -21.0 (3.4) | 154,207 | -5.5 (4.3) |
| 25 to 44 years old | 256,454 | -16.4 (2.5) | 482,892 | -2.9 (2.2) |
| 45 to 64 years old | 318,977 | -10.1 (2.1) | 500,223 | -0.6 (2.2) |
| 65 years old or older | 202,205 | -2.6 (2.6) | 289,936 | 1.0 (2.5) |
| **BUILDING TYPE** | 445,128 | -10.4 (1.3) | 847,988 | 0.5 (1.4) |
| One-family, detached | 305,565 | -8.6 (7.7) | 530,987 | -0.5 (1.9) |
| One-family, attached | 29,337 | -9.6 (1.6) | 50,915 | -8.7 (5.5) |
| 2 apartments | 11,460 | -12.1 (5.4) | 31,085 | 7.1 (7.8) |
| 3 or 4 apartments | 15,282 | -10.3 (8.5) | 39,093 | -6.7 (5.9) |
| 5 to 9 apartments | 16,172 | -18.5 (6.7) | 37,593 | 9.2 (7.2) |
| 10 to 19 apartments | 14,060 | -23.5 (5.9) | 35,688 | 4.3 (7.8) |
| 20 to 49 apartments | 12,347 | 2.4 (9.4) | 27,652 | 4.4 (8.4) |
| 50 or more apartments | 21,788 | -9.8 (7.9) | 38,025 | 6.5 (6.9) |
| Other (boat, van, etc.) | 19,117 | -13.8 (6.8) | 56,950 | 5.2 (6.6) |
| **TENURE** | 435,317 | -10.3 (1.3) | 416,998 | -0.6 (1.2) |
| Owned with a mortgage | 215,368 | -12.0 (1.8) | 315,502 | -4.5 (1.9) |
| Owned free and clear | 105,735 | -1.4 (3.0) | 157,812 | 7.2 (3.1) |
| Rented | 105,641 | -13.5 (2.7) | 242,926 | 0.4 (2.6) |
| Occupied without payment of rent | 8,573 | -35.7 (7.5) | 16,260 | -15.5 (9.2) |

†General Educational Development

Source: U.S. Census Bureau, American Community Survey, 2015 Envelope Mandatory Messaging Test

Note: The percent change [(e-c)/c * 100] was derived from weighted counts for the demographic variable and its subcategories for the Control Sort (c) and Test (e) treatments. Standard errors are in parentheses.