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

November 6, 2023

2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS23-RER-05

MEMORANDUM FOR Donna M. Daily
Chief, American Community Survey Office

From: Paul C. Beatty *Paul C. Beatty*
Chief, Center for Behavioral Science Methods

Prepared by: Kathleen M. Kephart *K.M.K.*
Center for Behavioral Science Methods
Broderick E. Oliver *B.O.*
Decennial Statistical Studies Division
Gregory Mills *G.M.*
Jason Lizarraga *J.L.*
American Community Survey Office

Subject: 2022 American Community Survey Content Test Evaluation Report: Household Roster

Attached is the 2022 American Community Survey (ACS) Content Test report for Household Roster. This report presents the methods and results of the test for a revised version of the Household Roster questions.

If you have any questions about this report, please contact Kathleen Kephart at 301-763-8891.

cc:
acso.re.workgroup.list@census.gov

Lauren Contard	DSSD	Michael Risley	DSSD
Lindsay Longsine	DSSD	Samantha Spiers	DSSD
Peter Massarone Jr.	DSSD	Hyon Shin	SEHSD

This page is intentionally blank.

DRB Clearance Number—CBDRB—FY23—ACSO003—B0066

American Community Survey Research and Evaluation Program

November 6, 2023

2022 American Community Survey Content Test Evaluation Report: Household Roster

FINAL REPORT

Kathleen M Kephart
Center for Behavioral Science Methods
Broderick E Oliver
Decennial Statistical Studies Division
Gregory Mills
Jason Lizarraga
American Community Survey Office

This page is intentionally blank.

DRB Clearance Number—CBDRB—FY23—ACSO003—B0066

TABLE OF CONTENTS

DRB Clearance Number—CBDRB—FY23—ACSO003—B0066

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The U.S. Census Bureau conducted the 2022 American Community Survey (ACS) Content Test from September through December of 2022. The 2022 ACS Content Test tested the wording, format, and placement of proposed new ACS questions and proposed revisions of current ACS questions for potential inclusion in the ACS data collection instruments. The tested questions came from 10 topics. This report presents the results of this field test for Household Roster.

In preparation for the 2022 Content Test, the Census Bureau, in consultation with the Office of Management and Budget (OMB) and the Interagency Council on Statistical Policy Subcommittee on the ACS, determined which proposals solicited from over 25 federal agencies would be tested in 2022. Approved proposals for new content or changes to existing content were tested according to the ACS content change process, which includes cognitive testing and field testing.

The 2022 ACS Content Test consisted of a nationally representative sample of 120,000 housing unit addresses, excluding Puerto Rico, Alaska, and Hawaii. The sample, which was independent of production ACS, was divided evenly among three treatments, a Control treatment and two test treatments.

Like production ACS, the data collection for the 2022 ACS Content Test was conducted in two phases: a self-response phase, which lasted up to nine weeks, followed by a nonresponse follow-up phase, conducted via Computer-Assisted Personal Interviewing (CAPI). The CAPI operation lasted about one month. For households where we received a response in the original Content Test interview, a Content Follow-Up telephone reinterview was conducted to measure response error.

The purpose of this test was to evaluate if a new set of rostering instructions for the paper questionnaire and revisions to the roster questions for the internet and interviewer-administered instruments would improve the rostering procedure. The primary goal was to determine if the Test version of the household roster question would increase representation from groups that have been underrepresented in the ACS, without reducing data quality.

At a housing unit level, the Test version captured a significantly higher percentage of complex households. With respect to data quality, there were multiple indications of higher data quality in the Test version than in the Control: 1) The item missing data rate for the first roster question on the paper questionnaire, which asks respondents to report the total number of people living at the address, was significantly lower in the Test version than in the Control; 2) The percentage of households where the reported number of people living at the address was not equal to the number of people listed by name on the roster was significantly lower in the Test version than in the Control; 3) The help text was accessed a significantly lower percentage of the time in the Test internet instrument than in the Control internet instrument.

An analysis of the roster actions in the CAPI and internet modes indicates less confusion and potentially a more accurate roster in the Test version than in the Control version: 1) a higher percentage of people were originally rostered on the first screen in the Test version; 2) a significantly lower percentage of people were deleted from the Test version; 3) a higher percentage of young children (0-4) were added during the second roster screen in the Test version; 4) a higher percentage of added people were ultimately kept in the Test version than in the Control version.

Given the above findings, the recommendation of the Census Bureau is to adopt the Test version of the roster questions for the paper, internet, and CAPI instruments.

1 BACKGROUND

The U.S. Census Bureau conducted the 2022 American Community Survey (ACS) Content Test from September to December of 2022. The 2022 ACS Content Test tested the wording, format, and placement of proposed new ACS questions and proposed revisions of current ACS questions for potential inclusion in the ACS data collection instruments. The questions came from these ten ACS topics, three of which, Sewer, Electric Vehicles, and Solar Panels are new:

- Household Roster
- Sewer
- Electric Vehicles
- Solar Panels
- Supplemental Nutrition Assistance Program (SNAP)
- Educational Attainment
- Health Insurance Coverage
- Disability
- Labor Force
- Income

This report presents the results of the field test for Household Roster.

1.1 Proposals for New and Revised ACS Questions

In June 2018, the Census Bureau solicited proposals for new or revised ACS content from over 25 federal agencies. For new questions, the proposals explained why these data were needed and why other data sources that provide similar information were not sufficient. Proposals for new content were reviewed to ensure that the requests met a statutory or regulatory need for data at small geographic levels or for small populations.

The Census Bureau, in consultation with the Office of Management and Budget (OMB) and the Interagency Council on Statistical Policy Subcommittee on the ACS, determined which proposals moved forward. Approved proposals for new content or changes to current content were tested via the ACS content change process. This process includes cognitive testing and field testing. An interagency team consisting of Census Bureau staff and representatives from other federal agencies participated in development and testing activities.

In accordance with OMB's Standards and Guidelines for Statistical Surveys (OMB, 2006) and the Census Bureau's Statistical Quality Standards (U.S. Census Bureau, 2022a), the Census Bureau conducted cognitive interviewing to pretest survey questions prior to field testing or implementing the questions in production.

1.2 Cognitive Testing

For the 2022 ACS Content Test, the Census Bureau contracted with Research Triangle Institute (RTI) International to conduct three rounds of cognitive testing.[1] Cognitive interviews were conducted virtually, in English and Spanish.[2] In the first round of cognitive testing, each topic tested one or two versions of the question. Based on the results of the first round, wording modifications to the questions were made and one or two versions per topic were tested in the second round. The interagency team used the results of both rounds of cognitive testing to recommend question content for the field test. For more information on the cognitive testing procedures and results from rounds one and two, see RTI International (2022a).

The third round of cognitive testing was conducted in Puerto Rico and in Group Quarters (GQ), as the 2022 ACS Content Test did not include field testing in these areas. Cognitive interviews in Puerto Rico were conducted in Spanish; GQ cognitive interviews were conducted in English. For more information on the cognitive testing procedures and results from the third round, see RTI International (2022b).

Three topics included in the cognitive testing were not included in the field test: Homeowners Association or Condominium Fees, Home Heating Fuel, and Means of Transportation to Work. For the most part, the changes to these questions are expected to either impact a small population or result in a small change in the data that would not be detectable in the Content Test. The subject matter experts recommended that cognitive testing was sufficient for these questions and that field testing was not necessary; the Interagency Council on Statistical Policy Subcommittee on the ACS agreed with this recommendation. Content changes for these topics will be implemented in production ACS in 2024.

1.3 Field Testing Household Roster in the 2022 ACS Content Test

1.3.1 Justification for Inclusion of Household Roster in the Content Test

The 2022 ACS Content Test was the first formal study by the Census Bureau of within household coverage in the ACS. The roster instructions used by the Census Bureau have changed very little since the late 1990s, while the complexity of household living arrangements has increased (Cherlin, 2010). The roster instructions tested in the 2022 ACS Content Test are the first major revision to the ACS roster instructions in over twenty years.

[1] For each test topic, subcommittees were formed to develop question wording and research requirements for cognitive testing. The subcommittees included representation from the Census Bureau and other federal agencies.

[2] Cognitive testing interviews were conducted virtually due to the COVID-19 pandemic. Interviews were attempted by videoconferencing first and were moved to phone interviews if there were technical problems with Skype or MS Teams.

Recent research produced by the Census Bureau's Undercount of Young Children Task Force reports that the 2009 ACS coverage ratio for all children ages 0 to 4 was 0.89, suggesting a substantial undercoverage problem (Jensen & Hogan 2017; Jensen, 2019). Further, the task force's research showed evidence of more undercoverage error for children who were not the biological or adopted child of the householder, such as grandchildren and children not related to the householder (for instance, foster children or the children of a roommate) (Jensen et al., 2018). These types of children may have more tenuous ties to the households in which they reside. The causes for the undercount of young children are unknown, but Census Bureau researchers conjecture that rostering error is one likely source.

In addition to undercounting young children, there is evidence of respondent confusion and burden when rostering a household. Cognitive testing conducted by Census Bureau staff indicates that respondents are sometimes confused about whom to include on the ACS roster (Ashenfelter et al., 2012; Clark, 2017).

The approach to obtaining a household roster using the automated instruments (internet and CAPI) in the ACS is to first ask the respondent to list everyone living or staying at the address, and then follow up with several coverage probes about anyone who may have been missed in this initial question. Research conducted in 2015 on the automated ACS instruments shows that most people are rostered using a simple instruction such as, "The following questions are about everyone who is living or staying at <address>. First, create a list of people." The coverage probes are questions that are posed after the initial roster question at the beginning of the automated ACS instruments, such as, "Does anyone else live or stay there?" and "Other than the people listed below, is there anyone else staying there even for a short time?"

It is important to note that over two million people (weighted count) were rostered as a result of one of the coverage probes (Clark, 2017). Without these probes, these people would have been omitted from the roster. Households with complex living situations have more difficulty with rostering and thus the Census Bureau relies on the coverage probes to refine the list of people living or staying there. However, many of the people added as a result of the current coverage probes were then deleted before the final roster, indicating that the coverage probes may need modification (Clark, 2017).

Despite responding in the affirmative to the coverage probes, 67.0 percent of respondents who said "yes" did not provide a name, and therefore, no one was added to the roster. This suggests that respondents either did not understand the coverage questions or did not want to provide the information for the additional person(s). Additionally, there is evidence of respondents likely not fully understanding the coverage probes, as respondents use the "back" button frequently while navigating the roster questions in the internet mode (Clark, 2017; Horwitz et al., 2013).

In addition to the question wording itself, some of the confusion about who to roster in a household may also be a result of the two-month rule that is used for the ACS, which is a modified de facto residence rule. The intent of the ACS current residence concept is to count everyone living or staying at the address on the day the survey is completed and to include people who might be away for a short vacation, business trip, or overnight sleepover. The use of the two-month reference in the instructions and rostering questions may be causing people who should be included in the household to be left off the roster when respondents incorrectly apply this reference period to a situation. For instance, there is an instruction to ignore the reference period if the person has no other place to live, but many respondents may not notice or may ignore this and leave these people off the roster.

1.3.2 Cognitive Testing Development for Household Roster

The roster questions were cognitively tested in English and Spanish for the paper and interviewer-administered modes by RTI over three rounds of testing (RTI International, 2022a and 2022b). The internet mode had joint cognitive and usability testing conducted by the Center for Behavioral Science Methods (CBSM) over two rounds (Olmsted-Hawala et al., 2023). The roster questions were modified based on this testing.

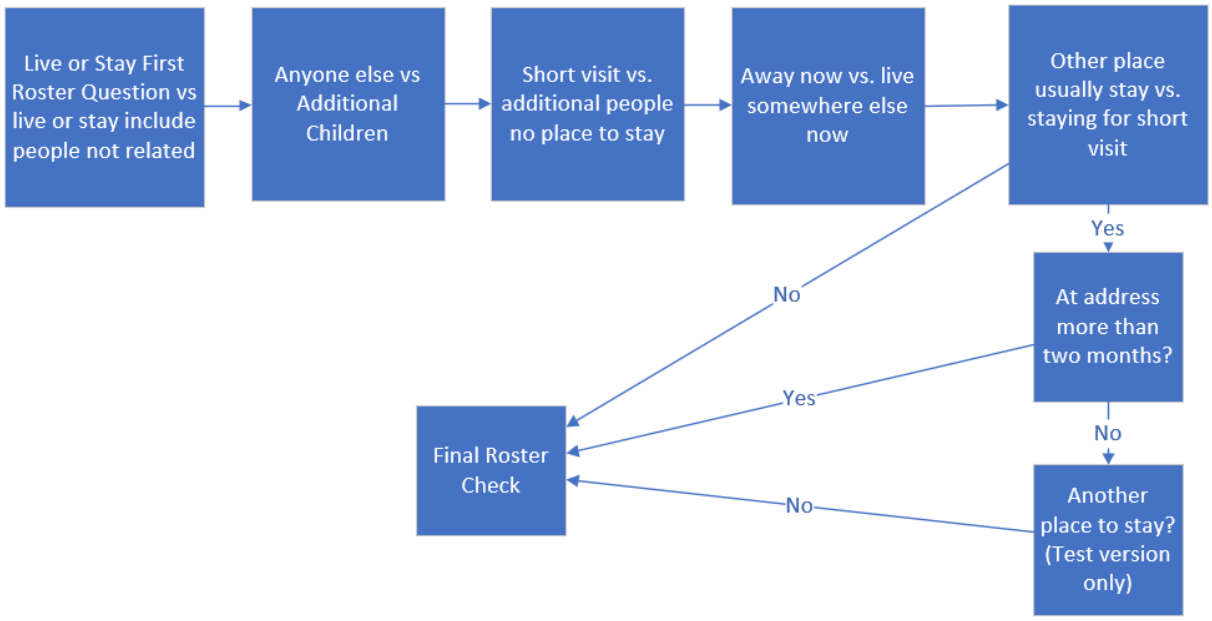
Cognitive testing was conducted to make sure that the wording used for the Content Test was clearly interpreted. Several versions of the roster wording were cognitively tested and the selected wording was found to be the least problematic of all cognitively tested versions (RTI, 2022a and 2022b). A primary goal of the Content Test is to determine if the test version results in the inclusion of more groups that have been underrepresented in the ACS, without reducing data quality in other ways.

