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

September 7, 2017

2017 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES # ACS17-RER-04

MEMORANDUM FOR	Victoria Velkoff Chief, American Community Survey Office
From:	David Waddington Chief, Social, Economic, and Housing Statistics Division (SEHSD)
Prepared by:	Christopher Mazur Social, Economic, and Housing Statistics Division (SEHSD)
Subject:	2016 American Community Survey Content Test Evaluation Report: Telephone Service

Attached is the final report for the 2016 American Community Survey (ACS) Content Test for Telephone Service. This report describes the results of the test for a revised version of the Telephone Service question.

If you have any questions about this report, please contact Ellen Wilson at 301-763-5789 or Christopher Mazur at 301-763-4103.

Attachment

cc:
Kathryn Cheza (ACSO)
Jennifer Ortman (ACSO)
Robert Sawyer (ACSO)
Patrick Cantwell (DSSD)
Elizabeth Poehler (DSSD)
Anthony Tersine (DSSD)
Nicole Scanniello (SEHSD)
Edward Welniak (SEHSD)
Ellen Wilson (SEHSD)

Intentionally Blank

American Community Survey Research and Evaluation Program

September 7, 2017

2016 American Community Survey Content Test Evaluation Report: Telephone Service

FINAL REPORT

Christopher Mazur
Social, Economic, and Housing Statistics Division

Andrew Roberts
Population Division

R. Chase Sawyer
American Community Survey Office

Intentionally Blank

TABLE OF CONTENTS

EXECUTIVE SUMMARY iv
1. BACKGROUND 1
1.1. Justification for Inclusion of Telephone Service in the Content Test 1
1.2. Question Development 2
1.3. Question Content 4
1.4. Research Questions 5
2. METHODOLOGY 6
2.1. Sample Design 6
2.2. Data Collection 7
2.3. Content Follow-Up 8
2.4. Analysis Metrics 8
2.4.1. Unit Response Rates and Demographic Profile of Responding Households 9
2.4.2. Item Missing Data Rates 11
2.4.3. Response Distributions 11
2.4.4. Benchmarks 11
2.4.5. Response Error 12
2.4.6. Standard Error Calculations 14
3. DECISION CRITERIA 14
4. LIMITATIONS 14
5. RESEARCH QUESTIONS AND RESULTS 16
5.1. Unit Response Rates and Demographic Profile of Responding Households 16
5.1.1. Unit Response Rates for the Original Content Test Interview 16
5.1.2. Unit Response Rates for the Content Follow-Up Interview 17
5.1.3. Demographic and Socioeconomic Profile of Responding Households 18
5.2. Item Missing Data Rates 20
5.3. Response Distributions 21
5.4. Benchmarks 22
5.5. Results for Analysis Specific to Telephone Service 23
5.5.1. Consistency between Telephone Service and Computer Type Questions 23
5.5.2. Consistency between Telephone Service Response with Valid Phone Number and Successful CFU Contact 24
6. CONCLUSIONS AND RECOMMENDATIONS 25
7. ACKNOWLEDGEMENTS 26
8. REFERENCES 27

Appendix A. Control and Test Versions of the Telephone Service Question in Internet Mode .. 28
Appendix B. NHIS Telephone Service Questions 29
Appendix C. Control and Test Versions of the Computer Type Question in Mail Mode 30
Appendix D. ACS FEFU Rules to Determine a Valid Phone Number 31
Appendix E. Unit Response Rates Supplemental Table 32
Appendix F. Consistency between Telephone Service Response with Valid Phone Number and Successful CFU Contact Supplemental Tables 33
Appendix G. Percent Distribution of Valid Telephone Service Question Responses to Other Information from Content Test 34

List of Tables

Table 1. Decision Criteria for Telephone Service 14
Table 2. Original Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode 16
Table 3. Mail Response Rates by Designated High (HRA) and Low (LRA) Response Areas 17
Table 4. Content Follow-Up Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode of Original Interview 18
Table 5. Response Distributions: Test versus Control Treatment 19
Table 6. Comparison of Average Household Size 19
Table 7. Comparison of Language of Response 20
Table 8. Item Missing Data Rate for Control and Test Treatment, by Mode 20
Table 9. Households Reporting Having Telephone Service, by Treatment and Mode 21
Table 10. Percent of Households without Telephone Service by Treatment and Benchmark 22
Table 11. Households Reporting No Telephone Service but Reported Having a Smartphone, by Treatment 23
Table 12. Households Reporting No Telephone Service that Provided a Valid Phone Number, by Treatment 24
Table 13. Households Reporting No Telephone Service that Provided a Valid Phone Number and Had a Successful CFU Contact, by Treatment 25
Table E-1. Unit Response Rates by Designated High (HRA) and Low (LRA) Response Areas. 32
Table F-1. Percent of Households that Reported Having Telephone Service that Provided a Valid Phone Number, by Treatment 33
Table F-2. Percent of Households that Reported Having Telephone Service that Provided a Valid Phone Number and Had a Successful CFU Contact, by Treatment 33
Table G-1. Respondent's Telephone Service and Smartphone Responses, by Treatment 34
Table G-2. Respondent's Telephone Service Response and Whether a Valid Phone Number was Provided, by Treatment 34

List of Figures

Figure 1. Current Production Telephone Service Question 4
Figure 2. Control Version of the Telephone Service Question 5
Figure 3. Test Version of the Telephone Service Question 5

Figure A-1. Control Version of the Telephone Service Question in the Internet Mode..... 28
Figure A-2. Test Version of the Telephone Service Question in the Internet Mode..... 28
Figure B-1. NHIS Question on Working Telephone Inside Home..... 29
Figure B-2. NHIS Question on Working Cell Phone 29
Figure C-1. Control Version of the Computer Type Question 30
Figure C-2. Test Version of the Computer Type Question..... 30

EXECUTIVE SUMMARY

Overview

From February to June of 2016, the U.S. Census Bureau conducted the 2016 American Community Survey (ACS) Content Test, a field test of new and revised content. The primary objective was to test whether changes to question wording, response categories, and definitions of underlying constructs improve the quality of data collected. Both new and revised versions of existing questions were tested to determine if they could provide data of sufficient quality compared to a control version as measured by a series of metrics including item missing data rates, response distributions, comparisons with benchmarks, and response error. The results of this test will be used to help determine the future ACS content and to assess the expected data quality of revised questions and new questions added to the ACS.

The 2016 ACS Content Test consisted of a nationally representative sample of 70,000 residential addresses in the United States, independent of the production ACS sample. The sample universe did not include group quarters, nor did it include housing units in Alaska, Hawaii, or Puerto Rico. The test was a split-panel experiment with one-half of the addresses assigned to the control treatment and the other half assigned to the test treatment. As in production ACS, the data collection consisted of three main data collection operations: 1) a six-week mailout period, during which the majority of self-response via internet and mailback were received; 2) a one-month Computer-Assisted Telephone Interview period for nonresponse follow-up; and 3) a one-month Computer-Assisted Personal Interview (CAPI) period for a sample of the remaining nonresponse. For housing units that completed the original Content Test interview, a Content Follow-Up (CFU) telephone reinterview was conducted to measure response error.

Telephone Service

A question collecting information on telephones or telephone service in a housing unit has been asked of occupied housing units beginning with the 1960 decennial census. The question has gone through several modifications since then. The need to revise the question in the past, and again now, comes from the evolution of telephone equipment, services, and usage. The rise of cell phone and smartphone usage, and other complex and varied telephone services and equipment, has changed how Americans view and use telephones in a household.

Research suggests that some respondents may not fully understand the current question. For example, the instruction singling out “cell phones” may have caused confusion that telephone service is not limited to cell phones, but that standard wired telephone equipment should also be included. This confusion likely led to anomalies in the estimates of households without telephone service in multiple counties.

Cognitive testing and consultation with topical experts led to the decision to make telephone service a stand-alone question, separated from its current placement in a question that asks about a series of physical components of a housing unit. For the Content Test, two versions of the stand-alone question were tested: one without any additional instruction (control treatment) and

one with an additional instruction to clarify what types of phones respondents should consider when answering the question (test treatment).

Research Questions and Results

This research was guided by questions concerning item missing data rates, response distributions, comparisons to benchmark data, and response reliability.

Item Missing Data Rates
Item missing data rates were compared between treatments on an overall basis and by each mode of data collection. Overall and in the mail mode, there was no statistical difference between the treatments. For the internet mode, the test treatment had a lower item missing data rate, while the control treatment had a lower item missing data rate for the CAPI mode.

Response Distributions
Overall, the proportion of cases with telephone service was statistically higher in the test treatment than it was in the control treatment. This result was also true in the mail and internet modes. The test treatment did not have a statistically higher proportion of cases with telephone service in the CAPI mode.

Benchmark Data
Benchmark comparisons were made to data from the March 2016 ACS production panel and the July-December 2015 National Health Interview Survey (NHIS) Wireless Substitution Early Release Program. Formal statistical comparisons were not made due to differences of methodology, question wording, and universes between the Content Test estimates and benchmark estimates. Instead, this analysis evaluated if the treatment estimates were grossly different from the benchmark data. Subjective analysis indicated the test treatment data were within an acceptable range after taking into consideration the factors that potentially could have led to any differences between the estimates.

Response Reliability
Most topics in the Content Test measured response error using metrics from the CFU. However, CFU was a telephone interview and required respondents to have telephone service. Therefore, response reliability for this topic consisted of evaluating when a respondent reported having no telephone service, but other information from the Content Test could possibly indicate otherwise.

First, the proportion of respondents that reported having no telephone service but provided a valid phone number in the original interview was evaluated. Of these respondents, it was then determined whether or not the Census Bureau was able to successfully contact them to partake in the CFU reinterview. If a successful contact was made, the household was identified as possibly having telephone service. There were no statistical differences between treatments in the rate of these inconsistencies.

Second, the proportion of respondents that reported having no telephone service but reported having a smartphone in the computer type question was evaluated. The test treatment had a statistically lower rate of this type of inconsistency. However, because different wording for the

computer type question was also tested in the Content Test, the response categories that referred to “smartphone” differed in each treatment and it is unclear to what extent the changes in the computer type and telephone service questions may have led to this inconsistency.

Conclusions

As telephone equipment, services, and usage change with the advancement of technology, it is imperative to periodically evaluate and update the question about telephone service. This research established that asking telephone service as a stand-alone question with additional instruction performed better than asking it as a stand-alone question without the additional instruction.