1.3.3 Question Content

The automated versions (internet and CAPI) of the Household Roster allow for more detailed probing than the paper version. The internet and CAPI versions consist of a series of screens where the initial roster is created, people are potentially added, and people are potentially removed. Figures 4 through 30 in the Appendix present a fictitious roster for these screens for the Control and Roster Test internet instruments; the CAPI instrument screens utilize the same language and flow.

Figure 1 summarizes the branching and major differences between the Control and the Roster Test version of the Household Roster internet and CAPI screens. Overall, the flow of the questions is mostly the same across Control and Roster Test versions; the Roster Test version has one additional screen. Additionally, some of the question stems were changed in the Roster Test version.

Figure 1. Control versus Roster Test: Branching of Screens



In the internet and CAPI instruments, the first roster question asks for a list of everyone living or staying at the address, but in the Test version it explicitly says, "including people not related to you." The Control version makes no reference to people being related or not.

The second question in the Control instrument asks if ANYONE ELSE lives or stays there. In the Test instrument the second question asks about "ADDITIONAL children living or staying there, for example babies, grandchildren, or foster children? These children could be related or unrelated to you." It shifts the focus explicitly to children who may not be the biological or adopted children of the respondent. The provided examples of children are groups that past research has found are more prone to undercoverage (Jensen et al., 2018).

The third question in the Control instrument asks if there is ANYONE ELSE staying there even for a short time; in the instruction it prompts respondents to "not include overnight guests … who have a residence somewhere else." In the Test instrument, the focus is on "ADDITIONAL people staying there, for example roommates and other people or families who have no other place to stay." With the new wording, the Test instrument calls out several living situations that are more prone to rostering errors such as whole separate families and unrelated roommates. It also places the criteria that they need to have no other place to stay right in the question stem instead of as an instruction that is easily overlooked.

The fourth question presents the entire roster and is an opportunity to remove people from the roster. In the Control instrument the focus is on people "away NOW for more than two months," but in cognitive testing participants were not clear if the two months had to be in the

future or past (RTI International, 2022a). To simplify the question, the Test version asks, "do any of these people live somewhere else now?"

The fifth question is also an opportunity to remove people from the roster. In the Control it asks if people have "some other place they usually stay," whereas the Test version asks, "are any of these people staying at <address> for a short visit or for an overnight stay?"

The sixth question asks if the person selected in the fifth question is staying at the address for at least two months; it is the same in the Control and Test versions. In the Test version, if the respondent says the person is not staying at the address for at least two months they are asked if the person has another place to stay. If they do not have another place to stay, they are kept on the roster, whereas in the Control version if they are not staying at the address for at least two months they are dropped from the roster.

The Control and Roster Test paper questionnaire versions of the Household Roster are shown in Figures 2 and 3, respectively. The paper version has limited space so there is only one roster question instead of a series of rostering questions. Paper is the only version of the instrument that collects a population count. For the Roster Test paper question, we removed the mention of "2 months," reformatted the instructions, and reworded the question to remind respondents to include themselves.

Figure 2. Control: Paper Version Household Roster

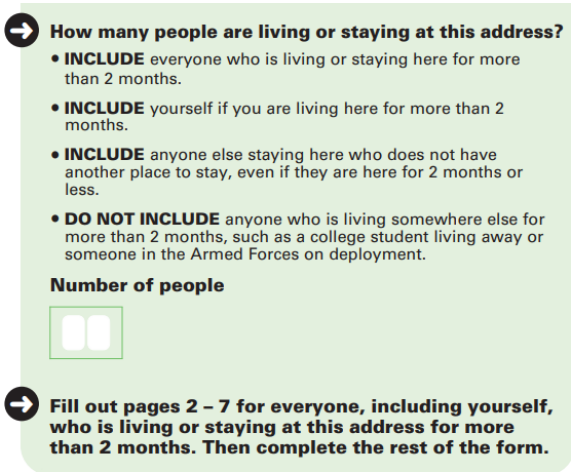
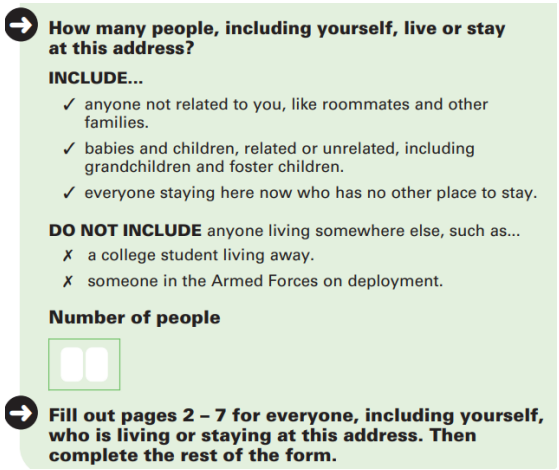


Figure 3. Roster Test: Paper Version Household Roster



1.3.4 Research Questions

The questions examined for this research are presented below.

1.3.4.1 Item Missing Data Rates for Household Roster

RQ1. *Is the item missing data rate for population count, which asks how many people live at the address, different on paper returns for the Control and Roster Test versions? (Analyze for paper mode).*

Note: The population count of how many people live or stay is not asked in the internet and CAPI modes. Therefore, the item missing data rates for the population count question is not applicable to these modes. In these modes, the instrument begins with collecting the names of everyone living or staying there. In the event they do not roster any names during the first question they cannot continue the interview.

1.3.4.2 Response Distributions for Household Roster

Count Discrepancy Research Questions

RQ2. *Are the percentage of cases with a count discrepancy for paper returns for the Control and Roster Test versions different?*

RQ3. *Are the percentage of cases of high count discrepancy for paper returns for the Control and Roster Test versions different?*

RQ4. *Are the percentage of cases of low count discrepancy for paper returns for the Control and Roster Test versions different?*

Household Characteristics Research Questions

RQ5. *Are the percentage of complex households for the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI).*

A *complex household* is defined as a household with a structure other than one in which: 1) a householder lives alone; 2) a householder lives with a married or unmarried partner without children; 3) a householder lives with a married or unmarried partner with biological or adopted children; or 4) the householder is a single parent with biological or adopted children.

Examples of complex households include:

- Blended families (i.e., stepchildren)
- Multi-generational households (grandparents, parents, and children)
- Family with other relatives (i.e., aunts, uncles, cousins, etc.)
- Skip generation (grandparents with grandchildren, no parents present)
- Family with other nonrelative(s)

RQ6. *Are the percentage of people with tenuous connections to the household for the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI).*

Note: A *tenuous connection* is defined based on each person's relationship to Person 1. This includes people who are not a partner or biological or adopted children of Person 1, such as a brother or sister, stepchildren, father or mother, parent-in-law, son-in-law or daughter-in-law, other relative, roommates, foster children, and other nonrelatives.

RQ7. *Is there a difference in the number of young children between the ages of 0-4 on the final roster for the Control and Roster Test versions? (Analyze overall and by mode: paper, internet, and CAPI).*

RQ8. *Is there a difference in household size distributions between Control and Roster Test versions? (Analyze overall and by mode: paper, internet, and CAPI).* (Compare the proportion of 1-person, 2-person, 3-person, 4-person, and 5-person or larger households).

RQ9. *Are the person characteristics between the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI*). (The person characteristics are age, race, sex, educational attainment, Hispanic origin, English-speaking ability, and Tenuous connection).

Roster Add or Delete Research Questions

RQ10. *Is there a difference in the percentage of people who were originally rostered on screen Roster_a between the Control and Roster Test versions, independent of if they stayed on the roster or were later deleted? (Analyze overall and by mode: internet and CAPI).*

RQ11. *Is there a difference in the percentage of people who were added, and kept on the final roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ12. *Is there a difference in the percentage of people who were added that are later deleted between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ13. *Is there a difference in the percentage of people who were deleted from the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ14. *Is there a difference in the percentage of households with an added household member between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ15. *Is there a difference in the percentage of households with a deleted household member between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ16. *Is there a difference in the percentage of young children (0-4 years) added to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI, and by screen Roster_A, Roster_B, and Roster_C).*

RQ17. *Is there a difference in the number of households that added a young child (0-4 years) to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI, and by screen Roster_A, Roster_B, Roster_C).*

RQ18. *Across modes and screens, is there a difference in the distribution of biological and adopted children versus all other relationship statuses (relative to Person 1) for young children (0-4 years) who were added between the Control and Roster Test versions?*

RQ19. *For the universe of households that added a person, is there a difference in the percentage of persons per household added between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ20. *For the universe of households that deleted a person, is there a difference in the percentage of persons per household deleted between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

RQ21. *Is there a difference in the characteristics of households that added someone to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).* (Characteristics include household size, tenure, building type, if no one in household over the age of 14 speaks English very well, and household type (complex versus non-complex).

RQ22. *Is there a difference in the characteristics of households that deleted someone from the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

Characteristics include household size, tenure, building type, if no one over the age of 14 speaks English very well, and household type (complex versus non-complex).

RQ23. *Is there a difference in the characteristics of the people that are added to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

Characteristics include relationship to reference person, age, race, Hispanic origin, sex, educational attainment, and English-speaking ability.

1.3.4.3 Response Reliability for Household Roster

The Content Followup (CFU) reinterview collected an independent roster by asking the same roster questions again. Respondents who received the Roster Test version were asked the Roster Test version roster questions again. Respondents who received the Control version were asked the Control version roster questions again. (See Section 2.3 for more information about the Content Followup operation.)

RQ24. *When we match the Content Roster Test roster to the CFU independent roster by name and date of birth for each person within the household, is the proportion of households with a mismatch between the Content Roster Test and CFU different between the Control and Roster Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

RQ25. *When we match at the person level between the Content Test roster and the CFU independent roster, what is the distribution of age groups (0-4, 5-17, 18-25, 25-65, 65 plus) for people who are not matched for the Control and Roster Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

RQ26. *When we match at the person level between the Content Roster Test roster and the CFU independent roster, what is the distribution of relationship type relative to Person 1 for people who are not matched for the Control and Roster Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

1.3.4.4 Other Metrics for Household Roster

We examine two metrics that provide an indication of respondent burden pertaining to the household roster question in the internet mode:

RQ27. *How often do respondents access the help screen when answering the household roster question?*

RQ28. *What are the breakoff rates for the household roster question?*

2 METHODOLOGY

2.1 Sample Design

The 2022 ACS Content Test consisted of a national sample of roughly 120,000 housing unit addresses, excluding Puerto Rico, Alaska, and Hawaii (due to cost constraints, only stateside housing units were included). The sample was independent of the ACS production sample; however, the sample design for the Content Test was largely based on the ACS production sample design, with some modifications to meet the test objectives. The ACS production sample design is described in Chapter 4 of the ACS and Puerto Rico Community Survey (PRCS) Design and Methodology report (U.S. Census Bureau, 2022b).

The sample design modifications included stratifying addresses into high and low self-response areas, oversampling addresses from the low self-response areas to ensure equal response from both strata, and selecting an initial sample of addresses, followed by a nearest neighbor method for selecting the remaining addresses for sample. The high and low self-response strata were defined based on ACS self-response rates from the 2018 and 2019 panels at the tract level.

In the sample selection process, we selected an initial sample of 40,000 addresses, then selected the two nearest neighbors for each initially selected address. If possible, we selected nearest neighbors that were in both the same Content Test sampling stratum as well as the same state, county, and sub-county area as the initially selected address. In total, three samples were selected, one for the Control treatment and two for the two test treatments. These three treatments are shown in Table 1.

The Control treatment contained production questions and questions from the three new topics: Solar Panels, Electric Vehicles, and Sewer. The Test treatment contained a test version question for all topics except Household Roster. Two of the new topics, Solar Panels and Sewer, only had one version of the test question; therefore, the same question was asked in the Control and test treatments. The other new topic, Electric Vehicles, had two versions; one was asked in the Control and Roster Test treatments and the other in the Test treatment.

The primary purpose of the Roster Test treatment was to test the household roster test question separately since changes in the amount and types of people included in the household could impact the results of person-level topics. Therefore, the analyses for Test Version 2 of the Health Insurance Coverage, Labor Force, and Income questions could have been impacted by

these changes. However, it was determined that the additional information gained from testing an additional version of the topics in the Roster Test treatment was worth the risk.[3]

Table 1. Questions by Treatment

Topic	Control Treatment	Test Treatment	Roster Test Treatment
Household Roster	Production	Production	Test Version
Solar Panels	Test Version	Test Version	Test Version
Electric Vehicles	Test Version 1	Test Version 2	Test Version 1
Sewer	Test Version	Test Version	Test Version
Educational Attainment	Production	Test Version	Production
Health Insurance Coverage	Production	Test Version 1	Test Version 2
Disability	Production	Test Version	Production
SNAP	Production	Test Version	Test Version†
Labor Force	Production	Test Version 1	Test Version 2
Income	Production	Test Version 1	Test Version 2

† The SNAP Test Version was in both test treatments to align with Labor Force and Income that also have a reference period change to the previous calendar year.

2.2 Data Collection

The 2022 ACS Content Test occurred in parallel with data collection activities for the September 2022 ACS production panel. Data collection for production ACS data consists of two main phases: an approximately two-month self-response data collection phase and a one-month follow-up phase.

During the self-response phase, addresses in sample are asked to self-respond by internet or mail. The Census Bureau sends addresses in sample up to five mailings to encourage self-response. This operation is followed by a one-month Computer-Assisted Personal Interviewing

[3] We examined differences in key household and person characteristics among the Control and Roster Test treatments to explore any indication of bias in the Health Insurance Coverage, Labor Force, and Income analyses. See Spiers et al. (2023) for more information.

(CAPI) operation, where Census Bureau field representatives attempt to complete a survey for a sub-sample of the remaining nonresponding addresses.
The following data collection protocols for the 2022 ACS Content Test remained the same as production ACS:

- Data were collected using the self-response modes of internet (in English and Spanish) and paper questionnaires for the first and second month of data collection.
- In the third month of data collection, a sub-sample of nonresponding addresses were selected for CAPI.
- During CAPI, Census Bureau field representatives conducted interviews in person and over the phone.
- Self-response via internet or paper was accepted throughout the three-month data collection period.

The following data collection protocols for the 2022 ACS Content Test differed from production ACS:

- There were no paper versions of the 2022 ACS Content Test questionnaires in Spanish.[4]
- If respondents called Telephone Questionnaire Assistance (TQA) and opted to complete the survey over the phone, the interviewers conducted the survey using the production ACS questionnaire.[5] Since the TQA interviews did not include test questions, they were excluded from the analysis of the 2022 ACS Content Test.
- The 2022 ACS Content Test did not include the Telephone Failed-Edit Follow-Up (FEFU) operation. In production, this operation follows up on households that provided incomplete information on the form or reported more than five people on the roster of a paper questionnaire.[6]
- The 2022 ACS Content Test used a telephone reinterview component to measure response reliability or response bias (depending upon the ACS topic). This telephone reinterview operation is discussed in Section 2.3 below.