We believe that providing additional clarification in the test version, particularly in the self-response modes of data collection, improved respondents’ understanding that the type of equipment that could make or receive phone calls is not limited to cell phones or landlines, but other phone devices as well. As advancements in telephone technology continue to evolve, the test version’s inclusionary instruction could help alleviate future confusion as to the types of telephone equipment and devices respondents need to consider when answering the question.

Based upon the prioritized decision criteria set forth in the research, the test version of the question performed at an acceptable level and is the version of the telephone service question that we recommend implementing moving forward in the ACS.

1. BACKGROUND

From February to June of 2016, the Census Bureau conducted the 2016 American Community Survey (ACS) Content Test, a field test of new and revised content. The primary objective was to test whether changes to question wording, response categories, and definitions of underlying constructs improve the quality of data collected. Both revised versions of existing questions and new questions were tested to determine if they could provide data of sufficient quality compared to a control version as measured by a series of metrics including item missing data rates, response distributions, comparisons with benchmarks, and response error. The results of this test will be used to help determine the future ACS content and to assess the expected data quality of revised questions and new questions added to the ACS.

The 2016 ACS Content Test included the following topics:

- Relationship
- Race and Hispanic Origin
- Telephone Service
- Computer and Internet Use
- Health Insurance Coverage
- Health Insurance Premium and Subsidy (new questions)
- Journey to Work: Commute Mode
- Journey to Work: Time of Departure for Work
- Number of Weeks Worked
- Class of Worker
- Industry and Occupation
- Retirement, Survivor, and Disability Income

This report discusses Telephone Service.

1.1. Justification for Inclusion of Telephone Service in the Content Test

A question collecting information on telephones or telephone service in a housing unit has been asked of occupied housing units beginning with the 1960 census. The question has evolved as telephone equipment and service have changed. The 1960 and 1970 censuses collected data on telephone availability. A unit was classified as having a telephone available if a phone number could be used to contact the occupants of the unit. The telephone could have been in another unit, in a common hall, or outside the building. In 1980 and 1990, a telephone had to be inside the house or apartment for the unit to be classified as having a telephone available. Units where the respondent used a telephone located inside the building but not in the respondent's living quarters were classified as not having a telephone.

In 2000, the focus of the telephone question changed from an equipment-based question to a service-based question. This was necessary because a household could own a telephone but have no service to make or receive phone calls. The rise of cell phone and smartphone usage, and other complex and varied telephone services and equipment, led to a need to revise the question again as Americans shifted and continue to shift how they think about and use telephones.

The National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics, is conducted annually to collect information on health status, health-related behaviors, and health care access and utilization through in-person household interviews. The survey also collects information about telephone service and the presence of a working wireless telephone in a household for the civilian noninstitutionalized U.S. population. Estimates of telephone coverage are released twice each year through the NHIS Wireless Substitution Early Release Program, along with comparable estimates from the NHIS for the previous three years.

The July-December 2015 release of the Wireless Substitution Early Release Program indicated that 48.3 percent of households had only wireless telephones between July and December 2015. During the same six-month period between July and December 2012, 38.2 percent of households had only wireless telephones. The 10.1 percentage point increase in a three-year period indicated the continued rapid shift towards wireless-only households and the need to ensure respondents fully understand the telephone service question on the ACS (Blumberg & Luke, 2015).

In addition, internal Census Bureau review of ACS telephone service data in 2011, 2012, and 2015 identified multiple geographies in which there was an unexpected increase in the number of households with no telephone service from one year to the next. As a result, ACS 1-year and 5-year telephone service data were suppressed for these geographies. Analysis of the issue suggested that some respondents might not have fully understood the current wording of the instruction and what the question intended to capture. The instruction, singling out "cell phones," may have caused confusion that telephone service is not limited to cell phones, but that standard wired telephone equipment commonly found in many homes is sufficient to have in order to answer the question correctly.

1.2. Question Development

Initial versions of the new and revised questions were proposed by federal agencies participating in the U.S. Office of Management and Budget (OMB) Interagency Committee for the ACS. The initial proposals contained a justification for each change and described previous testing of the question wording, the expected impact of revisions to the time series and the single-year as well as five-year estimates, and the estimated net impact on respondent burden for the proposed revision.[1] For proposed new questions, the justification also described the need for the new data, whether federal law or regulation required the data for small areas or small population groups, if other data sources were currently available to provide the information (and why any alternate sources were insufficient), how policy needs or emerging data needs would be addressed through the new question, an explanation of why the data were needed with the geographic precision and frequency provided by the ACS, and whether other testing or production surveys had evaluated the use of the proposed questions.

The Census Bureau and the OMB, as well as the Interagency Council on Statistical Policy Subcommittee, reviewed these proposals for the ACS. The OMB determined which proposals moved forward into cognitive testing. After OMB approval of the proposals, topical

[1] The ACS produces both single and five-year estimates annually. Single year estimates are produced for geographies with populations of 65,000 or more and five-year estimates are produced for all areas down to the block-group level, with no population restriction.

subcommittees were formed from the OMB Interagency Committee for the ACS, which included all interested federal agencies that use the data from the impacted questions. These subcommittees further refined the specific proposed wording that was cognitively tested.

The Census Bureau contracted with Westat to conduct three rounds of cognitive testing. The results of the first two rounds of cognitive testing informed decisions on specific revisions to the proposed content for the stateside Content Test (Stapleton and Steiger, 2015). In the first round, 208 cognitive interviews were conducted in English and Spanish and in two modes (self-administered on paper and interviewer-administered on paper). In the second round of testing, 120 cognitive interviews were conducted for one version of each of the tested questions, in English and Spanish, using the same modes as in the first round.

A third round of cognitive testing involved only the Puerto Rico Community Survey (PRCS) and Group Quarters (GQ) versions of the questionnaire (Steiger, Anderson, Folz, Leonard, & Stapleton, 2015). Cognitive interviews in Puerto Rico were conducted in Spanish; GQ cognitive interviews were conducted in English. The third round of cognitive testing was carried out to assess the revised versions of the questions in Spanish and identify any issues with questionnaire wording unique to Puerto Rico and GQ populations.[2] The proposed changes identified through cognitive testing for each question topic were reviewed by the Census Bureau, the corresponding topical subcommittee, and the Interagency Council on Statistical Policy Subcommittee for the ACS. The OMB then provided final overall approval of the proposed wording for field testing.[3]

Cognitive testing and question development specific to telephone service led the topical interagency subcommittee on telephone service to make several recommendations for the final two versions of the question for the Content Test. Prior to 2008, the telephone service question was a stand-alone question in the ACS. Beginning in 2008, in its current form, telephone service is one of six items in a single question that also asks if a respondent's housing unit has certain plumbing and kitchen components. With a greater sense that telephone equipment is not necessarily a fixed component of a housing unit, the subcommittee decided the current structure of how telephone service is asked is difficult for respondents to grasp. The subcommittee decided to separate the telephone service question from the plumbing and kitchen components series and make it a stand-alone question for both the control and test treatments in order to create a smoother question structure and make the intent of the question easier to understand by respondents.

The remaining issue was to determine whether the question should include an additional instruction after the question wording to give more clarity to respondents on what types of telephone equipment can be considered when answering the question. As discussed in Section 1.1, evidence suggests some respondents misunderstand the current instruction "Include cell phones" because the wording is exclusive to one type of telephone. Therefore, the subcommittee wanted to ensure the instruction was inclusive of multiple types of telephones and also short and simple to understand. The instruction is intended to reduce respondent burden by providing

[2] Note that the field testing of the content was not conducted in Puerto Rico or in GQs. See the Methodology section for more information.

[3] A cohabitation question and domestic partnership question were included in cognitive testing but ultimately we decided not to move forward with field testing these questions.

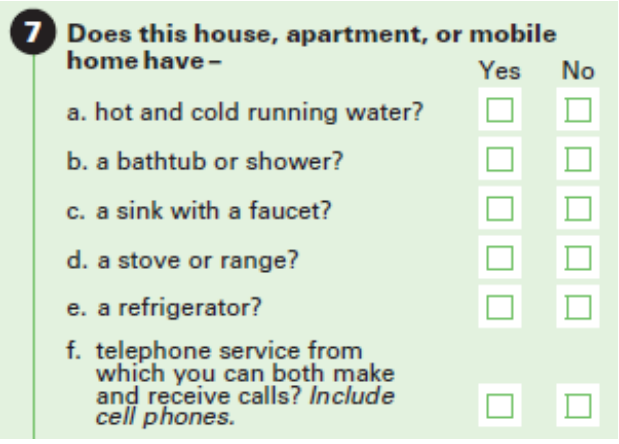
clarity on the types of devices and equipment that are acceptable to correctly answer the question as to whether or not someone in a household has the ability to make and receive phone calls. The subcommittee decided to test a version of the revised question wording that included the revised instruction "Include calls using cell phones, land lines, or other phone devices."

As a result of these decisions, and the strong belief that the telephone service question needed to be a stand-alone question to reduce respondent burden, two versions of the question were tested in the Content Test, neither of which is the current ACS production question.

1.3. Question Content

The current production version of the telephone service question was not used in the Content Test for this topic. To gain a better understanding of the differences of the current ACS production version of the telephone service question and the two versions used in the Content Test, an image how telephone service is currently asked is shown in Figure 1 as it appears in the mail mode.

Figure 1. Current Production Telephone Service Question



7 Does this house, apartment, or mobile home have –

	Yes	No
a. hot and cold running water?	☐	☐
b. a bathtub or shower?	☐	☐
c. a sink with a faucet?	☐	☐
d. a stove or range?	☐	☐
e. a refrigerator?	☐	☐
f. telephone service from which you can both make and receive calls? *Include cell phones.*	☐	☐

The two versions of the telephone service question used in the Content Test are shown in Figures 2 and 3 as they appeared in the mail mode. Both the control and test versions of the question have the same question wording. However, the test version contains the additional instruction on types of telephones and equipment respondents should consider when answering the question.

Figure 2. Control Version of the Telephone Service Question

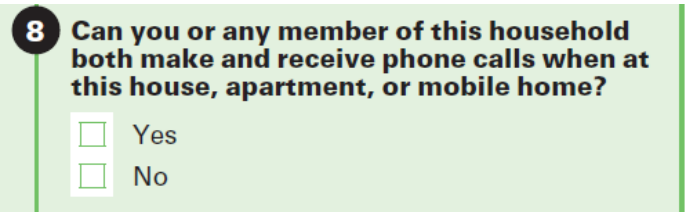
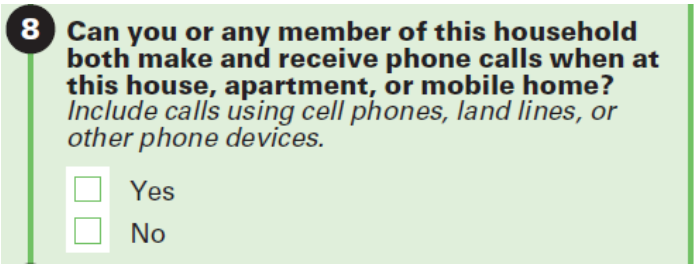


Figure 3. Test Version of the Telephone Service Question



The internet and Computer-Assisted Personal Interview (CAPI) question versions had the same content formatted accordingly for each mode.[4] Like current ACS production, the telephone service question was not asked in the Computer-Assisted Telephone Interview (CATI) operation given that interviewers were already on the phone with the respondent, thus the assumption was that the respondent had telephone service in the housing unit and the answer was automatically set to "yes."