[4] In 2019, 412 Spanish questionnaires were mailed back out of all mailable cases. Based upon this rate, we projected that only 8 Spanish questionnaires would be mailed back in the 2022 Content Test, which would not be cost-effective.

[5] The interviewer did not know which treatment the caller was in and therefore administered the production questionnaire. In 2019, less than one percent (0.6%) of cases responded by TQA and had no other response in a different mode. Based upon this rate, we projected about 744 TQA-only responses would be excluded from the 2022 ACS Content Test analysis.

[6] The information obtained from the FEFU improves accuracy in a production environment but confounds the evaluation of respondent behavior in the Content Test environment. For paper questionnaires, where the household size is six or more (up to 12), we only collected name, age, and sex of these additional persons, but not detailed information as we do in the FEFU operation for ACS production.

For detailed information about ACS data collection procedures, consult the ACS and PRCS Design and Methodology Report (U.S. Census Bureau, 2022b).

2.3 Content Follow-Up Operation

To measure response reliability or response bias, a Content Follow-Up (CFU) reinterview was attempted with every household with an original Content Test interview that met the CFU eligibility requirements. Among the requirements were that the household must be occupied, and the household must have a valid telephone number. See the CFU requirements document for the complete list of eligibility requirements (Spiers, 2021a).

2.3.1 Content Test Follow-Up Protocol

As in previous ACS Content Tests, a case was sent to the CFU operation no sooner than two weeks (14 calendar days) after the original interview and had to be completed within three weeks after being sent to the CFU. This timing attempted to balance two competing needs: (1) to minimize the possibility of real changes in answers due to a change in life circumstances between the two interviews; (2) to minimize the possibility of the respondent repeating their previous answer based on their recollection of the original interview response, rather than considering the most appropriate answer.

All CFU reinterviews were conducted by telephone. At the first contact with a household, interviewers asked to speak with the original respondent. If that person was not available, interviewers scheduled a callback at a time when the original respondent was expected to be available. If this respondent could not be reached at the time of the second contact, the interviewer requested to speak with any other eligible household member (a household member who is 15 years or older). CFU reinterviews for the Content Test were conducted in either English or Spanish.

The CFU data collection instrument included the questions being tested for the 2022 ACS Content Test and some production ACS questions for context. It also included questions on public assistance from the 2022 Current Population Survey Annual Social and Economic Supplement (CPS ASEC) to measure response bias in the income from the public assistance question.

The CFU collected an independent household roster by re-asking the household roster questions along with Relationship, Sex, Age, and Date of Birth. The remaining CFU questions were only asked of the original household roster members. Only the Control and Roster Test

panels collected an independent household roster. The Test panel used the original household roster to ask housing and detailed person questions.[7]

2.3.2 Content Test Follow-Up for Household Roster

The CFU collected a second household roster by asking the original roster question version the household received during the Content Test. The relationship to Person 1, Sex, Age, and Date of Birth questions were also asked of the new roster members and then the interview transitioned to only ask the remaining follow-up questions of the original interview roster members. The interviewers were trained to prevent respondents from trying to "correct" the roster when there was a discrepancy between the original and CFU household rosters.

2.4 Analysis Metrics

The sample addresses for the Control and test treatments were selected in a manner so that their response propensities and response distributions (on particular characteristics) would be the same. Similar distributions allow us to conclude that any difference in the metrics used to analyze Household Roster is attributable to differences in the wording and format. We tested these unit-level assumptions in both the original interview and the CFU reinterview. See Section 2.4.1 for details. The metrics that we used to evaluate Household Roster are presented in Section 2.4.2.

For the 2022 ACS Content Test, typical production ACS edits were not made because the primary concern of this test was how changes to existing questions and differences between versions of new questions affected the unaltered responses provided directly by respondents. For this reason, responses were not imputed either. A few edits were applied to the non-topic data, such as calculating a person's age based on his or her date of birth, but such edits were minimal.[8]

All estimates from the ACS Content Test were weighted. The final content test weights took into account the initial probability of selection (the base weight) and CAPI sub-sampling. The weights used in the CFU analysis also included an adjustment for CFU non-response.[9]

Comparisons between the Control and Test version of Household Roster were conducted using two-tailed t-tests at the α=0.1 level of significance. The Content Test sample size was chosen to

[7] The Test panel did not need to collect an independent household roster. The independent roster was needed to calculate the response reliability metrics for the Household Roster topic, which only used data from the Control and Roster Test treatments.

[8] This only refers to edits made to the data sets before analysis. During the analysis phase, additional edits, such as collapsing categories, were made based on the needs of the individual question.

[9] The Content Test weight creation process does not include all the steps followed in the ACS, including the noninterview adjustment for the original interview and calibration to housing unit and population controls (see U.S. Census Bureau, 2022b, Chapter 11). For more information on the 2022 Content Test weighting procedure, see Risley and Oliver (2022) and Keathley (2022).

provide enough statistical power (0.80) to detect a difference in the gross difference rates (measuring differences in adds and deletes from the household roster) of at least two percentage points between the Control and Roster Test groups for the Household Roster question. In statistical tests involving multiple comparisons, we controlled for the overall Type I error rate by adjusting the resulting p-values using the Hochberg method (Hochberg, 1988).[10]

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, 2022b, Chapter 12). We calculated the variance for each rate and difference using the formula below. The standard error of an estimate (X_0) is the square root of the variance:

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

where:

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

2.4.1 Unit-Level Analysis

The unit response rate is important, as it provides an indication of the quality of the survey data. As part of our analysis, we examined unit-level (i.e., address-level) responses for the Control and test versions in the original interviews and CFU reinterviews. These results are provided in a separate report (Spiers et al., 2023).[11]

2.4.2 Topic-Level Analysis

To evaluate the proposed changes to the household roster question, we calculated a variety of metrics, presented in Sections 2.4.2.1 through 2.4.2.4.

2.4.2.1 Item Missing Data Rates

We calculated the item missing data rate for the population count question on the front of the paper questionnaire as the proportion of eligible housing units for which a required response to this question was missing. A high item missing data rate can be indicative of a question that lacks clarity, is sensitive, or is simply too difficult to answer.

Note: This metric was not applicable to the internet and CAPI modes, where the respondent is required to provide at least one name on the first roster screen in order to continue with the survey.

[10] Use the MULTTEST Procedure in SAS®.

[11] As part of the 2022 ACS Content Test, we analyzed respondent burden. The results of this analysis are also contained in Virgile et al. (2023).

2.4.2.2 Response Distributions

To assess how changes to household roster question affected the resulting estimates, we compared the response distributions for the Control and Roster Test versions for various constructs such as count discrepancy, complex households, tenuous connection to the household, and household size.

2.4.2.3 Response Reliability

Survey responses are subject to error. Response error occurs for a variety of reasons, such as flaws in the survey design, misunderstanding of the questions, misreporting by respondents, and interviewer effects.

The accuracy of the household roster is important to the accuracy of ACS estimates. To evaluate the reliability of the household roster collected in the original interview, the CFU operation collected an independent household roster for eligible households, by re-asking the household roster questions along with Relationship, Sex, Age, and Date of Birth. We measured response reliability via different roster match metrics between the roster collected in the original interview and the roster collected in the CFU reinterview.

2.4.2.4 Other Metrics

Two additional questions investigate if the Roster Test version of the household roster question increases respondent burden. The metrics examined pertain to how often respondents who answer online seek online help and how often they break off from the questions.

3 DECISION CRITERIA

Before field testing Household Roster, a team of subject matter experts identified and prioritized which of the research questions presented in Section 1.3.4 would determine which version of Household Roster would be recommended for inclusion in the ACS. The decision criteria for Household Roster are presented in Table 2.

Table 2. Decision Criteria for Household Roster

Priority	Research Question	Decision Criteria
1	7	The number of children between the ages of 0-4 on the final roster is significantly higher in the Roster Test version compared to the Control version.
2a	24	The proportion of households with a mismatch between the Content Roster Test and CFU roster is lower in the Roster Test version than the Control version.
2b	6	The percentage of people with tenuous connections to the household is significantly higher in the Roster Test version compared to the Control version.
3	5	The percentage of complex households for the Roster Test version is significantly higher than the Control version.
4	2, 3, 4	The percentage of cases with a count discrepancy for the Roster Test version is lower than the number of cases with a count discrepancy for the Control version.
5	1	The item missing data rates for the Roster Test version are the same or lower than the item missing data rates for the Control version.

Research questions not included in the decision criteria are for informational purposes only.

4 ASSUMPTIONS AND LIMITATIONS

4.1 Assumptions

- The sample addresses for the Control and Roster Test treatments were selected in a manner so that their response propensities and response distributions would be the same. This assumption of homogeneity allows us to conclude that any difference between treatments is attributable to differences in wording and format. See Section 5 for more details.

- There was no difference between treatments in mail delivery timing or subsequent response time. The treatments had the same sample size and used the same postal sort and mailout procedures. Previous research indicated that postal procedures alone could cause a difference in response rates at a given point in time between experimental treatments of different sizes, with response for the smaller treatments lagging (Heimel, 2016).

- We assume that the frequency of real changes in answers due to a change in life circumstances between the original interview and CFU reinterview were similar between treatments.

4.2 Limitations

- Housing units from Alaska, Hawaii, and Puerto Rico were not included in the sample for the 2022 ACS Content Test. The results of the Content Test may not extend to the housing unit population in these areas.

- The paper questionnaire was only available in English and was not available in Spanish like in production. The Content Test results related to the English paper questionnaire may not extend to Spanish paper questionnaire.

- For paper questionnaires, where the household size is six or more (up to 12), we only collected name, age, and sex of these additional persons. Detailed information for these persons in ACS production are collected in the FEFU operation. We did not include the FEFU operation because the information collected from it improves accuracy and could confound respondent behavior in the Content Test environment.

- We did not have response data for some partial internet responses (179 cases) because of a server issue. These cases were excluded from the analyses.

- TQA responses were excluded from the analysis of the 2022 ACS Content Test response data because survey responses completed via the TQA operation were only conducted using the ACS production data collection instrument.

- CAPI interviewers were assigned 2022 ACS Content Test cases as well as regular production cases. The potential risk of this approach is the introduction of a cross-contamination or carry-over effect among Control and test treatments and production due to the same interviewer administering multiple versions of the same question item (despite their training to read questions verbatim).

- Because of budget constraints, the CAPI workload could not exceed 28,000 housing units. This workload was less than what was subsampled originally because we over-sampled addresses in low response areas. Limiting the CAPI workload caused an increase in the variances for the analysis metrics used.

- The CFU reinterviews were conducted by phone only, whereas the original interviews were completed online, by mail, by phone in CAPI, and in person in CAPI. Hence, some of the differences observed between the original interviews and the CFU interviews may be the result of mode effects.

- Not all households who provided a response in the original interview were eligible for the CFU reinterview (see Section 2.3 for more information). As a result, 2.5 percent (standard error 0.2) of households from the original Control interviews, 2.5 percent (standard error 0.2) of households from the original Test interviews, and 3.0 percent (standard error 0.2) of households from the original Roster Test interviews were not eligible for the CFU reinterview. These rates were not significantly different between treatments (chi-square p-value 0.11).

- We reinterviewed the same person who responded in the original interview when possible but accepted interviewing a different person from the same household after two unsuccessful attempts at reaching the original person. Therefore, differences in results between the original interview and CFU reinterview for these cases could partly be from different people answering the questions. We interviewed a different household member in CFU for 7.3 percent (standard error 0.4) of CFU Control cases, 9.4 percent (standard error 0.5) of CFU Test cases, and 8.5 percent (standard error 0.5) of CFU Roster Test cases. These rates were significantly different between treatments (chi-square p-value 0.01) with the rate of CFU Test cases (t-test p-value <0.01) and CFU Roster Test cases (t-test p-value 0.04) being significantly higher than the rate of CFU Control cases.

- We examined potential differences between CFU respondents and nonrespondents within some socioeconomic and demographic characteristics because there were differences in the 2016 CFU reinterview (Spiers, 2021b). For all treatments combined, there were significant differences between CFU respondents and nonrespondents for *household size, tenure, age, race, Hispanic origin, language of original interview response,* and *high and low response areas*. These differences are similar to the ones found in the 2016 CFU (Spiers, 2021b).

- The 2022 ACS Content Test did not include the production weighting adjustments for unit nonresponse or population controls which are designed to minimize nonresponse and under-coverage bias. As a result, any estimates derived from the Content Test data did not provide the same level of inference as the production ACS and cannot be compared to production estimates.

5 RESULTS

This section of the report presents the results of various metrics used to evaluate Household Roster. The comparisons presented assume homogeneity of the response distributions for the three treatments, prior to the field test. We tested this assumption using unit-level (i.e., address level) analyses. The results are presented in Spiers et al. (2023).

In general, the overall unit response rates were not significantly different between treatments, nor were the response rate portions by mode. Additionally, when examining demographic and socioeconomic distributions, none of the response distributions were significantly different between treatments.

There is no evidence of underlying CFU response rate issues that would negatively affect topic-level response error analyses. When examining demographic and socioeconomic distributions, none of the overall response distributions were significantly different between the Control and Test version except for tenure.

5.1 Benchmark Results for Household Roster

There are no benchmark measures for household roster. Hence, this metric is not applicable.

5.2 Item Missing Data Rate Results for Household Roster

RQ1. *Is the item missing data rate for population count, which asks how many people live at the address, different on paper returns for the Control and Roster Test versions? (Analyze for paper mode).*

The paper form asks respondents to report the total number of people living at the address. This question appears on the front page of the paper questionnaire and is preceded by instructions that explain who should be included in the count and who should be excluded. The paper mode has limited space and a full roster of names and additional roster questions cannot be asked, instead the paper version collects only a population count. It is crucial to prime respondents to think of all people who live or stay at the address before they start providing names one-by-one and answering person-level questions on the next page.

The results of this household-level analysis in Table 3 indicate that the item missing data rate for the reported population count is significantly lower for the Roster Test version (Test) (which has modified instructions before the population count question) than the Control version.

Table 3. Item Missing Data Rate for Reported Population Count on Paper Instrument

Mode	Control	Test	Difference	P-value
Paper	5.1(0.5)	3.0(0.4)	-2.1(0.7)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.
An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.3 Response Distribution Results for Household Roster

RQ2. *Are the percentage of cases with a count discrepancy for paper returns for the Control and Roster Test versions different?*

RQ3. *Are the percentage of cases of high count discrepancy for paper returns for the Control and Roster Test versions different?*

RQ4. *Are the percentage of cases of low count discrepancy for paper returns for the Control and Test versions different?*

As discussed in RQ1, the paper form asks respondents to first report the total number of people living at the address and then asks the respondent to roster the people (report their names and demographics). In the paper instrument:

- A *count discrepancy* occurs when the number of people listed by name on the roster does not equal the reported number of people living at the address.
- A *high count discrepancy* occurs when the number of people listed by name on the roster is greater than the reported number of people living at the address.
- A *low count discrepancy* occurs when the number of people listed by name on the roster is less than the reported number of people living at the address.

The results of the household-level analysis in Table 4 (for the paper instrument) indicate that the count discrepancy rate for the Test version is significantly lower than those of the Control version (about 0.8 percent lower). This means that the percentage of households where the reported number of people living at the address was not equal to the number of people listed by name was significantly lower in the Test version than in the Control.

The main source of this difference is an improved low count discrepancy for the Test version. In the event of a low count discrepancy, when a higher number of people are reported in the first question than the number of names later provided, interview followup is required. This means there is a potential data collection cost savings as well as an improvement in data quality in the Test version.

Table 4. Count Discrepancy Rates

Mode	Discrepancy Metric	Control	Test	Difference	P-value
Paper	Count discrepancy	2.8(0.4)	2.0(0.3)	-0.8(0.5)	0.09*
Paper	High Count Discrepancy	2.0(0.3)	1.6(0.2)	-0.3(0.4)	0.42
Paper	Low Count Discrepancy	0.8(0.2)	0.3(0.1)	-0.4(0.2)	0.02*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ5. *Are the percentage of complex households for the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI).*

A *complex household* is defined as a household other than one in which: 1) a householder lives alone; 2) a householder lives with a married or unmarried partner without children; 3) a householder lives with a married or unmarried partner with biological or adopted children; or 4) the householder is a single parent with biological or adopted children.

Examples of complex households:

- Blended families (i.e., stepchildren)
- Multi-generational households (grandparents, parents, and children)
- Family with other relatives (i.e., aunts, uncles, cousins, etc.)
- Skip generation (grandparents with grandchildren, no parents present)
- Family with other nonrelative(s)

The results of the household-level analysis in Table 5 indicate that the Test version has a significantly higher percentage of complex households than the Control version across all modes and among internet respondents. There are no significant differences in the paper and CAPI modes.

The paper mode only added an instruction to include people not related to you, it did not modify the question wording like the automated modes. Given how often instructions are overlooked it is not surprising that we did not see significantly more complex households in the paper mode.

In the CAPI mode, interviewers often can provide additional clarification and mitigate the effectiveness of wording changes between experimental and Control versions across multiple questions.

Table 5. Percentage of Complex Households

Mode	Control	Test	Difference	P-value
Across Mode	14.7(0.3)	15.6(0.4)	0.9(0.5)	0.08*
Internet	14.0(0.4)	15.2(0.5)	1.1(0.7)	0.09*
Paper	10.0(0.8)	10.0(0.6)	0.1(0.9)	0.97
CAPI	20.1(1.1)	20.5(1.1)	0.4(1.5)	0.80

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number CBDRB-FY23- ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ6. *Are the percentage of people with tenuous connections to the household for the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI).*

A *tenuous connection* is defined based on each person's relationship to person 1, the reference person. This relationship includes people who are not a partner or biological or adopted children, such as a brother or sister, stepchildren, father or mother, parent-in-law, son-in-law or daughter-in-law, other relative, roommates, foster children, and other nonrelatives.

The results of the person-level analysis in Table 6 indicate that the Control and Test versions do not differ significantly on the percentage of persons with tenuous connections to the reference person in any of the modes examined. When we added new language to include people not related to you, the hypothesis was that we would see an increase in the percentage of "tenuously attached people."

For this analysis a tenuous connection was defined as anything other than a partner or child of the reference person. However, future analysis may find a significant difference if the definition of tenuous is further restricted only to unrelated people or extended family.

Table 6. Percentage of Persons with a Tenuous Connection

Mode	Control	Test	Difference	P-value
Across Mode	10.0(0.3)	10.2(0.3)	0.2(0.4)	0.57
Internet	9.2(0.3)	9.4(0.3)	0.2(0.4)	0.68
Paper	7.5(0.6)	7.3(0.5)	-0.3(0.7)	0.70
CAPI	13.6(0.8)	14.0(0.7)	0.4(1.1)	0.74

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ7. *Is there a difference in the number of children between the ages of 0-4 on the final roster for the Control and Roster Test versions? (Analyze overall and by mode: paper, internet, and CAPI).*

The results of the person-level analysis in Table 7 indicate that the Control and Test versions do not differ significantly on the number of young children, ages 0 to 4, on the final roster in any of the modes examined.

Table 7. Number of Young Children (0-4) on Final Roster

Mode	Control	Test	Difference	P-value
Across Mode	12,720,000(486,385)	12,520,000(477,556)	-200,000(656,026)	0.76
Internet	8,848,000(384,274)	8,672,000(391,793)	-176,000(519,410)	0.73
Paper	552,400(78,232)	675,800(105,648)	123,400(122,221)	0.31
CAPI	3,323,000(291,189)	3,174,000(275,073)	-149,000(398,971)	0.71

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0071. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ8. *Is there a difference in household size distributions between Control and Roster Test versions? (Analyze overall and by mode: paper, internet, and CAPI).* (Compare the proportion of 1-person, 2-person, 3-person, 4-person, and 5-person or larger households).

We hypothesized that there would be a significant increase in the proportion of larger households in the Test version since it encouraged participants to include people not related to them and additional families they live with. However, we did not find a significant difference in the proportion of households at any size between Control and Test. See the results in Table 8.

Table 8. Household Size Distribution: Control Treatment versus Test Treatment

Mode	Rao-Scott Chi-Square P-value
Overall	$p = 0.58$
Internet	$p = 0.74$
Paper	$p = 0.60$
CAPI	$p = 0.54$

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance.

RQ9. *Are the person characteristics between the Control and Roster Test versions different? (Analyze overall and by mode: paper, internet, and CAPI).* (The person characteristics are age, race, sex, educational attainment, Hispanic origin, English-speaking ability, and Tenuous connection).

The results of the Rao-Scott Chi-Square comparisons for the distributions of the person characteristics of interest are shown in Table 9. The only characteristics where the distributions differ significantly between the Control and Test version are Sex and English-Speaking Ability in the paper mode.

To determine the direction of the differences between the Control and Test versions for Sex and English-speaking ability (paper mode), we compared their respective percentages via two-tailed t-tests. The results are shown in Table 10.

The Test version has a significantly higher percentage of males than the Control version (marginal difference) and a significantly higher percentage of people who speak a foreign language, but also speak English well.

Both of these populations are subject to undercoverage error, so this is a positive result. We do not have an explanation for why we found improved coverage for these populations. Keep in mind that the number of respondents who respond in the paper mode is significantly less than those who respond in the internet mode.

Table 9. Person Characteristics Distribution: Control Treatment versus Test Treatment

Mode	Response Distribution	Rao-Scott Chi-Square P-value
Overall	***Age***	$p = 0.40$
Internet		$p = 0.62$
Paper		$p = 0.99$
CAPI		$p = 0.62$
Overall	***Race***	$p = 0.57$
Internet		$p = 0.51$
Paper		$p = 0.76$
CAPI		$p = 0.24$
Overall	***Sex***	$p = 0.30$
Internet		$p = 0.49$
Paper		$p = 0.10$*
CAPI		$p = 0.97$
Overall	***Educational Attainment***	$p = 0.67$
Internet		$p = 0.68$
Paper		$p = 0.55$
CAPI		$p = 1.00$
Overall	***Hispanic Origin***	$p = 0.93$
Internet		$p = 0.28$
Paper		$p = 0.90$
CAPI		$p = 0.61$
Overall	***English-Speaking Ability***	$p = 0.89$
Internet		$p = 0.37$
Paper		$p = 0.01$*
CAPI		$p = 0.61$
Overall	***Tenuous Connection***	$p = 0.57$
Internet		$p = 0.68$
Paper		$p = 0.70$
CAPI		$p = 0.74$

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance.

Table 10. Person Characteristics: Source of Difference

Mode	Characteristic	Category	Control	Test	Difference	P-value
Paper	Sex	Male	45.1(0.6)	46.4(0.6)	1.3(0.8)	0.10*
		Female	55.0(0.6)	53.6(0.6)	-1.3(0.8)	0.10*
Paper	English-Speaking Ability	Well	83.8(2.4)	90.2(1.4)	6.4(2.6)	0.01*
		Not Well	16.2(2.4)	9.8(1.4)	-6.4(2.6)	0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ10. *Is there a difference in the percentage of people who were originally rostered on screen Roster_A between the Control and Roster Test versions, independent of if they stayed on the roster or were later deleted? (Analyze overall and by mode: internet and CAPI).*

The results of the person-level analysis in Table 11 indicate that the percentage of people who were originally rostered via the Roster_A screen (see Figures 6 and 7 in the Appendix) is significantly higher in the Test version than the Control version, across mode (internet and CAPI combined) and in the internet mode.

This result is positive, especially when coupled with the fact the percentage of deletes across all screens was lower in the Test version, meaning a higher percentage of people were added on Roster_A, and kept on the final roster. (See RQ11.)

Table 11. Originally Rostered Persons

Mode	Control	Test	Difference	P-value
Across Mode	98.0(0.2)	98.6(0.1)	0.7(0.2)	<0.01*
Internet	97.7(0.2)	98.5(0.2)	0.8(0.3)	<0.01*
CAPI	98.8(0.3)	99.1(0.2)	0.2(0.3)	0.52

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ11. *Is there a difference in the percentage of people who were added and kept on the final roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of the person-level analysis in Table 12 indicate that a significantly higher percentage of people were added on the Roster_B and Roster_C screens (see Figures 8-15 in the Appendix) and kept in the Test version than the Control version, across mode and in the internet mode.

This is an important finding, as past research has found a large percentage of people added after the first roster screen are just deleted before the final roster (Clark, 2017). The Test version seems to have helped overcome this issue; 86.6 percent of people added after the first screen were kept on the final roster.

Table 12. Added and Kept on Final Roster

Mode	Control	Test	Difference	P-value
Across Mode	62.1(3.2)	86.6(3.0)	24.6(4.2)	<0.01*
Internet	61.1(3.5)	91.6(2.1)	30.5(4.2)	<0.01*
CAPI	67.5(9.2)	64.8(10.1)	-2.7(13.3)	0.84

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ12. *Is there a difference in the percentage of people who were added that were later deleted between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of the person-level analysis in Table 13 indicate that a significantly higher percentage of persons were added via the Roster_B and Roster_C screens (see Figures 8-15) and later deleted in the Control version than the Test version, across mode and in the internet mode.

Note: The denominator for this analysis is people who were not originally listed on the Roster_A screen but were added on the Roster_B or Roster_C screens and then deleted before the final roster.

Table 13. Added and Later Deleted

Mode	Control	Test	Difference	P-value
Across Mode	38.0(3.2)	13.4(3.0)	-24.6(4.2)	<0.01*
Internet	38.9(3.5)	8.4(2.1)	-30.5(4.2)	<0.01*
CAPI	32.5(9.2)	35.2(10.1)	2.7(13.3)	0.84

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ13. *Is there a difference in the percentage of people who were deleted from the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of the person-level analysis in Table 14 indicate that a significantly higher percentage of persons were deleted from the roster in the Control version than the Test version, across mode and in the internet mode.

Note: The denominator for this analysis includes all people regardless of which screen they were added on (Roster_A, Roster_B, or Roster_C).

Table 14. Deleted from Roster

Mode	Control	Test	Difference	P-value
Across Mode	3.0(0.2)	2.3(0.1)	-0.6(0.2)	<0.01*
Internet	3.2(0.2)	2.5(0.2)	-0.8(0.2)	<0.01*
CAPI	2.1(0.3)	1.9(0.3)	-0.2(0.4)	0.52

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ14. *Is there a difference in the percentage of households with an added household member between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of this household-level analysis in Table 15 indicate that there is no significant difference between the Control and Test versions in the percentage of households where at least one person was added to the roster (regardless of whether they were kept on the roster or later deleted).

Table 15. Households with an Added Member

Mode	Control	Test	Difference	P-value
Across Mode	2.2(0.2)	2.0(0.2)	-0.2(0.3)	0.46
Internet	2.5(0.2)	2.3(0.2)	-0.3(0.3)	0.40
CAPI	1.2(0.3)	1.3(0.2)	0.1(0.4)	0.87

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ15. *Is there a difference in the percentage of households with a deleted household member between Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of this household-level analysis in Table 16 indicate that the Control version has a significantly higher percentage of households where at least one person was deleted from the roster than the Test version, across mode and in the internet mode.

Table 16. Households with a Deleted Member

Mode	Control	Test	Difference	P-value
Across Mode	5.7(0.3)	4.7(0.3)	-1.0(0.4)	0.01*
Internet	6.2(0.3)	5.0(0.3)	-1.2(0.4)	<0.01*
CAPI	4.2(0.6)	3.8(0.5)	-0.3(0.7)	0.64

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ16. *Is there a difference in the number of young children (0-4 years) added to the roster between the Control and Roster Test versions? (Analyze for internet and CAPI and by screen: Roster_A, Roster_B, and Roster_C).*

The results of the person-level analysis in Table 17 indicate that the number of children, 0-4 years of age, originally rostered via the Roster_A screen and the number of children subsequently added to the roster via the Roster_C screen does not differ significantly between the Control and the Test version. However, the number of children, 0-4 years of age added to the roster via the Roster_B screen is significantly higher in the Test version, across mode (internet and CAPI combined) and in the internet mode. Across all screens, there are no significant results.