1.4. Research Questions

The following research questions were formulated to guide the analyses of the Telephone Service question. The analyses assessed how the test version of the question performed compared to the control version in the following ways: how often the respondents answered the question, the consistency and accuracy of the responses, and how the responses affected the resulting estimates.

1. How does the proportion of housing units with no telephone service in each treatment compare with the proportion reported in the March 2016 ACS production panel and the most current NHIS Wireless Substitution Early Release Program?

2. Is there a difference in the item missing data rate between treatments? If so, which question version results in the lower rate?

3. Are there differences in the item missing data rates by mode between treatments?

4. Is the proportion of cases that report having telephone service greater in the test version than in the control version?

[4] Images of the control and test versions of the telephone service question in the internet mode can be found in Appendix A.

5. For each data collection mode, is the proportion of cases that report having telephone service greater in the test version than in the control version?

6. For each treatment, how often do cases report "No" to the telephone service question but report having a smartphone in the computer type question? Does either treatment have a higher rate of inconsistency in this sense, or are they about the same?

7. Of cases that marked "No" to telephone service on the original response, is the proportion of cases providing a valid phone number comparable between the two treatments? Of cases that marked "No" to telephone service but provided a valid phone number, is the proportion that could be contacted in the Content Follow-Up (CFU) telephone reinterview comparable between the two treatments?

2. METHODOLOGY

2.1. Sample Design

The 2016 ACS Content Test consisted of a nationally representative sample of 70,000 residential addresses in the United States, independent of the production ACS sample. The Content Test sample universe did not include GQs, nor did it include housing units in Alaska, Hawaii, or Puerto Rico.[5] The sample design for the Content Test was largely based on the ACS production sample design with some modifications to better meet the test objectives.[6] The modifications included adding an additional level of stratification by stratifying addresses into high and low self-response areas, oversampling addresses from low self-response areas to ensure equal response from both strata, and sampling units as pairs.[7] The high and low self-response strata were defined based on ACS self-response rates at the tract level. Sampled pairs were formed by first systematically sampling an address within the defined sampling stratum and then pairing that address with the address listed next in the geographically sorted list. Note that the pair was likely not neighboring addresses. One member of the pair was randomly assigned to receive the control version of the question and the other member was assigned to receive the test version of the question, thus resulting in a sample of 35,000 control cases and 35,000 test cases.

As in the production ACS, if efforts to obtain a response by mail or telephone were unsuccessful, attempts were made to interview in person a sample of the remaining nonresponding addresses (see Section 2.2 Data Collection for more details). Addresses were sampled at a rate of 1-in-3, with some exceptions that were sampled at a higher rate.[8] For the Content Test, the development of workload estimates for CATI and CAPI did not take into account the oversampling of low

[5] Alaska and Hawaii were excluded for cost reasons. GQs and Puerto Rico were excluded because the sample sizes required to produce reliable estimates would be overly large and burdensome, as well as costly.

[6] The ACS production sample design is described in Chapter 4 of the ACS Design and Methodology report (U.S. Census Bureau, 2014).

[7] Tracts with the highest response rate based on data from the 2013 and 2014 ACS were assigned to the high response stratum in such a way that 75 percent of the housing units in the population (based on 2010 Census estimates) were in the high response areas; all other tracts were designated in the low response strata. Self-response rates were used as a proxy for overall cooperation. Oversampling in low response areas helps to mitigate larger variances due to CAPI subsampling. This stratification at the tract level was successfully used in previous ACS Content Tests, as well as the ACS Voluntary Test in 2003.

[8] The ACS production sample design for CAPI follow-up is described in Chapter 4, Section 4.4 of the ACS Design and Methodology report (U.S. Census Bureau, 2014).

response areas. This oversampling resulted in a higher than expected workload for CATI and CAPI and therefore required more budget than was allocated. To address this issue, the CAPI sampling rate for the Content Test was adjusted to meet the budget constraint.

2.2. Data Collection

The field test occurred in parallel with the data collection activities for the March 2016 ACS production panel, using the same basic data collection protocol as production ACS with a few differences as noted below. The data collection protocol consisted of three main data collection operations: 1) a six-week mailout period, during which the majority of internet and mailback responses were received; 2) a one-month CATI period for nonresponse follow-up; and 3) a one-month CAPI period for a sample of the remaining nonresponse. Internet and mailback responses were accepted until three days after the end of the CAPI month.

As indicated earlier, housing units included in the Content Test sample were randomly assigned to a control or test version of the questions. CATI interviewers were not assigned specific cases; rather, they worked the next available case to be called and therefore conducted interviews for both control and test cases. CAPI interviewers were assigned Content Test cases based on their geographic proximity to the cases and therefore could also conduct both control and test cases.

The ACS Content Test's data collection protocol differed from the production ACS in a few significant ways. The Content Test analysis did not include data collected via the Telephone Questionnaire Assistance (TQA) program since those who responded via TQA used the ACS production TQA instrument. The Content Test excluded the telephone Failed Edit Follow-Up (FEFU) operation.[9] Furthermore, the Content Test had an additional telephone reinterview operation used to measure response reliability. We refer to this telephone reinterview component as the Content Follow-Up, or CFU. The CFU is described in more detail in Section 2.3.

ACS production provides Spanish-language versions of the internet, CATI, and CAPI instruments, and callers to the TQA number can request to respond in Spanish, Russian, Vietnamese, Korean, or Chinese. The Content Test had Spanish-language automated instruments; however, there were no paper versions of the Content Test questionnaires in Spanish.[10] Any case in the Content Test sample that completed a Spanish-language internet, CATI, or CAPI response was included in analysis. However, if a case sampled for the Content Test called TQA to complete an interview in Spanish or any other language, the production interview was conducted and the response was excluded from the Content Test analysis. This was due to the low volume of non-English language cases and the operational complexity of translating and implementing several language instruments for the Content Test. CFU interviews for the Content Test were conducted in either Spanish or English. The practical need to limit the

[9] In ACS production, paper questionnaires with an indication that there are more than five people in the household or questions about the number of people in the household, and self-response returns that are identified as being vacant or a business or lacking minimal data are included in FEFU. FEFU interviewers call these households to obtain any information the respondent did not provide.

[10] In the 2014 ACS, respondents requested 1,238 Spanish paper questionnaires, of which 769 were mailed back. From that information, we projected that fewer than 25 Spanish questionnaires would be requested in the Content Test.

language response options for Content Test respondents is a limitation to the research, as some respondents self-selected out of the test.

2.3. Content Follow-Up

For housing units that completed the original interview, a CFU telephone reinterview was also conducted to measure response error.[11] A comparison of the original interview responses and the CFU reinterview responses was used to answer research questions about response error and response reliability.[12]

A CFU reinterview was attempted with every household that completed an original interview for which there was a telephone number. A reinterview was conducted no sooner than two weeks (14 calendar days) after the original interview. Once the case was sent to CFU, it was to be completed within three weeks. This timing balanced two competing interests: (1) conducting the reinterview as soon as possible after the original interview to minimize changes in truth between the two interviews, and (2) not making the two interviews so close together that the respondents were simply recalling their previous answers. Interviewers made two call attempts to interview the household member who originally responded, but if that was not possible, the CFU reinterview was conducted with any other eligible household member (15 years or older).

The CFU asked basic demographic questions and a subset of housing and detailed person questions that included all of the topics being tested, with the exception of Telephone Service, and any questions necessary for context and interview flow to set up the questions being tested.[13] All CFU questions were asked in the reinterview, regardless of whether or not a particular question was answered in the original interview. Because the CFU interview was conducted via telephone, the wording of the questions in CFU followed the same format as the CATI nonresponse interviews. Housing units assigned to the control version of the questions in the original interview were asked the control version of the questions in CFU; housing units assigned to the test version of the questions in the original interview were asked the test version of the questions in CFU. The only exception was for retirement, survivor, and disability income, for which a different set of questions was asked in CFU.[14]

2.4. Analysis Metrics

This section describes the metrics used to assess the revised Telephone Service question. The metrics include the item missing data rate, response distributions, comparisons to benchmarks, response error, and other metrics. This section also describes the methodology used to calculate unit response rates and standard errors for the test.

---

[11] Throughout this report, the "original interview" refers to responses completed via paper questionnaire, internet, CATI, or CAPI.

[12] While Telephone Service was not a question asked in the CFU reinterview, the ability to contact a respondent for a CFU reinterview was used as part of the response reliability analysis specific to Telephone Service.

[13] Because the CFU interview was conducted via telephone the Telephone Service question was not asked. We assume that CFU respondents have telephone service.

[14] Refer to the 2016 ACS Content Test report on Retirement Income for a discussion on CFU questions for survivor, disability, and retirement income.

All Content Test data were analyzed without imputation due to our interest in how question changes or differences between versions of new questions affected "raw" responses, not the final edited variables. Some editing of responses was done for analysis purposes, such as collapsing response categories or modes together or calculating a person's age based on his or her date of birth.

All estimates from the ACS Content Test were weighted. Analysis involving data from the original interviews used the final weights that take into account the initial probability of selection (the base weight) and CAPI subsampling. For analysis involving data from the CFU interviews, the final weights were adjusted for CFU nonresponse to create CFU final weights.

The significance level for all hypothesis tests is $\alpha = 0.1$. Since we are conducting numerous comparisons between the control and test treatments, there is a concern about incorrectly rejecting a hypothesis that is actually true (a "false positive" or Type I error). 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. When adjusting for multiple comparisons, the Holm-Bonferroni method was used (Holm, 1979).

2.4.1. Unit Response Rates and Demographic Profile of Responding Households

The unit response rate is generally defined as the proportion of sample addresses eligible to respond that provided a complete or sufficient partial response.[15] Unit response rates from the original interview are an important measure to look at when considering the analyses in this report that compare responses between the control and test versions of the survey questionnaire. High unit response rates are important in mitigating potential nonresponse bias.