Table 17. Number of Young Children (0-4 years old) Added to the Roster

Screen	Mode	Control	Test	Difference	P-value
Roster A	**Overall**	11,820,000(479,531)	11,210,000(444,910)	-610,000(654,969)	0.36
	Internet	8,653,000(391,481)	8,079,000(360,818)	-574,000(511,286)	0.26
	CAPI	3,164,000(258,261)	3,134,000(274,423)	-30,000(370,427)	0.94
Roster B	**Overall**	245,800(87,250)	556,700(97,313)	310,900(133,000)	0.02*
	Internet	161,000(50,093)	529,600(97,513)	368,600(111,774)	<0.01*
	CAPI	84,820(76,230)	27,190(9,609)	-57,630(76,655)	0.45
Roster C	**Overall**	108,000(64,983)	74,910(51,103)	-33,090(84,173)	0.69
	Internet	33,900(14,147)	62,840(50,653)	28,940(53,268)	0.59
	CAPI	74,150(63,491)	12,080(7,018)	-62,070(64,674)	0.34
Across Screens	**Overall**	12,170,000(481,196)	11,850,000(464,774)	-320,000(645,860)	0.61
	Internet	8,848,000(384,274)	8,672,000(391,793)	176,000(519,410)	0.73
	CAPI	3,323,000(291,189)	3,174,000(275,073)	149,000(398,971)	0.71

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0071. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ17. *Is there a difference in the number of households that added a young child (0-4 years) to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI, and by screen Roster_A, Roster_B, Roster_C).*

The results of the household-level analysis in Table 18 indicate that there is no significant difference between the Control and Test version with respect to the number households that added a young child, 0-4 years of age via the Roster_A screen, or the Roster_C screen. However, the number of households that added a child, 0-4 years of age via the Roster_B screen is significantly higher in the Test version, across mode (internet and CAPI combined) and in the internet mode. Across all screens, there are no significant differences in the number of young children between the Test and Control versions on the final roster.

Table 18. Number of Households with Young Children (0-4) Added to the Roster

Screen	Mode	Control	Test	Difference	P-value
Roster A	**Overall**	9,280,000(348,595)	8,820,000(319,669)	-460,000(486,297)	0.35
	Internet	6,775,000(293,901)	6,386,000(265,955)	-389,000(390,363)	0.32
	CAPI	2,504,000(185,860)	2,435,000(210,801)	-69,000(281,364)	0.80
Roster B	**Overall**	164,400(47,753)	506,400(84,430)	342,000(98,124)	<0.01*
	Internet	134,600(43,096)	479,300(84,538)	344,700(95,805)	<0.01*
	CAPI	29,780(25,631)	27,190(9,609)	-2,590(27,103)	0.92
Roster C	**Overall**	80,390(42,562)	48,180(26,548)	-32,210(50,977)	0.53
	Internet	31,100(13,726)	36,110(25,636)	5,010(29,692)	0.87
	CAPI	49,290(40,381)	12,080(7,018)	-37,210(41,111)	0.37
Across Screens	**Overall**	9,510,000(346,186)	9,333,000(335,825)	-177,000(484,235)	0.72
	Internet	6,927,000(288,398)	6,871,000(288,799)	-56,000(399,027)	0.89
	CAPI	2,583,000(195,618)	2,462,000(210,361)	-121,000(290,720)	0.68

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0071. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ18. *Across modes and screens, is there a difference in the distribution of biological and adopted children versus all other relationship statuses (relative to Person 1) for young children (0-4 years) who were added between the Control and Roster Test versions?*

The results in Table 19 of a Rao-Scott Chi Square test comparing the distribution of biological and adopted children versus all other relationship statuses (relative to person 1) for the Control version versus the Test version (across mode and screens) shows that there is no significant difference between the distributions.

Table 19. Biological and Adopted Children Distribution: Control versus Test Treatment

Mode	Rao-Scott Chi-Square P-value
Across Mode and Screens	$p = 0.62$

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance.

RQ19. *For the universe of households that added a person, is there a difference in the number of persons per household added between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of the household-level analysis in Table 20 indicate that there is no significant difference between the Control and Test version across mode and in the internet mode with respect to the average number of persons added per household for the universe of households that added a person. However, in the CAPI mode, the average number of adds per household is significantly larger in the Control version.

Table 20. Average Number of Adds in Households with Adds

Mode	Control	Test	Difference	P-value
Across Mode	1.5(0.1)	1.5(0.1)	<0.1 (0.1)	0.89
Internet	1.5(0.1)	1.6(0.1)	0.1(0.1)	0.40
CAPI	1.7(0.2)	1.2(0.1)	-0.4(0.2)	0.10*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ20. *For the universe of households that deleted a person, is there a difference in the number of persons per household deleted between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).*

The results of the household-level analysis in Table 21 indicate that there is no significant difference between the Control and Test version in any of the modes examined with respect to the average number of persons deleted per household for the universe of households that deleted a person, across mode and in the internet mode.

Table 21. Average Number of Deletes in Households with Deletes

Mode	Control	Test	Difference	P-value
Across Mode	1.3(<0.1)	1.3(<0.1)	-0.1(<0.1)	0.14
Internet	1.4(<0.1)	1.3(<0.1)	-0.1(<0.1)	0.15
CAPI	1.3(0.1)	1.2(0.1)	<0.1(0.1)	0.66

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ21. *Is there a difference in the characteristics of households that added someone to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).* (Characteristics include household size, tenure, building type, if no one in household over the age of 14 speaks English very well, and household type (complex versus non-complex)).

The results of this household-level analysis are presented across mode (internet and CAPI combined) in Table 22 and by mode in Tables 23 (internet mode) and 24 (CAPI mode). The p-values have been adjusted within household characteristic to control the type I familywise error rate.

The only significant results are in the CAPI mode for Household Size and Building Type. Household sizes of 5 or more and one-family houses in the Control were significantly more likely to have an added person than the Test version. For small apartment buildings, a significantly higher percentage of households had an add in the Test version than in the Control. We do not have an explanation for this, but it affects a relatively small number of cases and doesn't seem to have an impact on overall coverage.

Table 22. Characteristics of Households that Added Persons to the Roster (Across Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	0.7(0.5)	<0.1(<0.1)	-0.7(0.5)	0.57
	2	24.1(3.4)	20.2(4.0)	-4.0(5.0)	0.57
	3	20.8(3.1)	31.6(4.2)	10.7(5.6)	0.27
	4	21.6(3.1)	18.9(3.5)	-2.6(4.6)	0.57
	5 or more	32.9(3.5)	29.4(3.4)	-3.5(4.9)	0.57
Household Tenure	Own	70.4(4.2)	61.8(4.3)	-8.6(6.0)	0.16
	Rent	29.6(4.2)	38.1(4.3)	8.5(6.0)	0.16
Building Type	Mobile Home	3.2(0.9)	4.3(1.7)	1.1(2.1)	0.59
	One-Family House	80.1(3.1)	72.9(3.6)	-7.2(4.5)	0.44
	Small Apt. Building	7.5(1.8)	15.6(3.2)	8.1(3.7)	0.15
	Large Apt. Building	9.2(2.4)	7.0(2.2)	-2.1(3.0)	0.59
	Other Type Residence	0.1(0.1)	0.2(0.1)	0.1(0.1)	0.59
Household Type	Complex	59.7(4.1)	57.5(4.7)	-2.2(6.3)	0.73
	Not Complex	40.3(4.1)	42.5(4.7)	2.2(6.3)	0.73
English-Speaking Ability	Well	79.3(6.1)	84.7(5.2)	5.4(8.6)	0.53
	Not Well	20.7(6.1)	15.3(5.2)	-5.4(8.6)	0.53

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 23. Characteristics of Households that Added Persons to the Roster (Internet Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	0.5(0.5)	<0.1(<0.1)	-0.5(0.5)	0.63
	2	27.0(3.8)	19.9(4.1)	-7.0(5.4)	0.57
	3	22.2(3.3)	34.9(4.4)	12.7(5.8)	0.14
	4	21.8(3.3)	15.7(3.3)	-6.2(4.6)	0.57
	5 or more	28.6(3.6)	29.5(3.7)	1.0(5.2)	0.85
Household Tenure	Own	71.1(4.7)	67.0(4.3)	-4.2(0.5)	0.53
	Rent	28.9(4.7)	33.1(4.3)	4.2(0.5)	0.53
Building Type	Mobile Home	3.7(1.1)	4.0(2.0)	0.3(2.5)	0.95
	One-Family House	78.9(3.5)	78.6(3.5)	-0.2(4.8)	0.95
	Small Apt. Building	7.1(2.0)	9.2(1.9)	2.1(2.7)	0.95
	Large Apt. Building	10.2(2.9)	8.0(2.5)	-2.2(3.5)	0.95
	Other Type Residence	0.1(0.1)	0.2(0.1)	0.1(0.2)	0.95
Household Type	Complex	58.1(4.5)	54.8(4.8)	-3.4(7.1)	0.63
	Not Complex	41.9(4.5)	45.2(4.8)	3.4(7.1)	0.63
English-Speaking Ability	Well	81.0(6.5)	89.1(5.7)	8.1(9.1)	0.37
	Not Well	19.0(6.5)	10.9(5.7)	-8.1(9.1)	0.37

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 24. Characteristics of Households that Added Persons to the Roster (CAPI Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	1.6(1.5)	<0.1(<0.1)	-1.6(1.5)	0.60
	2	7.0(3.8)	21.3(10.1)	14.3(10.4)	0.60
	3	12.7(7.5)	15.3(9.3)	2.7(12.4)	0.83
	4	20.0(9.5)	34.9(12.5)	14.9(14.3)	0.60
	5 or more	58.7(10.9)	28.4(7.1)	-30.2(12.3)	0.07*
Household Tenure	Own	66.6(10.5)	38.6(11.8)	-28.0(16.7)	0.10
	Rent	33.4(10.4)	61.4(11.8)	28.0(16.7)	0.10
Building Type	Mobile Home	<0.1(<0.1)	5.8(3.1)	5.8(3.1)	0.13
	One-Family House	87.4(5.5)	46.2(12.0)	-41.2(12.2)	<0.01*
	Small Apt. Building	9.6(4.6)	45.5(12.5)	35.9(12.9)	<0.01*
	Large Apt. Building	3.0(2.3)	2.5(1.9)	-0.5(3.0)	0.87
	Other Type Residence	------	------	------	------
Household Type	Complex	69.3(10.6)	71.0(11.5)	1.8(14.7)	0.90
	Not Complex	30.7(10.6)	29.0(11.5)	-1.8(14.7)	0.90
English-Speaking Ability	Well	73.2(13.5)	64.3(15.2)	-8.8(19.3)	0.65
	Not Well	26.8(13.5)	35.7(15.2)	8.8(19.3)	0.65

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance. The (----) symbol indicates that the statistic could not be computed because the categories in both treatments have a frequency of zero.

RQ22. *Is there a difference in the characteristics of households that deleted someone from the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).* Characteristics include household size, tenure, building type, if no one in household over the age of 14 speaks English very well, and household type (complex versus non-complex).

The results of this household-level analysis are presented across mode (internet and CAPI combined) in Table 25 and by mode in Tables 26 (internet mode) and 27 (CAPI mode). The p-values have been adjusted within household characteristic to control the type I familywise error rate.

The only significant results are in the internet mode for household tenure where more renter households deleted someone in the Test version than in the Control. Given the high mobility of renter households, it is possible that they have more people coming and going than homeowners. For example, someone may have rostered a person who was couch surfing during Roster_A, Roster_B, or Roster_C, but the person was ultimately removed because they did not meet the criteria of staying at least two months OR having no other place to stay.

Table 25. Characteristics of Households that Deleted Persons from the Roster (Across Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	16.4(1.8)	15.2(2.0)	-1.2(2.9)	0.94
	2	35.3(2.3)	34.1(2.6)	-1.2(3.8)	0.94
	3	30.1(2.5)	29.8(3.1)	-0.3(4.0)	0.94
	4	9.1(1.5)	12.9(1.9)	3.9(2.5)	0.59
	5 or more	9.2(1.4)	8.1(1.6)	-1.1(2.0)	0.94
Household Tenure	Own	81.4(1.6)	84.5(2.1)	3.1(2.3)	0.18
	Rent	18.6(1.6)	15.5(2.1)	-3.1(2.3)	0.18
Building Type	Mobile Home	1.9(0.7)	3.6(1.3)	1.7(1.4)	0.87
	One-Family House	85.3(1.8)	85.2(2.4)	-0.2(2.9)	0.95
	Small Apt. Building	7.4(1.3)	8.6(1.8)	1.2(2.1)	0.95
	Large Apt. Building	5.2(1.3)	2.5(0.7)	-2.6(1.5)	0.37
	Other Type Residence	0.2(0.2)	<0.1(<0.1)	-0.2(0.2)	0.95
Household Type	Complex	12.5(1.7)	14.1(1.9)	1.6(2.4)	0.52
	Not Complex	87.5(1.7)	85.9(1.9)	-1.6(2.4)	0.52
English-Speaking Ability	Well	88.3(3.6)	82.1(5.0)	-6.3(6.2)	0.31
	Not Well	11.7(3.6)	17.9(5.0)	6.3(6.2)	0.31

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 26. Characteristics of Households that Deleted Persons from the Roster (Internet Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	12.9(1.6)	11.0(2.1)	-2.0(2.9)	1.00
	2	36.0(2.4)	34.5(2.9)	-1.5(4.2)	1.00
	3	32.6(2.9)	32.6(3.2)	<0.1(4.2)	1.00
	4	9.1(1.7)	14.5(2.3)	5.3(2.9)	0.34
	5 or more	9.3(1.7)	7.5(1.7)	-1.9(2.4)	1.00
Household Tenure	Own	84.5(1.8)	88.4(2.1)	4.0(2.2)	0.07*
	Rent	15.5(1.8)	11.6(2.1)	-4.0(2.2)	0.07*
Building Type	Mobile Home	1.0(0.4)	1.3(0.6)	0.3(0.6)	0.89
	One-Family House	89.1(1.7)	89.5(2.1)	0.4(2.6)	0.89
	Small Apt. Building	5.5(1.0)	6.5(1.8)	1.0(2.0)	0.89
	Large Apt. Building	4.1(1.2)	2.6(0.8)	-1.5(1.5)	0.89
	Other Type Residence	0.2(0.2)	<0.1(<0.1)	-0.2(0.2)	0.89
Household Type	Complex	12.0(1.9)	13.7(2.2)	1.7(2.7)	0.53
	Not Complex	88.0(1.9)	86.3(2.2)	-1.7(2.7)	0.53
English-Speaking Ability	Well	94.0(2.1)	82.8(7.1)	-11.2(7.4)	0.13
	Not Well	6.0(2.1)	17.2(7.1)	11.2(7.4)	0.13

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 27. Characteristics of Households that Deleted Persons from the Roster (CAPI Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Household Size	1	33.8(5.8)	34.3(8.3)	0.5(10.0)	0.98
	2	31.9(6.2)	32.1(8.0)	0.2(10.3)	0.98
	3	17.2(5.6)	16.9(7.0)	-0.3(9.0)	0.98
	4	8.9(3.3)	6.1(1.8)	-2.8(4.0)	0.98
	5 or more	8.2(3.7)	10.6(5.4)	2.4(0.5)	0.98
Household Tenure	Own	66.0(6.1)	66.6(6.1)	0.6(9.4)	0.95
	Rent	34.0(6.1)	33.4(6.1)	-0.6(9.4)	0.95
Building Type	Mobile Home	6.4(3.6)	14.1(6.2)	7.7(7.5)	0.91
	One-Family House	66.5(6.2)	65.4(8.6)	-1.1(10.9)	0.92
	Small Apt. Building	16.7(5.6)	18.4(4.6)	1.7(7.1)	0.92
	Large Apt. Building	10.3(4.2)	2.1(1.2)	-8.2(4.2)	0.19
	Other Type Residence	------	------	------	------
Household Type	Complex	14.8(5.0)	15.6(5.4)		0.92
	Not Complex	85.2(5.0)	84.4(5.4)		0.92
English-Speaking Ability	Well	71.5(11.4)	80.6(6.4)	9.0(12.8)	0.48
	Not Well	28.5(11.4)	19.4(6.4)	-9.0(12.8)	0.48

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance. The (----) symbol indicates that the statistic could not be computed because the categories in both treatments have a frequency of zero.

RQ23. *Is there a difference in the characteristics of the people that are added to the roster between the Control and Roster Test versions? (Analyze overall and by mode: internet and CAPI).* Characteristics include relationship to reference person, age, race, Hispanic origin, sex, educational attainment, and English-speaking ability.

The results of this household-level analysis are presented across mode (internet and CAPI combined) in Table 28 and by mode in Tables 29 (internet mode) and 30 (CAPI mode). The p-values have been adjusted within person characteristic to control the type I familywise error rate.

Significant results are found across mode and in the internet mode in these person characteristics: age, Hispanic origin, educational attainment, and relationship type to the reference person (tenuous or not tenuous). A higher percentage of added people were under 18 and a lower percentage of added people were over 65 in the Test version. Given the second coverage probe (Roster_B) explicitly asks about children it makes sense that more children are added on this screen in the Test version than the Control. It is also possible that such a higher percentage of added people had less than a high school education because of the higher number of these people who were under 18.

While the Control version had a significantly higher percentage of tenuously connected people added after the first screen, it is possible that these people were already included on the first roster question in the Test instrument, since that version prompts them to “include people not related to you.” As a result, fewer people need to be added in the Test instrument because the first screen is more complete.

Table 28. Characteristics of Persons Added to the Roster (Across Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Age	<18	36.2(4.3)	57.2(3.9)	21.0(6.1)	<0.01*
	18-29	18.0(3.0)	16.3(2.8)	-1.7(4.5)	0.70
	30-49	21.9(2.7)	14.7(2.3)	-7.1(3.6)	0.15
	50-64	11.0(2.4)	8.0(2.4)	-3.0(3.3)	0.70
	65 or older	12.9(2.5)	3.8(1.3)	-9.1(2.9)	<0.01*
Race	White alone	61.6(4.6)	53.2(4.6)	-8.4(6.6)	0.54
	Black alone	16.0(3.0)	13.7(2.3)	-2.4(3.9)	0.54
	Some other race alone	15.3(3.0)	29.2(5.1)	13.9(5.9)	0.07
	Two or more races	7.0(3.1)	3.9(1.0)	-0.4(4.9)	0.54
Sex	Male	51.6(3.1)	51.2(3.4)	-0.4(4.9)	0.93
	Female	48.4(3.1)	48.8(3.4)	0.4(4.9)	0.93
Hispanic Origin	Yes	16.1(2.5)	32.8(5.1)	16.7(5.4)	<0.01*
	No	83.9(2.5)	67.2(5.1)	-16.7(5.4)	<0.01*
English-Speaking Ability	Well	63.2(10.8)	77.4(8.2)	14.3(14.4)	0.32
	Not Well	36.8(10.8)	22.6(8.2)	-14.3(14.4)	0.32
Educational Attainment	High school or less	63.3(4.3)	80.9(4.2)	17.6(5.9)	<0.01*
	Some college or Assoc.	18.4(3.7)	9.2(2.9)	-9.3(4.9)	0.11
	Bachelor's or higher	18.3(3.2)	9.9(3.6)	-8.3(5.1)	0.11
Relationship to Reference Person	Tenuous	52.8(4.4)	40.2(4.4)	-12.6(6.6)	0.05*
	Not Tenuous	47.2(4.4)	59.8(4.4)	12.6(6.6)	0.05*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 29. Characteristics of Persons Added to the Roster (Internet Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Age	<18	31.6(3.9)	58.3(4.3)	26.7(6.2)	<0.01*
	18-29	16.4(3.1)	15.2(3.1)	-1.2(4.9)	0.81
	30-49	24.6(2.9)	13.9(2.4)	-10.6(4.0)	0.02*
	50-64	12.5(2.8)	8.5(2.7)	-4.0(3.8)	0.59
	65 or older	14.9(2.8)	4.0(1.5)	-10.9(3.2)	<0.01*
Race	White alone	68.3(4.0)	54.0(5.3)	-14.3(6.8)	0.12
	Black alone	12.0(2.1)	11.4(2.4)	-0.6(3.5)	0.87
	Some other race alone	16.4(3.5)	30.1(5.7)	13.6(6.6)	0.12
	Two or more races	3.3(1.4)	4.5(1.3)	1.3(1.7)	0.87
Sex	Male	53.3(3.3)	54.1(3.3)	0.8(5.0)	0.88
	Female	46.7(3.3)	45.9(3.3)	-0.8(5.0)	0.88
Hispanic Origin	Yes	12.6(2.0)	32.5(6.1)	19.8(6.0)	<0.01*
	No	87.3(2.0)	67.5(6.1)	-19.8(6.0)	<0.01*
English-Speaking Ability	Well	58.3(12.9)	80.1(10.6)	21.7(17.8)	0.22
	Not Well	41.7(12.9)	19.9(10.6)	-21.8(17.8)	0.22
Educational Attainment	High school or less	58.2(4.7)	79.6(5.1)	21.4(6.8)	<0.01*
	Some college or Assoc.	19.3(4.1)	9.4(3.5)	-9.9(5.7)	0.08*
	Bachelor's or higher	22.6(3.9)	11.1(4.3)	-11.5(6.0)	0.08*
Relationship to Reference Person	Tenuous	52.6(4.6)	38.2(4.6)	-14.3(7.1)	0.04*
	Not Tenuous	47.4(4.6)	61.8(4.6)	14.3(7.1)	0.04*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 30. Characteristics of Persons Added to the Roster (CAPI Mode)

Characteristic	Categories	Control	Test	Difference	Adj. P-value
Age	<18	58.7(11.8)	50.2(11.3)	-8.5(17.9)	0.83
	18-29	25.6(8.8)	22.8(8.5)	-2.8(13.1)	0.83
	30-49	8.8(4.0)	20.0(8.0)	11.3(8.9)	0.83
	50-64	3.8(2.0)	4.9(2.5)	1.1(2.9)	0.83
	65 or older	3.1(1.8)	2.0(1.6)	-1.1(2.4)	0.83
Race	White alone	27.6(12.4)	48.6(13.0)	21.0(18.0)	0.49
	Black alone	36.5(12.8)	27.6(10.0)	-8.9(16.9)	0.60
	Some other race alone	9.4(4.4)	23.8(9.2)	14.4(10.0)	0.44
	Two or more races	26.4(15.0)	0.1(<0.1)	26.4(15.0)	0.31
Sex	Male	42.5(10.3)	33.6(9.7)	-8.9(13.3)	0.50
	Female	57.5(10.3)	66.4(9.7)	8.9(13.3)	0.50
Hispanic Origin	Yes	34.0(7.9)	35.0(9.4)	1.0(12.5)	0.94
	No	66.0(7.9)	65.0(9.4)	-1.0(12.5)	0.94
English-Speaking Ability	Well	81.7(9.5)	69.5(14.4)	-12.3(16.6)	0.46
	Not Well	18.2(9.5)	30.5(14.4)	12.3(16.6)	0.46
Educational Attainment	High school or less	84.0(8.4)	87.4(5.1)	3.3(9.5)	0.73
	Some college or Assoc.	15.1(8.3)	8.2(4.0)	-6.9(9.0)	0.73
	Bachelor's or higher	0.8(0.9)	4.4(2.9)	3.6(3.1)	0.72
Relationship to Reference Person	Tenuous	54.3(11.0)	52.4(12.0)	-1.9(15.3)	0.90
	Not Tenuous	45.7(11.0)	47.6(12.0)	1.9(15.3)	0.90

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.4 Response Reliability Results for Household Roster

This section of the report presents metrics that provide an indication of how well the roster obtained in the original interview and the roster obtained in independent CFU reinterview match. The analyses only include households that completed both an original interview and a CFU reinterview, and where the last name and complete date of birth (month, day, year) of all persons rostered were provided. Households that did not meet these criteria were excluded from these analyses.

RQ24. *When we match the Content Test roster to the CFU independent roster by name and date of birth for each person within the household, is the proportion of households with a mismatch between the Content Roster Test and CFU different than between the Control and Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

Households where the rostered persons in the original interview and the CFU reinterview matched one-to-one by last name (exact spelling) and date of birth (exact month, day, and year) were classified as a household roster match. Otherwise, they were classified as a household roster mismatch. The household roster mismatch results are presented in Table 31.

The household roster mismatch percentages for the Control and Test versions do not differ significantly across mode or in any of the three modes examined. The relatively high household roster mismatch percentages are mostly a result of minor inconsistencies in the spelling of the last name between the two interviews. Spelling adjudication was not part of this analysis.

Table 31. Households with Roster Mismatch

Mode	Control	Test	Difference	P-value
Across Mode	31.3(1.1)	32.1(1.0)	0.7(1.5)	0.62
Internet	30.3(1.1)	31.6(1.0)	1.3(1.5)	0.36
Paper	27.8(1.9)	29.3(2.0)	1.6(2.8)	0.58
CAPI	38.2(2.8)	35.8(2.5)	-2.4(3.8)	0.53

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*)indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ25. *When we match at the person level between the Content Test roster and the CFU independent roster, what is the distribution of age groups (0-4, 5-17, 18-24, 25-64, 65 plus) for people who are not matched for the Control and Roster Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

Table 32 presents the results of a Rao-Scott Chi-Square comparison of the age distribution (0-4, 5-17, 18-24, 25-64, 65 plus) for the Control and Test versions for people who are not matched by last name and date of birth. The age distributions do not differ significantly between the Control and Test versions across mode or in any of the three modes examined. The roster mismatch rate in the Test version is comparable to the Control.

Table 32. Person Mismatch Age: Distributions Compared

Mode	Rao-Scott Chi-Square P-value
Overall	$p = 0.27$
Internet	$p = 0.30$
Paper	$p = 0.68$
CAPI	$p = 0.78$

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance.

RQ26. *When we match at the person level between the Content Test roster and the CFU independent roster, what is the distribution of relationship type relative to Person 1 for people who are not matched for the Control and Roster Test versions? (Analyze overall and by mode of original interview: paper, CAPI, and internet).*

The results in Table 33 of a Rao-Scott Chi-Square comparison of the distributions of the relationship type relative to Person 1 for the Control and Test versions for people who were not matched (by last name and date of birth) indicate that the distributions do not differ significantly across mode or in any of the three modes examined.

Table 33. Person Mismatch Relationship Type to Reference Person: Distributions Compared

Mode	Rao-Scott Chi-Square P-value
Overall	$p = 0.41$
Internet	$p = 0.50$
Paper	$p = 0.30$
CAPI	$p = 0.88$

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance.

5.5 Other Metrics for Household Roster

This section of the report presents select results of a separate study on respondent burden for the 2022 ACS Content Test (Virgile et al., 2023). The results, which provide an indication of respondent burden for household roster, are unweighted. See the Virgile et al. (2023) for details.

RQ27. *How often do respondents access the help screen when answering the household roster question?*

Table 34 presents the access rates for the help screens for Household Roster. The help screen access rate for the Test version is significantly lower than that of the Control version, which may indicate less confusion from respondents.

Table 34. Help Screen Access Rates for Household Roster

Control	Test	Difference	P-value
3.8(0.2)	3.1(0.2)	0.7(0.2)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ28. *What are the breakoff rates for the household roster question?*

Table 35 presents the survey breakoff rates for Household Roster. There is no significant difference between the breakoff rates for the Control and Test versions. The relatively low breakoff rates for both treatments indicate that only a small percent of respondents broke away from the survey upon encountering the Household Roster question.

Table 35. Survey Breakoff Rates for Household Roster

Control	Test	Difference	P-value
1.1(0.1)	1.0(<0.1)	<0.1(<0.1)	0.94

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

6 CONCLUSIONS AND RECOMMENDATIONS

The purpose of this content test was to evaluate if a new set of rostering instructions for the paper questionnaire and revisions to the roster questions for the internet and interviewer-administered instruments would improve the rostering procedure. On the whole, the Test version resulted in similar or better data quality indicators than the Control version.

The six decision criteria that we used to evaluate the Test version of household roster are provided below along with the result.

Decision Criteria 1: *The number of children between the ages of 0-4 on the final roster is significantly higher in the Roster Test version compared to the Control version.*

Research Question 7: There was not a significant difference between the Control and Test version in the number of young children (0-4 years of age) on the final roster. However, with respect to the universe of young children added to the roster (RQ16), a significantly larger number of them were added to the second undercoverage probe (Roster_B) in the Test version, which was modified to explicitly ask about additional children living or staying at the address.

Decision Criteria 2: *The proportion of households with a mismatch between the Content Roster Test and CFU roster is lower in the Roster Test version than the Control version.*

Research Question 24: There was not a significant difference in the match rates for households between the Control and test versions.