For both control and test treatments, we calculated the overall unit response rate (all modes of data collection combined) and unit response rates by mode: internet, mail, CATI, and CAPI. We also calculated the total self-response rate by combining internet and mail modes together. Some Content Test analyses focused on the different data collection modes for topic-specific evaluations, thus we felt it was important to include each mode in the response rates section. In addition to those rates, we calculated the response rates for high and low response areas because analysis for some Content Test topics was done by high and low response areas. Using the Census Bureau's Planning Database (U.S. Census Bureau, 2016), we defined these areas at the tract level based on the low response score.

The universe for the overall unit response rates consists of all addresses in the initial sample (70,000 addresses) that were eligible to respond to the survey. Some examples of addresses ineligible for the survey were a demolished home, a home under construction, a house or trailer that was relocated, or an address determined to be a permanent business or storage facility. The universe for self-response (internet and mail) rates consists of all mailable addresses that were eligible to respond to the survey. The universe for the CATI response rate consists of all nonrespondents at the end of the mailout month from the initial survey sample that were eligible to respond to the survey and for whom we possessed a telephone number. The universe for the

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

CAPI response rates consists of a subsample of all remaining nonrespondents (after CATI) from the initial sample that were eligible to respond to the survey. Any nonresponding addresses that were sampled out of CAPI were not included in any of the response rate calculations.

We also calculated the CFU interview unit response rate overall and by mode of data collection of the original interview and compared the control and test treatments, because response error analysis (discussed in Section 2.4.5) relies upon CFU interview data. Statistical differences between CFU response rates for control and test treatments will not be taken as evidence that one version is better than the other. For the CFU response rates, the universe for each mode consists of housing units that responded to the original questionnaire in the given mode (internet, mail, CATI, or CAPI) and were eligible for the CFU interview. We expected the response rates to be similar between treatments; however, we calculated the rates to verify that assumption.

Another important measure to look at in comparing experimental treatments is the demographic profile of the responding households in each treatment. The Content Test sample was designed with the intention of having respondents in both control and test treatments exhibit similar distributions of socioeconomic and demographic characteristics. Similar distributions allow us to compare the treatments and conclude that any differences are due to the experimental treatment instead of underlying demographic differences. Thus, we analyzed distributions for data from the following response categories: *age, sex, educational attainment,* and *tenure.* The topics of *race, Hispanic origin,* and *relationship* are also typically used for demographic analysis; however, those questions were modified as part of the Content Test, so we could not include them in the demographic profile. Additionally, we calculated *average household size* and the *language of response* for the original interview.[16]

For response distributions, we used chi-square tests of independence to determine statistical differences between control and test treatments. If the distributions were significantly different, we performed additional testing on the differences for each response category. To control for the overall Type I error rate for a set of hypotheses tested simultaneously, we performed multiple-comparison procedures with the Holm-Bonferroni method (Holm, 1979). A family for our response distribution analysis was the set of p-values for the overall characteristic categories (*age, sex, educational attainment,* and *tenure*) and the set of p-values for a characteristic's response categories if the response distributions were found to have statistically significant differences. To determine statistical differences for *average household size* and the *language of response* of the original interview we performed two-tailed hypothesis tests.

For all response-related calculations mentioned in this section, addresses that were either sampled out of the CAPI data collection operation or that were deemed ineligible for the survey were not included in any of the universes for calculations. Unmailable addresses were also excluded from the self-response universe. For all unit response rate estimates, differences, and demographic response analysis, we used replicate base weights adjusted for CAPI sampling (but not adjusted for CFU nonresponse).

[16] Language of response analysis excludes paper questionnaire returns because there was only an English questionnaire.

2.4.2. Item Missing Data Rates

Respondents leave items blank for a variety of reasons including not understanding the question (clarity), their unwillingness to answer a question as presented (sensitivity), and their lack of knowledge of the data needed to answer the question. The item missing data rate (for a given item) is the proportion of eligible units, housing units for household-level items or persons for person-level items, for which a required response (based on skip patterns) is missing.

For telephone service, the item missing data rates for the control and test treatments were calculated by dividing the number of invalid responses by all occupied housing units. Invalid responses included any household-level record that was a "don't know," "refuse," "blank," or a mail multiple. A mail multiple was any mail questionnaire in which both the "yes" and "no" responses were marked. Once the rates were calculated, a two-tailed t-test was conducted to test for significant differences between the two treatments.

2.4.3. Response Distributions

Comparing response distributions between the control version of a question and the test version of a question allows us to assess whether the question change affects the resulting estimates. Comparisons were made using Rao-Scott chi-squared tests (Rao & Scott, 1987) for distribution or t-tests for single categories when the corresponding distributions are found to be statistically different.

It was expected that the test treatment would result in a greater proportion of households having telephone service due to this version of the question including an additional instruction to clarify what types of telephone equipment should be included when a respondent answers the question. A t-test was conducted to compare both "yes" and "no" response categories between the control and test question versions. Proportion estimates were calculated as:

$$\text{Category proportion} = \frac{\text{weighted count of valid responses in category}}{\text{weighted count of all valid responses}}$$

2.4.4. Benchmarks

For the topic of Telephone Service, an estimate for the "percent of households without telephone service" is commonly used as a key measurement for understanding the lack of telephone service among households. Therefore, this benchmark estimate was used to informally evaluate data from both the control and test treatments to the March 2016 ACS production panel and the July-December 2015 NHIS Wireless Substitution Early Release Program. This informal evaluation examined whether results for the treatments grossly differed from other reliable resources, with the understanding that formal statistical comparisons could not be made due to such differences as methodology, question wording, and universes between data sources.

The March 2016 ACS production panel was used as a benchmark because both treatments in the Content Test had different question wording than the current telephone service question in ACS production. By using this source as a benchmark, the sampling and data collection processes

described in Sections 2.1 and 2.2 to collect telephone service data with the same universe and from the same month and survey for which the Content Test was being conducted could be taken advantage of. This production panel benchmark data were unedited to be consistent with the Content Test, with the exception of automatically editing the telephone service response to "yes" for CATI cases.[17] Any interviews completed through TQA were dropped from analysis.

Estimates from the July-December 2015 NHIS Wireless Substitution Early Release Program were used as a second benchmark. At the time of the analysis, the July-December 2015 NHIS Wireless Substitution Early Release Program was the most current release available. Though the NHIS primarily collects data focused on health care issues, it asks questions to determine whether a working phone is located inside a home that is not a cell phone and whether anyone in the household has a working cell phone.[18] These data are then tabulated to determine telephone status (e.g., landline with wireless, landline without wireless, wireless-only, etc.) for households, adults, and children.

The NHIS is a national, cross-sectional household interview survey conducted throughout the year. Informal, general comparisons using these benchmark data were made with caution due to several differences between the NHIS and the Content Test data. Most importantly, the NHIS uses a two-part question asking first about telephones inside the home and second about cell phones. Second, the universe for the NHIS is the civilian, noninstitutionalized population, which includes respondents from group quarters such as college dormitories. The universe in the ACS for the telephone service question does not include respondents living in noninstitutionalized group quarters, just those exclusively living in housing units. Lastly, the NHIS is a voluntary survey in which all data are collected through an in-person interview conducted by interviewers using a CAPI instrument. The ACS consists of four response modes, including self (mail and internet) as well as interviewer-administered response modes (CATI and CAPI).

2.4.5. Response Error

Response error occurs for a variety of reasons, such as flaws in the survey design, misunderstanding of the questions, misreporting by respondents, or interviewer effects. There are two components of response error: response bias and simple response variance. Response bias is the degree to which respondents consistently answer a question incorrectly. Simple response variance is the degree to which respondents answer a question inconsistently. A question has good response reliability if respondents tend to answer the question consistently. Re-asking the same question of the same respondent (or housing unit) allows us to measure response variance.

For most topics evaluated during the 2016 ACS Content test, simple response variance was measured by comparing the CFU reinterview with valid responses to the corresponding original interview. However, because CFU is a telephone interview and requires the respondent to have telephone service, the telephone service question was not included in the CFU.[19]

---

[17] Telephone service responses were automatically set to "yes" for CATI cases for the March 2016 ACS production panel and both treatments for the informal benchmark analysis. This editing process was done to align to ACS production procedures where the telephone service question is not asked for CATI cases since it is assumed the household has telephone service.

[18] Images of the NHIS telephone service questions can be found in Appendix B.

[19] When a respondent was reached for a telephone interview, it was assumed the household had telephone service. In these cases, the telephone service question was not asked and the response to the question was automatically set to "yes" for the household.

For telephone service, the consistency of responses to the question was assessed using other information provided during the Content Test. Analysis was conducted to determine how often households reported having no telephone service but did report having a smartphone in the computer type question. The computer type question immediately follows the telephone service question in both the control and test treatments. However, the computer type question was also tested as part of the Content Test and because the categorical responses for the question in the control treatment did not specify a singular "smartphone" category like the test treatment, comparisons between the two treatments were made with caution.[20] This was necessary because "handheld computers" and "other handheld wireless computers" were grouped into the same response category as "smart mobile phones" in the control treatment. As a result, respondents that may not have a smartphone but did have one of these other devices with no telephone service could be correctly answering "yes" to this question.

Further analysis was also conducted to assess the consistency between the telephone service question response, the presence of a valid telephone number, and the ability to use the telephone number to make contact with a respondent for a CFU reinterview. First, the proportion of households that reported having no telephone service but provided a valid telephone number in the original interview was calculated. The ACS FEFU rules were followed to determine if a telephone number was valid or not.[21] Second, of these households that reported having no telephone service but had a valid telephone number, a calculation was made to determine the proportion of households that the Census Bureau could successfully contact for a CFU reinterview, which would indicate that the household possibly does have telephone service. A contact was considered successful if the case had one of the following outcome codes:

- Fully complete
- Sufficient partial[22]
- Sufficient partial set at closeout
- Sample unit eligible but unavailable through closeout
- Unconverted language problem[23]
- Hostile breakoff
- Refusal
- Insufficient partial[24]

[20] Images of the control and test versions of the computer type question in the mail mode can be found in Appendix C.
[21] The ACS FEFU rules for determining a valid telephone number can be found in Appendix D.
[22] A case is deemed a "sufficient partial" when the respondent gets to the first question in the detailed person questions section for the first person in the household.
[23] A case is deemed an "unconverted language problem" when a respondent is contacted but the Census Bureau could not get an interviewer that spoke the same language as the respondent to conduct the interview.
[24] A case is deemed an "insufficient partial" when a respondent is contacted but the interview does not proceed to the first question in the detailed person questions section for the first person in the household.