Decision Criteria 3: *The percentage of people with tenuous connections to the household is significantly higher in the Roster Test version compared to the Control version.*

Research Question 6: There was not a significant difference in the percentage of people with a tenuous connection between the Control and Test versions. Given that the denominator for this analysis is person level and the universe of tenuously attached people is relatively small, it is not surprising this difference was not statistically significant. The definition of “tenuous” may have been too broad as it included every relationship to person 1, except partners and children. It is important to note that there being more complex households in the Test version than the Control suggests the Test version did help pick up more tenuously attached people at least at a household level.

Decision Criteria 4: *The percentage of complex households for the Roster Test version is significantly higher than the Control version.*

Research Question 5: The percentage of complex households in the Test version was significantly higher than the Control version.

Decision Criteria 5: *The percentage of cases with a count discrepancy for the Roster Test version is lower than the percentage of cases with a count discrepancy for the Control version.*

Research Questions 2, 3, & 4: The Test version had a significantly lower percentage of cases with a count discrepancy than the Control version. Meaning that, in the paper mode, a higher percentage of cases reported a matching number of people and names on their roster in the Test version than the Control.

Decision Criteria 6: *The item missing data rates for the Roster Test version are the same or lower than the item missing data rates for the Control version.*

Research Question 1: The item missing data rate for the population count question in the paper mode, which asks how many people live at the address, was significantly lower in the Test version than the Control version.

The paper questionnaire must be evaluated by several different metrics since it is limited in the number of roster questions it can ask. Because of space limitations, the paper questionnaire collects only the population count (the number of people living or staying at the address) and then collects the name and demographic information for each person in the household.

There are no under or overcoverage follow-up questions in the paper mode, as there are in the internet and CAPI modes. However, this research found evidence of higher data quality in the Test version than the Control for the paper mode. In particular, the item missing rate for the population count in the Test version was significantly lower than the Control by over two percentage points. There was also a lower count discrepancy, which in turn will lead to less followup for count discrepancies in data collection.

Most comparisons of the characteristics of people and households between the Test and Control versions found no significant differences. However, there were a few significant findings, which suggest improved coverage for populations where undercoverage exists. The percentage of complex households was significantly higher in the Test version than the Control version. With respect to differences in characteristics at the person level, there was a significantly higher percentage of men and people who spoke a foreign language and reported speaking English well in the Test version than in the Control. The ACS classifies someone as speaking a foreign language if they report using a language other than English at home. Past research has found that undercoverage is more likely for men than women, as well as for people for whom English is not the only language spoken in the home.

In the internet and CAPI modes there was significantly less churn in the Test version than the Control. A higher percentage of people were added during the first screen, and across all screens fewer people were later deleted before the final roster in the Test version. The first screen added the words “including people who are not related to you” into the question stem in the Test version. This is one possible explanation for the difference. This may also explain

why there was a significantly higher percentage of complex households in the Test version than the Control version.

While the first three decision criteria failed to meet statistical significance, this was potentially a result of small cell sizes leading to low statistical power. The failure to meet these criteria does not indicate poor data quality given the measures are similar with the Control version. At a housing unit level, the Test version did capture a significantly higher percentage of complex households, despite not detecting a difference in the percentage of tenuous connections at the person level. The fifth (count discrepancy) and sixth decision criteria (item missing rates) were met and indicate that the roster treatment had higher data quality, in the paper mode.

When looking at the roster actions in the CAPI and internet modes, a pattern of less confusion and potentially more accurate rosters also appears in that: 1) a higher percentage of people were originally rostered on the first screen in the Test version, 2) a significantly lower percentage of people were deleted from the Test version, 3) a higher percentage of young children (0-4) were added during the second roster screen in the Test version, and 4) a higher percentage of added people were ultimately kept in the Test version than the control.

The recommendation of the Census Bureau is to adopt the Test version of the roster procedure for paper, internet, and CAPI.

7 ACKNOWLEDGMENTS

The 2022 ACS Content Test would not have been possible without the participation and assistance of many individuals from the Census Bureau and other agencies. We sincerely appreciate their contributions:

- Census Bureau staff in the American Community Survey Office, Application Development and Services Division, Center for Behavioral Science Methods, 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 OMB's Interagency Working Group for the ACS and the topical subcommittees formed by the Working Group for each topic tested in the 2022 ACS Content Test.

- Staff in OMB's Statistical and Science Policy Office.

8 REFERENCES

Ashenfelter, K., Quach, V., Holland, T., Nichols, E., & Lakhe, S. (2012). *Report for Round 3 of Usability Testing of the 2011 American Community Survey Online Instrument: Focus on Login and Roster Features*. U.S. Census Bureau. Retrieved June 14, 2018 from https://www.census.gov/srd/papers/pdf/ssm2012-10.pdf

Cherlin, A. J. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and Family*, 72(3), 403-419. Demographic Trends in the United States: A Review of Research in the 2000s.

Clark, S. (2017). *Analysis of the household roster questions on the ACS. U.S*. Census Bureau. Retrieved June 14, 2018 from https://www.census.gov/library/working-papers/2017/acs/2017_Clark_01.html

Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 75 (4), 800-802. Retrieved January 17, 2017, from https://doi.org/10.2307/2336325

Heimel, S. (2016). *Postal tracking research on the May 2015 ACS panel*. U.S. Census Bureau. 2016 American Community Survey Research and Evaluation Report Memorandum Series #ACS16-RER-01.

Horwitz, R., Tancreto, J.G., Zelenak, M.F., Davis, M. (2013). *Use of paradata to assess the quality and functionality of the American Community Survey internet instrument*. U.S. Census Bureau. Retrieved June 14, 2018 from https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_01.pdf

Jensen, Eric B., and Howard R. Hogan (2017). "The coverage of young children in demographic surveys." Statistical Journal of the IAOS 33, no. 2 (2017): 321-333. Retrieved February 8, 2022 from https://content.iospress.com/download/statistical-journal-of-the-iaos/sji170376?id=statistical-journal-of-the-iaos%2Fsji170376

Jensen, E., Schwede, L., Griffin, D., & Konicki, S. (2018). "Investigating the 2010 Undercount of Young Children- Analysis of Complex Households." Washington, D.C.: U.S. Census Bureau. Retrieved February 8, 2022 from https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-2010-undercount-children-complex_households.pdf

Jensen, E. (2019). *Investigating the 2010 undercount of young children: Examining coverage in demographic surveys*. U.S. Census Bureau. Retrieved February 8, 2022 from: https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-2010-undercount-children-examining_coverage_demo.pdf

Keathley, D. (2022). *Specifications for selecting and weighting the 2022 American Community Survey Content Test sample.* U.S. Census Bureau. DSSD 2022 American Community Survey Memorandum Series #ACS22-S-21.

Office of Management and Budget (2006). *Standards and guidelines for statistical surveys*. Retrieved February 24, 2022, from https://www.whitehouse.gov/wpcontent/uploads/2021/04/standards_stat_surveys.pdf

Olmsted-Hawala, E., Berger, M., Kephart K., Otero Class B., & Nichols E. (2023). *Usability and Cognitive Testing of the Rostering Screens for the American Community Survey in English and Spanish*. Retrieved November 7, 2023, from Usability and Cognitive Testing of the Rostering Screens for the American Community Survey in English and Spanish (census.gov)

Risley, M., & Oliver, B. (2022). *High-level sampling and weighting requirements for the 2022 American Community Survey Content Test.* U.S. Census Bureau. DSSD 2022 American Community Survey Memorandum Series #ACS22-MP-01.

RTI International (2022a). *2022 ACS Content Test: Round 1 and Round 2 Cognitive Testing.* Retrieved July 19, 2022, from https://www.census.gov/content/dam/Census/library/working-papers/2022/acs/2022_Wilson_01.pdf

RTI International (2022b). *Cognitive Testing for the 2022 ACS Content Test: Round 3 Briefing* Retrieved August 29, 2022 from https://www.census.gov/library/working-papers/2022/acs/2022_Wilson_02.html

Spiers, S. (2021a). *Requirements for the Content Follow-Up Reinterview Survey in the 2022 American Community Survey Content Test*. U.S. Census Bureau. American Community Survey Memorandum Series #ACS22-MP-02.

Spiers, S. (2021b). Coverage and nonresponse bias in the 2016 American Community Survey Content Follow-Up reinterview. *Proceedings of the 2021 Joint Statistical Meetings, Survey Research Methods Section* (pp. 1664-1677). American Statistical Association.

Spiers, S., Baumgardner, S., & Mojica, A. (2023). *2022 American Community Survey Content Test evaluation report: Unit-level response*. Retrieved November 7, 2023, from 2022 American Community Survey Content Test Evaluation Report: Unit-Level Response (census.gov).

U.S. Census Bureau. (2022a). *U.S. Census Bureau statistical quality standards*. Retrieved September 16, 2022, from https://www2.census.gov/about/policies/quality/quality-standards.pdf

U.S. Census Bureau. (2022b). *American Community Survey and Puerto Rico Community Survey Design and Methodology, Version 3.0.* Retrieved December 15, 2022, from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

Virgile, M., Weng, C. F., Mills, G., & Spiers, S. (2023). *2022 American Community Survey Content Test evaluation report: Respondent burden.* Retrieved November 28, 2023 from, 2022 American Community Survey Content Test Evaluation Report: Respondent Burden (census.gov)

9 APPENDIX

The text below describes the path taken in answering the Household Roster questions in the internet and CAPI modes, for the Control and Test versions. The accompanying figures follow.

- **Resp_Name** (in Figures 4 and 5) is the first roster screen displayed, and it is where the reference person, person 1, enters his or her name, telephone number, and email address.
- **Roster_A** (in Figures 6 and 7) is the next roster screen displayed, and it is where the initial household roster is created. This screen includes the reference person. The respondent is asked to list everyone, including people not related to them, who are living or staying at the address. To add more than five people, the respondent has to click the link, "click here to add more people."
- **Roster_B** (in Figures 8 and 9) is the next roster screen displayed, and it is where the respondent is probed about ADDITIONAL children who might be living at the address (e.g., babies, grandchildren, foster children).
    - **Add_1** (in Figures 10 and 11) screen only appears if a "Yes" response was checked in Roster_B. This screen is where the additional children are entered.
- **Roster_C** (in Figures 12 and 13) is the next roster screen displayed, and it is where the respondent is probed about ADDITIONAL people staying at the address, even for a short time (e.g., roommates and other people or families who have no other place to stay).
    - **Add_2** (in Figures 14 and 15) screen only appears if a "Yes" response was checked in Roster_C. The screen is where the additional people staying at the address are listed.
- **Away_Now** (in Figures 16 and 17) is the next roster screen displayed, and it is where the respondent is probed about whether any of the persons rostered so far live somewhere else now (e.g., a college student or someone in the Armed Forces on deployment).
    - **Remove_One** (in Figures 18 and 19) screen only appears if a "Yes" response was checked in Roster_Away. The respondent is asked to select the names of persons who are living somewhere else now. These persons are removed from the roster.
- **Another_Home** (in Figures 20 and 21) is the next roster screen displayed, and it is where the respondent is probed about whether any of the current persons rostered (minus those removed in Remove_One) are at the address for a short visit or overnight stay. and only included 2 months in the followup question if respondents said "yes" to the question on the Another_Home screen in the internet or interviewer modes.

- **Another_Home_Who** (in Figures 22 and 23) screen only appears if a "Yes" response was checked in Another_Home. The respondent is asked to select the names of persons who are staying at the address for a short visit or overnight stay.
- **More_Than_Two** (in Figures 24-27) screen only appears for persons selected in Another_Home_Who. A separate screen appears for each person selected and references the specific person's name. In Figures 24 and 25, the screen name is More_Than_Two_2/4. The top number, 2 is the total number of persons selected in Another_Home_Who. The bottom number, 4 means that Erin is the fourth person listed in the roster list. The respondent is asked if <name> is staying at the address for more than two months. Since a "Yes" is selected for Erin, Erin is retained on the roster.

  Notice that for Frank (in Figures 26 and 27), a "No" was selected, meaning Frank is not staying at the address for more than two months. This triggers a new screen in the Test version only, the Roster_Stay_# screen.

- **Roster_Stay** (in Figure 29) only appears for persons whom a "No" response was selected in More_Than_Two_#/# (i.e., the person is not staying at the address for more than two months). This screen is only applicable to the Test version.
- **Roster_Check** (in Figures 28 and 30) is the final roster screen, and it is where the FINAL roster is displayed.