2.4.6. Standard Error Calculations

We estimated the variances of the estimates using the Successive Differences Replication (SDR) method with replicate weights, the standard method used in the ACS (see U.S. Census Bureau, 2014, Chapter 12). We calculated the variance for each rate and difference using the formula below. The standard error of the estimate (X_0) is the square root of the variance:

$$\mathrm{Var}(X_0) = \frac{4}{80}\sum_{r=1}^{80}(X_r - X_0)^2$$

where:

X_0 = the estimate calculated using the full sample,
X_r = the estimate calculated for replicate r.

3. DECISION CRITERIA

Before fielding the 2016 ACS Content Test, we identified which of the metrics would be given higher importance in determining which version of the question would be recommended for inclusion in the ACS moving forward. The following table identifies the research questions and associated metrics in priority order.

Table 1. Decision Criteria for Telephone Service

Research Question	Decision Criteria, in order of priority
4	Overall, the proportion of cases that report having telephone service should be higher in the test treatment than in the control treatment.
2	Overall, the item missing data rate for the test treatment should be the same or lower than the control treatment.
5	For each mode, the proportion of cases that report having telephone service should be higher in the test treatment than in the control treatment.
3	For each mode, the item missing data rate for the test treatment should be the same or lower than the control treatment.
1	The proportion of housing units with no telephone service in the test treatment should be lower than the March 2016 ACS production panel and the NHIS Wireless Substitution Early Release Program benchmark estimates.
7	The proportion of cases that originally reported having no telephone service but were reached for CFU should be the same or lower in the test treatment than in the control treatment.
6	The proportion of cases that originally reported having no telephone service but reported having a smartphone in the computer type question should be the same or lower in the test treatment than in the control treatment.

4. LIMITATIONS

CATI and CAPI interviewers were assigned control and test treatment cases, as well as production cases. The potential risk of this approach is the introduction of a cross-contamination

or carry-over effect due to the same interviewer administering multiple versions of the same question item. Interviewers are trained to read the questions verbatim to minimize this risk, but there still exists the possibility that an interviewer may deviate from the scripted wording of one question version to another. This could potentially mask a treatment effect from the data collected.

Interviews were only conducted in English and Spanish. Respondents who needed language assistance in another language were not able to participate in the test. Additionally, the 2016 ACS Content Test was not conducted in Alaska, Hawaii, or Puerto Rico. Any conclusions drawn from this test may not apply to these areas or populations.

For statistical analysis specific to the mail mode, there may be bias in the results because of unexplained unit response rate differences between the control and test treatments.

We were not able to conduct demographic analysis by relationship status, race, or ethnicity because these topics were tested as part of the Content Test.

The Content Test does not include the production weighting adjustments for seasonal variations in ACS response patterns, nonresponse bias, and under-coverage bias. As a result, any estimates derived from the Content Test data do not provide the same level of inference as the production ACS and cannot be compared to production estimates.

In developing initial workload estimates for CATI and CAPI, we did not take into account the fact that we oversampled low response areas as part of the Content Test sample design. Therefore, workload and budget estimates were too low. In order to stay within budget, the CAPI workload was subsampled more than originally planned. This caused an increase in the variances for the analysis metrics used.

An error in addressing and assembling the materials for the 2016 ACS Content Test caused some Content Test cases to be mailed production ACS questionnaires instead of Content Test questionnaires. There were 49 of these cases that returned completed questionnaires, and they were all from the test treatment. These cases were excluded from the analysis. Given the small number of cases affected by this error, there is very little effect on the results.

As mentioned in Section 2.4.5, response error could not be measured using CFU metrics for telephone service because CFU was a telephone interview and required the respondent to have telephone service. Instead, the quality of responses to the telephone service question was assessed by analyzing the consistency between question responses from the original interview, the presence of a valid telephone number from the original interview, and the ability of the Census Bureau to have a successful contact using the provided telephone number during CFU. However, the assumption was made that a valid telephone number had service within the respondent’s household. While this could be the case, there was the possibility a valid telephone number provided by the respondent was a phone number either external to the household, such as a work or friend’s, or even a fake phone number.

5. RESEARCH QUESTIONS AND RESULTS

This section presents the results from the analyses of the 2016 ACS Content Test data for the Telephone Service question. An analysis of unit response rates is presented first followed by topic-specific analyses. For the topic-specific analyses, each research question is restated, followed by corresponding data and a brief summary of the results.

5.1. Unit Response Rates and Demographic Profile of Responding Households

This section provides results for unit response rates for both control and test treatments for the original Content Test interview and for the CFU interview. It also provides results of a comparison of socioeconomic and demographic characteristics of respondents in both control and test treatments.

5.1.1. Unit Response Rates for the Original Content Test Interview

The unit response rate is generally defined as the proportion of sample addresses eligible to respond that provided a complete or sufficient partial response. We did not expect the unit response rates to differ between treatments. This is important because the number of unit responses should also affect the number of item responses we receive for analyses done on specific questions on the survey. Similar item response universe sizes allow us to compare the treatments and conclude that any differences are due to the experimental treatment instead of differences in the populations sampled for each treatment.

Table 2 shows the unit response rates for the original interview for each mode of data collection (internet, mail, CATI, and CAPI), all modes combined, and both self-response modes (internet and mail combined) for the control and test treatments. When looking at the overall unit response rate (all modes combined) the difference between control (93.5 percent) and test (93.5 percent) is less than 0.1 percentage points and is not statistically significant.

Table 2. Original Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode

Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
All Modes	19,400	93.5 (0.3)	19,455	93.5 (0.3)	<0.1 (0.4)	0.98
Self-Response	13,131	52.9 (0.5)	13,284	53.7 (0.5)	-0.8 (0.6)	0.23
Internet	8,168	34.4 (0.4)	8,112	34.1 (0.4)	0.4 (0.6)	0.49
Mail	4,963	18.4 (0.3)	5,172	19.6 (0.3)	-1.2 (0.5)	0.01*
CATI	872	8.7 (0.4)	880	9.2 (0.4)	-0.4 (0.6)	0.44
CAPI	5,397	83.5 (0.7)	5,291	83.6 (0.6)	<0.1 (0.9)	0.96

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

When analyzing the unit response rates by mode of data collection, the only modal comparison that shows a statistically significant difference is the mail response rate. The control treatment

had a higher mail response (19.6 percent) than the test treatment (18.4 percent) by 1.2 percentage points. As a result of this difference, we looked at how mail responses differed in the high and low response areas. Table 3 shows the mail response rates for both treatments in high and low response areas.[25] The difference in mail response rates appears to be driven by the difference of rates in the high response areas.

It is possible that the difference in the mail response rates between control and test is related to the content changes made to the test questions. There are some test questions that could be perceived as being too sensitive by some respondents (such as the test question relating to same-sex relationships) and some test questions that could be perceived to be too burdensome by some respondents (such as the new race questions with added race categories). In the automated modes (internet, CATI, and CAPI) there is a higher likelihood of obtaining a sufficient partial response (obtaining enough information to be deemed a response for calculations before the respondent stops answering questions) than in the mail mode. If a respondent is offended by the questionnaire or feels that the questions are too burdensome they may just throw the questionnaire away, and not respond by mail. This could be a possible explanation for the unit response rate being lower for test than control in the mail mode.

We note that differences between overall and total self-response response rates were not statistically significant. As most analysis was conducted at this level, we are confident the response rates were sufficient to conduct topic-specific comparisons between the control and test treatments and that there are no underlying response rate concerns that would impact those findings.

Table 3. Mail Response Rates by Designated High (HRA) and Low (LRA) Response Areas

	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
HRA	2,082	20.0 (0.4)	2,224	21.5 (0.4)	-1.5 (0.6)	0.02*
LRA	2,881	13.8 (0.3)	2,948	14.1 (0.3)	-0.3 (0.4)	0.43
Difference		6.2 (0.5)		7.4 (0.4)	-1.1 (0.7)	0.11

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

5.1.2. Unit Response Rates for the Content Follow-Up Interview

Table 4 shows the unit response rates for the CFU interview by mode of data collection of the original interview and for all modes combined, for control and test treatments. Overall, the differences in CFU response rates between the treatments are not statistically significant. The rate at which CAPI respondents from the original interview responded to the CFU interview is lower for test (34.8 percent) than for control (37.7 percent) by 2.9 percentage points. While the protocols for conducting CAPI and CFU were the same between the test and control treatments, we could not account for personal interactions that occur in these modes between the respondent and interviewer. This can influence response rates. We do not believe that the difference suggests

[25] Table E-1 (including all modes) can be found in Appendix E.

any underlying CFU response issues that would negatively affect topic- specific response reliability analysis for comparing the two treatments.

Table 4. Content Follow-Up Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode of Original Interview

Original Interview Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
All Modes	7,867	44.8 (0.5)	7,903	45.7 (0.6)	-0.8 (0.8)	0.30
Internet	4,078	51.9 (0.6)	4,045	52.5 (0.7)	-0.6 (0.8)	0.49
Mail	2,202	46.4 (0.9)	2,197	44.2 (0.9)	2.1 (1.3)	0.11
CATI	369	48.9 (1.9)	399	51.5 (2.5)	-2.5 (2.9)	0.39
CAPI	1,218	34.8 (1.2)	1,262	37.7 (1.1)	-2.9 (1.6)	0.07*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

5.1.3. Demographic and Socioeconomic Profile of Responding Households

One of the underlying assumptions of our analyses in this report is that the sample for the Content Test was selected in such a way that responses from both treatments would be comparable. We did not expect the demographics of the responding households for control and test treatments to differ. To test this assumption, we calculated distributions for respondent data for the following response categories: *age, sex, educational attainment,* and *tenure.*[26] The response distribution calculations can be found in Table 5. Items with missing data were not included in the calculations. After adjusting for multiple comparisons, none of the differences in the categorical response distributions shown below is statistically significant.

[26] We were not able to conduct demographic analysis by relationship status, race, or ethnicity because these topics were tested as part of the Content Test.

Table 5. Response Distributions: Test versus Control Treatment

Item	Test Percent	Control Percent	Adjusted P-Value
AGE	(n=43,236)	(n=43,325)	0.34
Under 5 years old	5.7 (0.2)	6.1 (0.2)	
5 to 17 years old	17.8 (0.3)	17.6 (0.3)	
18 to 24 years old	8.6 (0.3)	8.1 (0.3)	
25 to 44 years old	25.1 (0.3)	26.2 (0.3)	
45 to 64 years old	26.8 (0.4)	26.6 (0.4)	
65 years old or older	16.0 (0.3)	15.4 (0.3)	
SEX	(n=43,374)	(n=43,456)	1.00
Male	48.8 (0.3)	49.1 (0.3)	
Female	51.2 (0.3)	50.9 (0.3)	
EDUCATIONAL ATTAINMENT#	(n=27,482)	(n=27,801)	1.00
No schooling completed	1.3 (0.1)	1.2 (0.1)	
Nursery to 11th grade	8.1 (0.3)	8.0 (0.3)	
12th grade (no diploma)	1.7 (0.1)	1.6 (0.1)	
High school diploma	21.7 (0.4)	22.3 (0.4)	
GED† or alternative credential	3.5 (0.2)	3.6 (0.2)	
Some college	21.0 (0.4)	20.2 (0.4)	
Associate's degree	8.8 (0.3)	9.1 (0.3)	
Bachelor's degree	20.9 (0.4)	20.3 (0.4)	
Advanced degree	13.1 (0.3)	13.7 (0.3)	
TENURE	(n=17,190)	(n=17,236)	1.00
Owned with a mortgage	43.1 (0.6)	43.2 (0.5)	
Owned free and clear	21.1 (0.4)	21.2 (0.4)	
Rented	33.8 (0.6)	34.0 (0.5)	
Occupied without payment of rent	1.9 (0.2)	1.7 (0.1)	

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
#For ages 25 and older
†General Educational Development
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing done at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

We also analyzed two other demographic characteristics shown by the responses from the survey: *average household size* and *language of response.* The results for the remaining demographic analyses can be found in Table 6 and Table 7.

Table 6. Comparison of Average Household Size

Topic	Test (n=17,608)	Control (n=17,694)	Test minus Control	P-value
Average Household Size (Number of People)	2.51 (<0.1)	2.52 (<0.1)	>-0.01 (<0.1)	0.76

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a two-tailed t-test at the α=0.1 level.

Table 7. Comparison of Language of Response

Language of Response	Test Percent (n=17,608)	Control Percent (n=17,694)	Test minus Control	P-value
English	96.1 (0.2)	96.2 (0.2)	<0.1 (0.3)	0.52
Spanish	2.7 (0.2)	2.6 (0.2)	<0.1 (0.2)	0.39
Undetermined	1.2 (0.1)	1.2 (0.1)	<0.1 (0.2)	0.62

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a two-tailed t-test at the α=0.1 level.

The Content Test was available in two languages, English and Spanish, for all modes except the mail mode. However, the language of response variable was missing for some responses, so we created a category called *undetermined* to account for those cases.

There are no detectable differences between control and test for *average household size* or *language of response*. There are also no detectable differences for any of the response distributions that we calculated. As a result of these analyses, it appears that respondents in both treatments do exhibit comparable demographic characteristics since none of the resulting findings is significant, which verifies our assumption of demographic similarity between treatments.

5.2. Item Missing Data Rates

Is there a difference in the item missing data rate between treatments? If so, which question version results in the lower rate?

Are there differences between the item missing data rates by mode between treatments?

Table 8 shows the item missing data rates for the control and test treatments overall and by each data collection mode. The item missing data rates were calculated as defined in Section 2.4.2. A two-tailed t-test was used to determine significant differences between treatments.

For both research questions, the universe was all occupied housing units.

Table 8. Item Missing Data Rate for Control and Test Treatment, by Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	16,754	1.2 (<0.1)	16,843	1.3 (<0.1)	-0.1 (0.1)	0.32
Internet	8,105	1.0 (0.1)	8,073	1.4 (0.2)	-0.4 (0.2)	0.06*
Mail	4,859	2.4 (0.3)	5,062	2.7 (0.3)	-0.2 (0.4)	0.54
CAPI	3,790	0.6 (0.2)	3,708	0.2 (0.1)	0.4 (0.2)	0.09*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

Overall, item missing data rates were not statistically different between the test and control treatments. Item missing data rates for the mail mode of data collection were also not statistically different between treatments. For the internet mode of data collection, the test treatment had an

item missing data rate of 1.0 percent, which was statistically lower than the control treatment's item missing data rate of 1.4 percent. For the CAPI mode of data collection, the control treatment's item missing data rate of 0.2 percent was statistically lower than the 0.6 percent item missing data rate for the test treatment.

5.3. Response Distributions

Is the proportion of cases that report having telephone service greater in the test version than in the control version?

For each data collection mode, is the proportion of cases that report having telephone service greater in the test version than in the control version?

Table 9 shows the households that reported having telephone service for the control and test treatments overall and by data collection mode in which the telephone service question was asked. The response distributions were calculated as defined in Section 2.4.3. A one-tailed t-test was used to determine if the proportion of cases that reported having telephone service was higher in the test treatment as compared to the control treatment.

For both research questions, the universe was all occupied housing units with a valid response. "Don't know," "refuse," "blank," and mail returns with both "Yes' and "No" answers selected were considered invalid responses and not part of the universe to answer these research questions.

Table 9. Households Reporting Having Telephone Service, by Treatment and Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	16,492	98.3 (0.2)	16,551	96.5 (0.2)	1.9 (0.2)	<0.01*
Internet	8,019	98.9 (0.1)	7,959	96.5 (0.2)	2.5 (0.3)	<0.01*
Mail	4,701	98.3 (0.2)	4,892	95.3 (0.4)	3.0 (0.5)	<0.01*
CAPI	3,772	97.7 (0.4)	3,700	97.2 (0.4)	0.5 (0.5)	0.15

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test > control) at the α=0.1 level.

Overall, the proportion of cases that reported having telephone service in the test treatment was 98.3 percent, a statistically higher proportion than the 96.5 percent of cases that reported having telephone service in the control treatment. For the two self-response modes of data collection, internet and mail, the test treatment also had a statistically higher proportion of cases that reported having telephone service than the control treatment. Within the CAPI mode of data collection, the proportion of cases that reported having telephone service was 97.7 percent in the test treatment, however this estimate was not statistically greater than the 97.2 percent of cases in the control treatment that reported having telephone service.

5.4. Benchmarks

How does the proportion of housing units with no telephone service in each treatment compare with the proportion reported in the March 2016 ACS production panel and the most current NHIS Wireless Substitution Early Release Program?

Since the wording in both treatments in the Content Test were different from the current ACS telephone service question wording, the March 2016 ACS production panel was used as a benchmark (see Section 2.4.4).

At the time of the analysis, the July-December 2015 NHIS Wireless Substitution Early Release Program was the most current release available for the second benchmark (see Section 2.4.4).

Table 10 shows the percent of households without telephone service from each treatment and benchmark. Formal statistical comparisons of differences were not made since the Content Test data did not include ACS production weighting and were not imputed nor adjusted for nonresponse or under-coverage bias.

However, the data presented in Table 10 for both treatments included CATI cases for which the telephone service response was automatically set to "yes." Inclusion of CATI cases was exclusive to the estimates produced for Table 10, and was done to align to ACS production procedures where the telephone service question is not asked for CATI cases since it is assumed the household has telephone service.

Table 10. Percent of Households without Telephone Service by Treatment and Benchmark

	Test	Control	ACS March 2016 Production Panel	NHIS July-December 2015
Percent of households without telephone service	1.6 (0.2)	3.4 (0.2)	2.9 (<0.1)	3.1 (0.2)

Sources: U.S. Census Bureau, 2016 American Community Survey (ACS) Content Test, March 2016 ACS production panel, National Health Interview Survey, July-December 2015 Wireless Substitution Early Release Program
Note: Standard errors are shown in parentheses.

While no formal statistical comparisons between these estimates were made, it is helpful to put the estimates into perspective with ACS production estimates of the percent of households without telephone service. Between 2011 and 2015, the percent of households without telephone service in the ACS 1-year estimates ranged from 2.3 percent to 2.6 percent, with all years' estimates having a margin of error of 0.1 percent.[27]

As mentioned in Section 2.4.4., differences between the benchmark data and ACS production data are important to understand. A key difference is that all NHIS data are collected by in-person surveys conducted by interviewers using a CAPI instrument, whereas the data collected in both the ACS Content Test and ACS production utilizes the mail, internet, and CAPI modes of data collection for the telephone service question.

[27] The 2011, 2012, 2013, 2014, and 2015 estimates and margins of error were derived by accessing ACS 1-year table ID number "DP04" entitled "Selected Housing Characteristics" from the U.S. Census Bureau's American FactFinder data dissemination platform located at https://factfinder.census.gov.

5.5. Results for Analysis Specific to Telephone Service

As mentioned in Section 2.4.5., response error could not be measured for telephone service using the reinterview due to CFU being a telephone interview that required the respondent to have telephone service. Instead, other analysis was conducted to assess the quality and reliability of responses to the telephone service question. This analysis provides information about the consistency between question responses from the original interview, the presence of a valid telephone number from the original interview, and the ability of the Census Bureau to make a successful contact with the respondent using the provided telephone number during CFU.

5.5.1. Consistency between Telephone Service and Computer Type Questions

For each treatment, how often do cases answer "No" to the telephone service question but report having a smartphone in the computer type question? Does either treatment have a higher rate of inconsistency in this sense, or are they about the same?

Table 11 shows the percent of households in each treatment that reported having no telephone service, but did report having a smartphone in the computer type question. Comparisons between the two treatments should be made with caution, as the categorical responses for the computer type question in the control treatment did not specify a singular "smartphone" category like the test treatment. Since the control treatment grouped "smart mobile phones" with "handheld computers" and "other handheld wireless computers" into the same response category, respondents that may not have a smartphone could still answer "yes" to this question if they have computer devices that would be included in this response category but do not accommodate phone calls.

Table 11. Households Reporting No Telephone Service but Reported Having a Smartphone, by Treatment

	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Households reporting no telephone service but reported having a smartphone	311	54.9 (4.3)	715	65.0 (2.4)	-10.1 (4.8)	0.03*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

Overall, the proportion of households that reported having no telephone service but reported having a smartphone in the computer type question was 54.9 percent in the test treatment, statistically lower than the 65.0 percent of households in the control treatment.

5.5.2. Consistency between Telephone Service Response with Valid Phone Number and Successful CFU Contact

Of cases that marked "No" telephone service on the original response, is the proportion of cases providing a valid phone number comparable between the two treatments? Of cases that marked "No" telephone service but provided a valid phone number, is the proportion that could be contacted in CFU comparable between the two treatments?

Table 12 shows the percent of households that reported having no telephone service but provided a valid phone number during the original Content Test interview. Valid phone numbers were those phone numbers that met the ACS FEFU rules (Appendix D) in determining the validity of a phone number.

Table 12. Households Reporting No Telephone Service that Provided a Valid Phone Number, by Treatment

	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Households reporting no telephone service that provided a valid phone number	311	82.4 (3.0)	715	84.8 (2.1)	-2.4 (3.6)	0.51

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance was tested based on a two-tailed t-test at the α=0.1 level.

There was no statistical difference between treatments of the proportion of households that reported no telephone service that provided a valid phone number.

In order to gain more understanding into the consistencies between the households that reported having no telephone service but had a valid phone number, these households were analyzed to determine whether they were contacted successfully during the CFU operation (see Section 2.4.5 for a definition of what constituted a successful contact). To ensure the universe included only those households for which a CFU contact was attempted at least once, any household that was part of the CFU operation but for which contact was never attempted was removed from the universe. Table 13 shows the percent of households with a successful CFU contact and for which the household provided a valid phone number during the original interview but the household originally reported it did not have telephone service.

Table 13. Households Reporting No Telephone Service that Provided a Valid Phone Number and Had a Successful CFU Contact, by Treatment

	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Households reporting no telephone service that provided a valid phone number and had a successful CFU contact	241	67.3 (4.4)	607	62.2 (2.7)	5.1 (5.4)	0.34

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance was tested based on a two-tailed t-test at the α=0.1 level.

There was no statistical difference between treatments in the proportion of households successfully contacted during CFU and had a valid phone number but reported having no telephone service in the original interview.

Additional analysis was conducted to assess the respondents that did report having telephone service and their consistency of providing a valid phone number and the Census Bureau's ability to successfully contact them for a CFU interview. See Appendix F and G for these supplemental tables.

6. CONCLUSIONS AND RECOMMENDATIONS

This report discusses findings from the ACS Content Test for two different versions of the telephone service question. Changes to the telephone service question were tested to clarify the intent of the question in order to make it easier for respondents to answer correctly.

The results were evaluated using a series of decision criteria established prior to testing. The metrics with the highest priority involved the response distribution and item missing data rates for each treatment. The test version had a higher proportion of households reporting having telephone service. The test treatment proportion was also found to be higher for the internet and mail modes of data collection. The test treatment did not have a statistically higher proportion of cases with telephone service in the CAPI mode. It is presumed that the additional clarification, particularly in the self-response modes of data collection, respondents in the test treatment were able to develop an understanding of the acceptable equipment that could make or receive phone calls is not limited to cell phones or landlines, but other phone devices as well. The results suggest that the additional instruction provided in the test version helped respondents to answer the question.

Overall, item missing data rates between the two treatments were not statistically different. When analyzed by mode of data collection, this also held true for the mail mode. Item missing data rates for the test treatment were statistically lower in the internet mode. As more respondents continue to use the internet mode to complete the ACS, this result is encouraging as it demonstrated that the test version might have made respondents more willing and able to answer the question by clarifying the question's intent.

In the CAPI mode, the control treatment had a lower item missing data rate. However, of the three data collection modes used to ask this particular question, the test version having the same or lower item missing data rates in two of the three modes was sufficient to recommend the test treatment. Furthermore, the CAPI mode is a field-representative mode of data collection and additional training on administering this question may help improve item missing data rates in the future.

Additional research was conducted to analyze the consistency on how respondents answered the telephone service question relative to other information from the Content Test that could possibly indicate a household had telephone service. Neither treatment had statistical differences in the proportion of households that originally reported having no telephone service but provided a valid phone number. Of the households for which a CFU attempt was made, there was no difference between treatments of the proportion of households that had a successful CFU contact but originally reported not having telephone service.

There was a difference between treatments in terms of the inconsistency of households reporting no telephone service but reporting having a smartphone in the computer type question. The control treatment was found to have a higher proportion of households with this inconsistency. However, the broader definitions for the response categories for the computer type question in the control treatment lead us to believe that some of the difference between treatments can be explained by these definitional differences.

Based upon the prioritized decision criteria for this analysis, results of the evaluation indicate that the test version of the telephone service question performed at an acceptable or better level than the control version to measure whether a household can make and receive phone calls. The additional instruction being inclusive and signifying “other phone devices” could help alleviate issues of confusion as telephone equipment and services continue to evolve in the future, even beyond the changes made for this testing. Therefore, the Census Bureau recommends the implementation of the test version on the ACS moving forward.

7. ACKNOWLEDGEMENTS

The 2016 ACS Content Test would not have been possible without the participation and assistance of many individuals from the Census Bureau and other agencies. Their contributions are sincerely appreciated and gratefully acknowledged.

- Census Bureau staff in the American Community Survey Office, Application Development and Services Division, Decennial Information Technology Division, Decennial Statistical Studies Division, Field Division, National Processing Center, Population Division, and Social, Economic, and Housing Statistics Division.
- Representatives from other agencies in the Federal statistical system serving on the Office of Management and Budget’s Interagency Working Group for the ACS and the Topical Subcommittees formed by the Working Group for each topic tested on the 2016 ACS Content Test.
- Staff in the Office of Management and Budget’s Statistical and Science Policy Office.

The authors would like to thank the following individuals for their contributions to the analysis and review of the Telephone Service report: *Elizabeth Poehler and Ellen Wilson*.

8. REFERENCES

Blumberg, S. J., & Luke, J. V. (2016). *Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2015*. National Center for Health Statistics. Retrieved November 18, 2016 from http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf

Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure," Scandinavian Journal of Statistics, Vol. 6, No. 2: 65-70. Retrieved on January 31, 2017 from https://www.jstor.org/stable/4615733?seq=1#page_scan_tab_contents

Rao, J. N. K., & Scott, A. J. (1987). "On Simple Adjustments to Chi-Square Tests with Sample Survey Data," The Annal of Statistics, Vol. 15, No. 1, 385-397. Retrieved on January 31, 2017 from http://projecteuclid.org/euclid.aos/1176350273

Stapleton, M., & Steiger, D. (2015). *Cognitive Testing of the 2016 American Community Survey Content Test Items: Summary Report for Round 1 and Round 2 Interviews.* Westat, Rockville, Maryland, January 2015.

Steiger, D., Anderson, J., Folz, J., Leonard, M., & Stapleton, M. (2015). *Cognitive Testing of the 2016 American Community Survey Content Test Items: Briefing Report for Round 3 Interviews.* Westat, Rockville, Maryland, June, 2015.

U.S. Census Bureau. (2014). *American Community Survey Design and Methodology (January 2014).* Retrieved February 1, 2017 from http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

U.S. Census Bureau. (2016). *2015 Planning Database Tract Data* [Data file]. Retrieved on January 31, 2017 from http://www.census.gov/research/data/planning_database/2015/

Appendix A. Control and Test Versions of the Telephone Service Question in Internet Mode

Figure A-1. Control Version of the Telephone Service Question in the Internet Mode

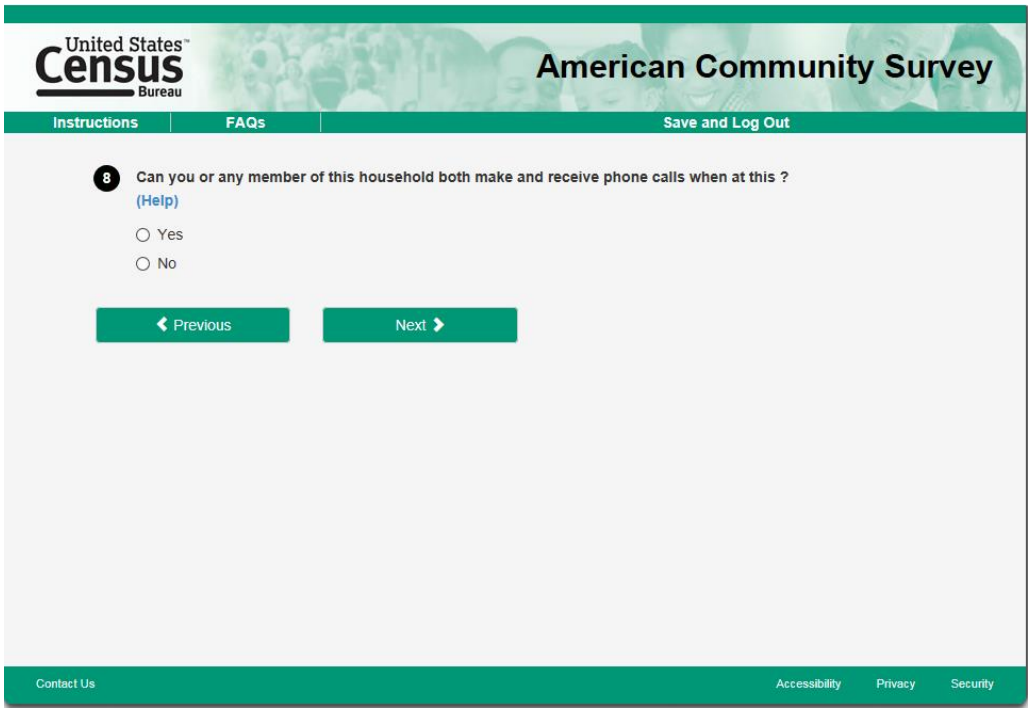
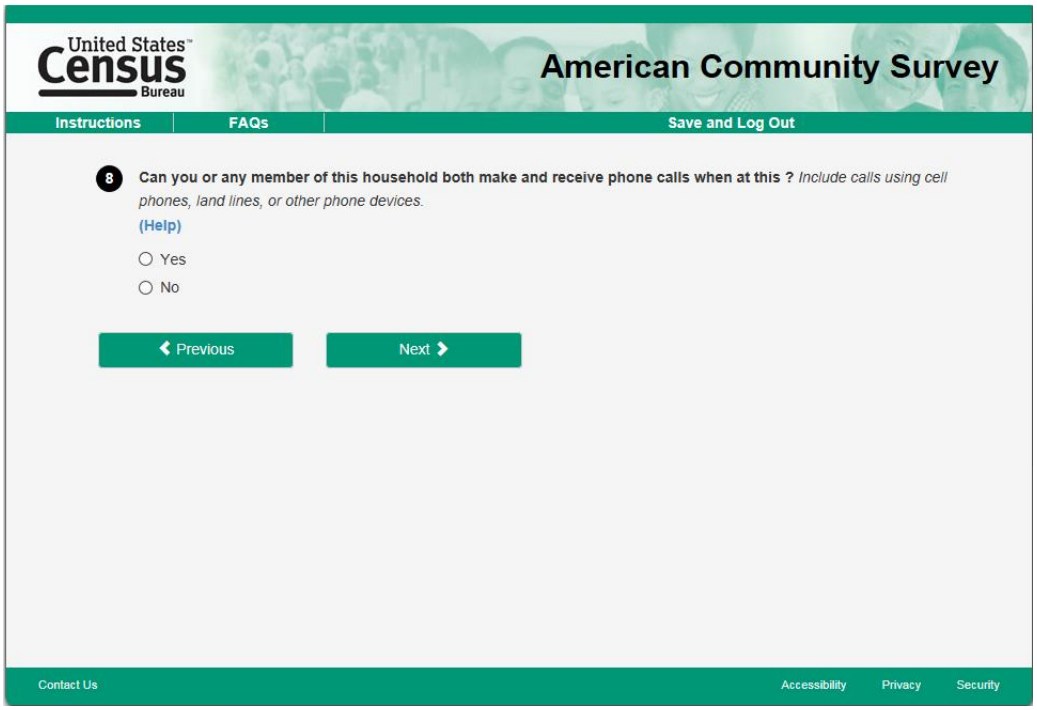


Figure A-2. Test Version of the Telephone Service Question in the Internet Mode



Appendix B. NHIS Telephone Service Questions

Figure B-1. NHIS Question on Working Telephone Inside Home

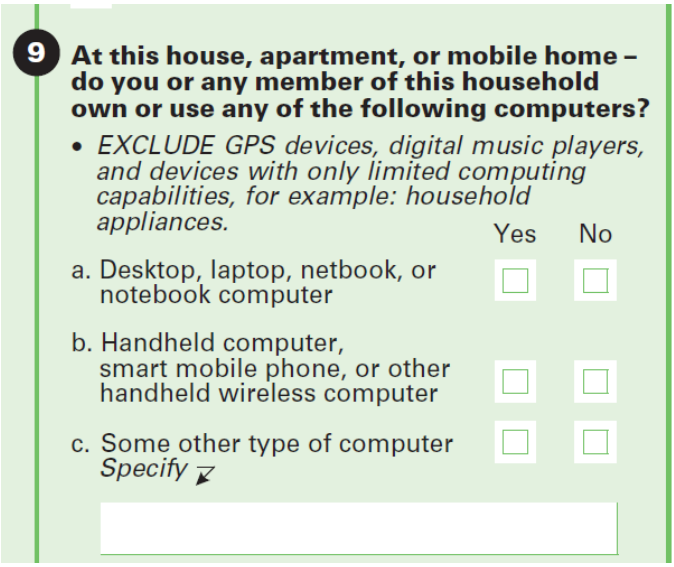
QuestionText:	?[F1] Is there at least one telephone INSIDE your home that is currently working and is not a cell phone?
1	Yes
2	No
7	Refused
9	Don't know

Figure B-2. NHIS Question on Working Cell Phone

QuestionText:	Do you or anyone in your family have a working cell phone?
1	Yes
2	No
7	Refused
9	Don't know

Appendix C. Control and Test Versions of the Computer Type Question in Mail Mode

Figure C-1. Control Version of the Computer Type Question

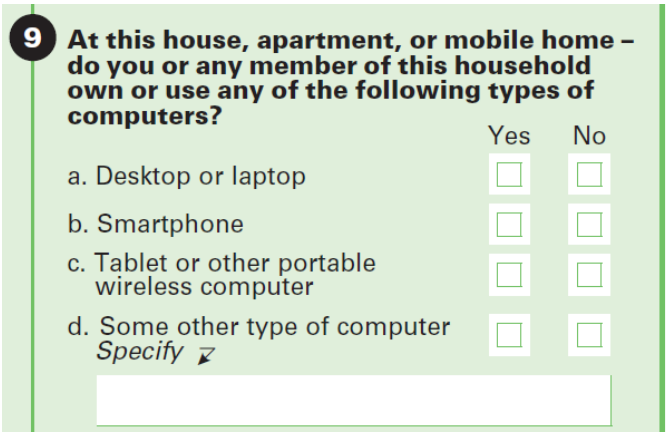


9 At this house, apartment, or mobile home – do you or any member of this household own or use any of the following computers?

- *EXCLUDE GPS devices, digital music players, and devices with only limited computing capabilities, for example: household appliances.*

	Yes	No
a. Desktop, laptop, netbook, or notebook computer	☐	☐
b. Handheld computer, smart mobile phone, or other handheld wireless computer	☐	☐
c. Some other type of computer *Specify*	☐	☐

Figure C-2. Test Version of the Computer Type Question



9 At this house, apartment, or mobile home – do you or any member of this household own or use any of the following types of computers?

	Yes	No
a. Desktop or laptop	☐	☐
b. Smartphone	☐	☐
c. Tablet or other portable wireless computer	☐	☐
d. Some other type of computer *Specify*	☐	☐

Appendix D. ACS FEFU Rules to Determine a Valid Phone Number

1. Create the variable ‘Phone’ by concatenating ‘Area Code’ and ‘Phone Number’
2. Set a valid respondent telephone number if the following conditions are true:
   a. Area Code does not equal ‘000’, ‘111’, ‘222’, ‘333’, ‘444’, ‘555’, ‘666’, ‘777’, ‘888,’ or ‘999’
   b. Area Code is greater than or equal to ‘200’
   c. Phone Number does not equal ‘1111111’, ‘2222222’, ‘3333333’, ‘4444444’, ‘5555555’, ‘6666666’, ‘7777777’, ‘8888888’, ‘9999999’, or ‘0000000’
   d. Phone Number is greater than or equal to ‘2000000’
   e. The length of the Phone Number is 10 digits
   f. Phone consists only of digits '0' through '9'

Appendix E. Unit Response Rates Supplemental Table

Table E-1. Unit Response Rates by Designated High (HRA) and Low (LRA) Response Areas

Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test Minus Control	P-Value
Total Response	19,400		19,455			
HRA	7,556	94.3 (0.4)	7,608	94.5 (0.3)	-0.2 (0.6)	0.72
LRA	11,844	91.5 (0.3)	11,847	91.0 (0.3)	0.5 (0.5)	0.29
Difference		2.7 (0.5)		3.5 (0.5)	-0.7 (0.7)	0.33
Self-Response	13,131		13,284			
HRA	6,201	59.7 (0.7)	6,272	60.6 (0.7)	-0.9 (0.9)	0.31
LRA	6,930	33.2 (0.4)	7,012	33.6 (0.4)	-0.4 (0.6)	0.55
Difference		26.5 (0.8)		27.0 (0.8)	-0.5 (1.2)	0.66
Internet	8,168		8,112			
HRA	4,119	39.6 (0.6)	4,048	39.1 (0.6)	0.5 (0.8)	0.51
LRA	4,049	19.4 (0.3)	4,064	19.5 (0.3)	0.1 (0.4)	0.87
Difference		20.2 (0.6)		19.6 (0.7)	0.6 (0.9)	0.52
Mail	4,963		5,172			
HRA	2,082	20.0 (0.4)	2,224	21.5 (0.4)	-1.5 (0.6)	0.02*
LRA	2,881	13.8 (0.3)	2,948	14.1 (0.3)	-0.3 (0.4)	0.43
Difference		6.2 (0.5)		7.4 (0.4)	-1.1 (0.7)	0.11
CATI	872		880			
HRA	296	9.0 (0.5)	301	9.6 (0.6)	-0.6 (0.8)	0.44
LRA	576	7.9 (0.4)	579	8.0 (0.3)	-0.1 (0.5)	0.85
Difference		1.1 (0.6)		1.6 (0.7)	-0.5 (0.9)	0.58
CAPI	5,397		5,291			
HRA	1,059	82.2 (1.0)	1,035	82.7 (0.9)	-0.5 (1.3)	0.69
LRA	4,338	85.8 (0.5)	4,256	85.0 (0.4)	0.8 (0.7)	0.23
Difference		-3.7 (1.1)		-2.3 (1.0)	-1.3 (1.5)	0.36

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

Appendix F. Consistency between Telephone Service Response with Valid Phone Number and Successful CFU Contact Supplemental Tables

Table F-1. Percent of Households that Reported Having Telephone Service that Provided a Valid Phone Number, by Treatment

Mode	Test Percent	Control Percent	Test minus Control	P-Value
Overall	96.8 (0.2)	97.0 (0.2)	-0.2 (0.3)	0.49
Internet	99.2 (0.1)	99.2 (0.1)	0.0 (0.2)	0.84
Mail	95.1 (0.3)	94.9 (0.3)	0.2 (0.5)	0.66
CAPI	95.0 (0.5)	95.8 (0.5)	-0.7 (0.7)	0.29

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance was tested based on a two-tailed t-test at the α=0.1 level.

Table F-2. Percent of Households that Reported Having Telephone Service that Provided a Phone Number and Had a Successful CFU Contact, by Treatment

Mode	Test Percent	Control Percent	Test minus Control	P-Value
Overall	73.5 (0.5)	72.3 (0.5)	1.2 (0.7)	0.07*
Internet	76.4 (0.6)	76.8 (0.6)	-0.4 (0.8)	0.60
Mail	73.4 (0.9)	69.8 (0.9)	3.6 (1.2)	<0.01*
CAPI	69.9 (1.0)	68.3 (1.1)	1.6 (1.4)	0.27

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

Appendix G. Percent Distribution of Valid Telephone Service Question Responses to Other Information from Content Test

Table G-1. Respondent's Telephone Service and Smartphone Responses, by Treatment

Respondent's Answer to Whether the Housing Unit Had Telephone Service	Respondent's Answer to Smartphone Response Category in Computer Type Question	Test Percent	Control Percent	Test minus Control	P-Value
Yes	Yes	81.4 (0.4)	81.0 (0.4)	0.4 (0.6)	0.53
No	Yes	1.0 (0.1)	2.4 (0.1)	-1.4 (0.2)	<0.01*
Yes	No	16.9 (0.4)	15.4 (0.4)	1.5 (0.6)	0.01*
No	No	0.7 (0.1)	1.2 (0.1)	-0.5 (0.1)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

Table G-2. Respondent's Telephone Service Response and Whether a Valid Phone Number was Provided, by Treatment

Respondent's Answer to Whether the Housing Unit Had Telephone Service	Did the Respondent Provide a Valid Phone Number?	Test Percent	Control Percent	Test minus Control	P-Value
Yes	Yes	95.2 (0.2)	93.6 (0.2)	1.7 (0.3)	<0.01*
No	Yes	1.4 (0.1)	3.0 (0.1)	-1.6 (0.2)	<0.01*
Yes	No	3.1 (0.2)	2.9 (0.2)	0.2 (0.3)	0.36
No	No	0.3 (0.1)	0.5 (0.1)	-0.2 (0.1)	0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.