Figure 4. Control: Resp_Name

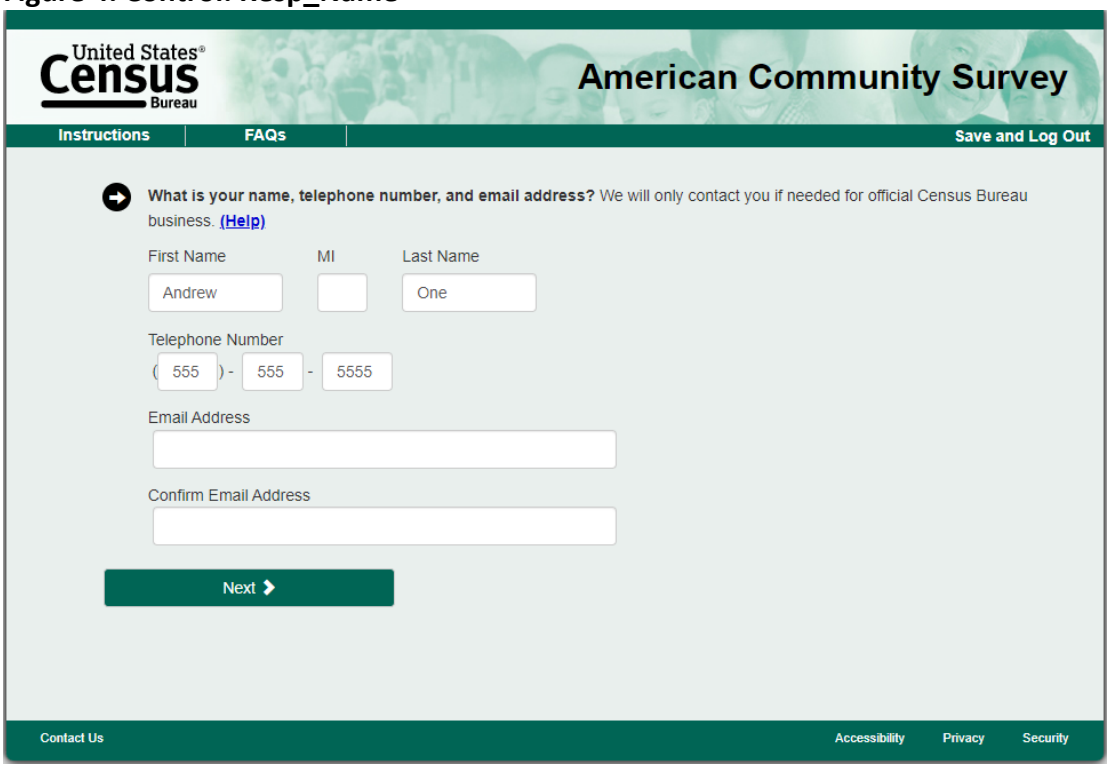


Figure 5. Roster Test: Resp_Name

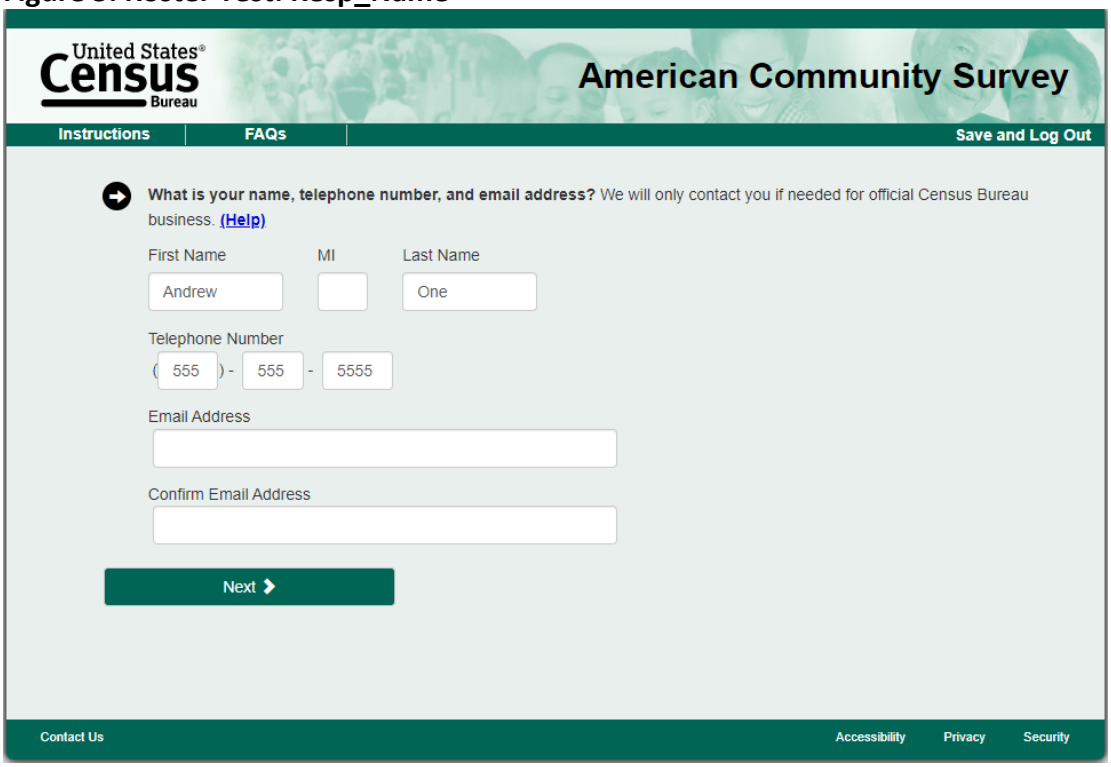


Figure 6. Control: Roster_A

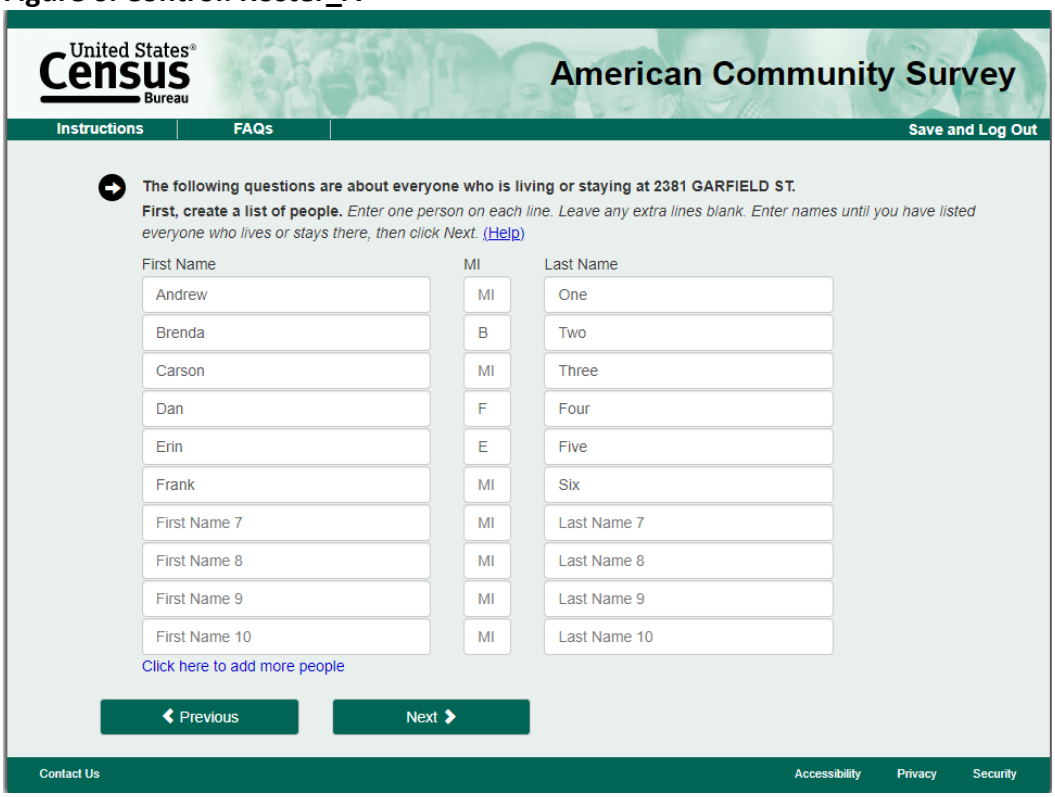


Figure 7. Roster Test: Roster_A

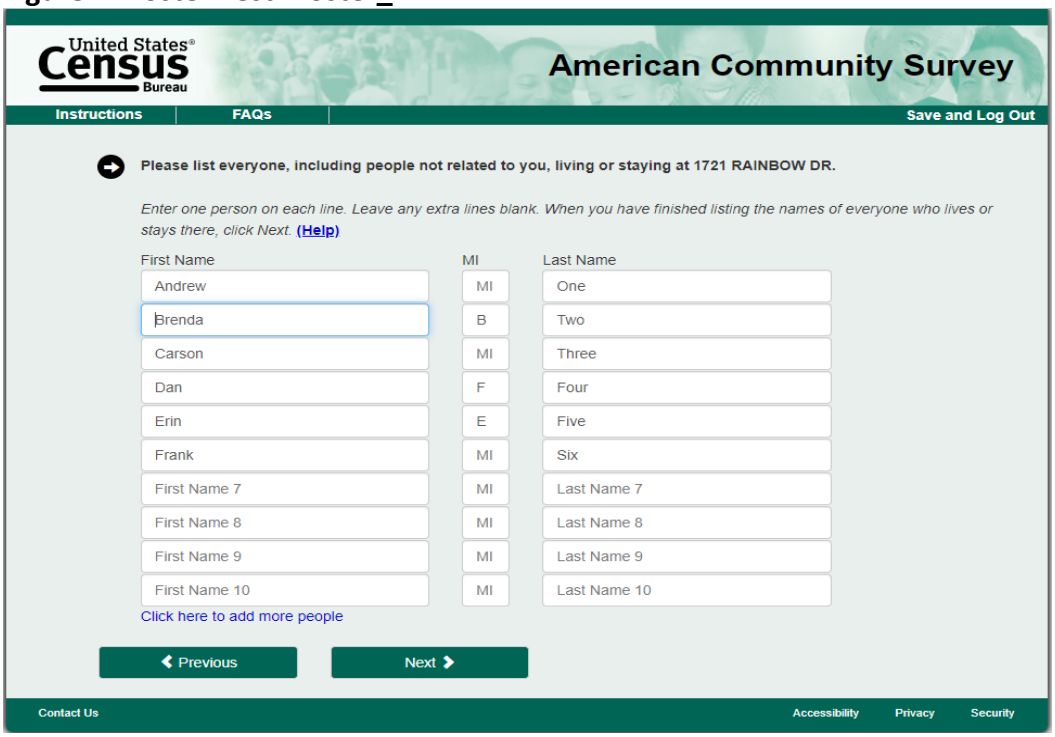


Figure 8. Control: Roster_B

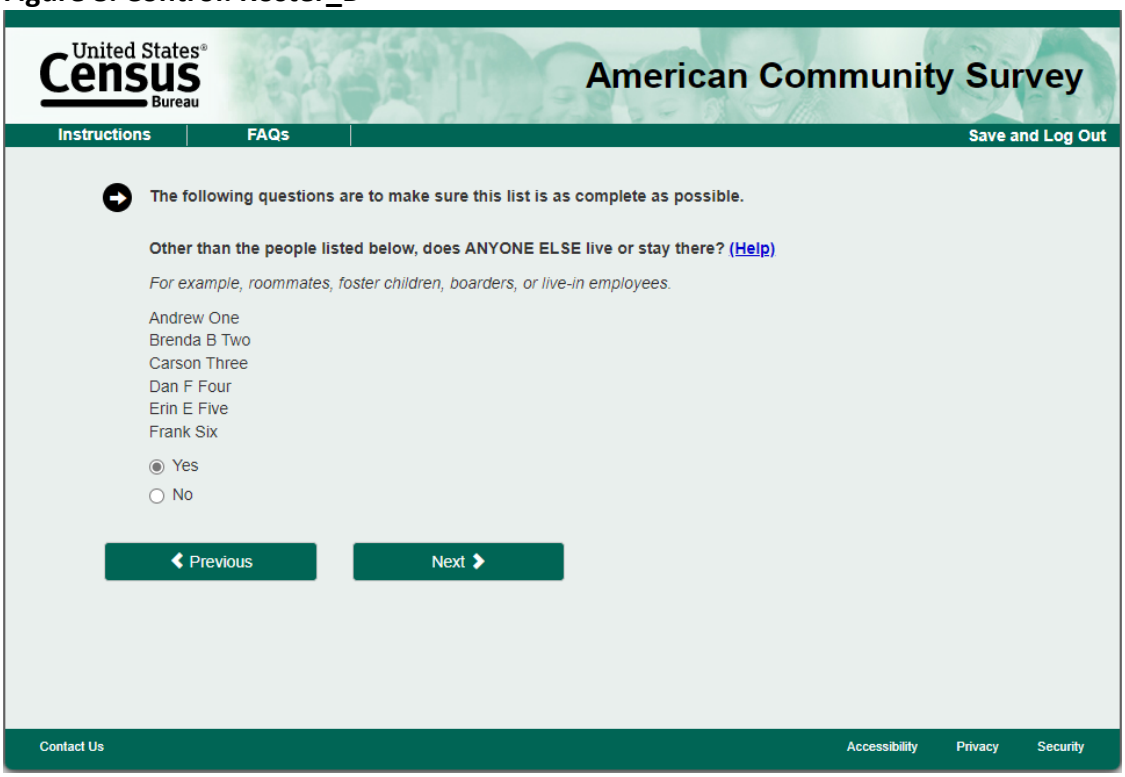


Figure 9. Roster Test: Roster_B

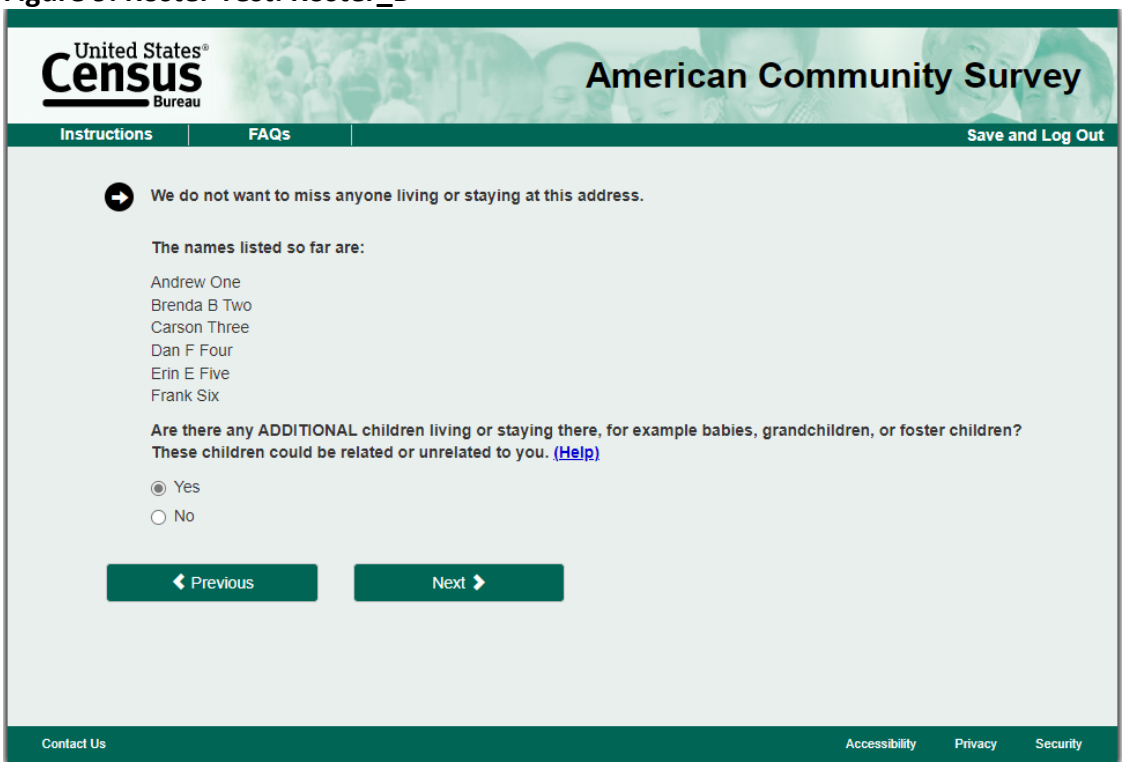


Figure 10. Control: Add_1

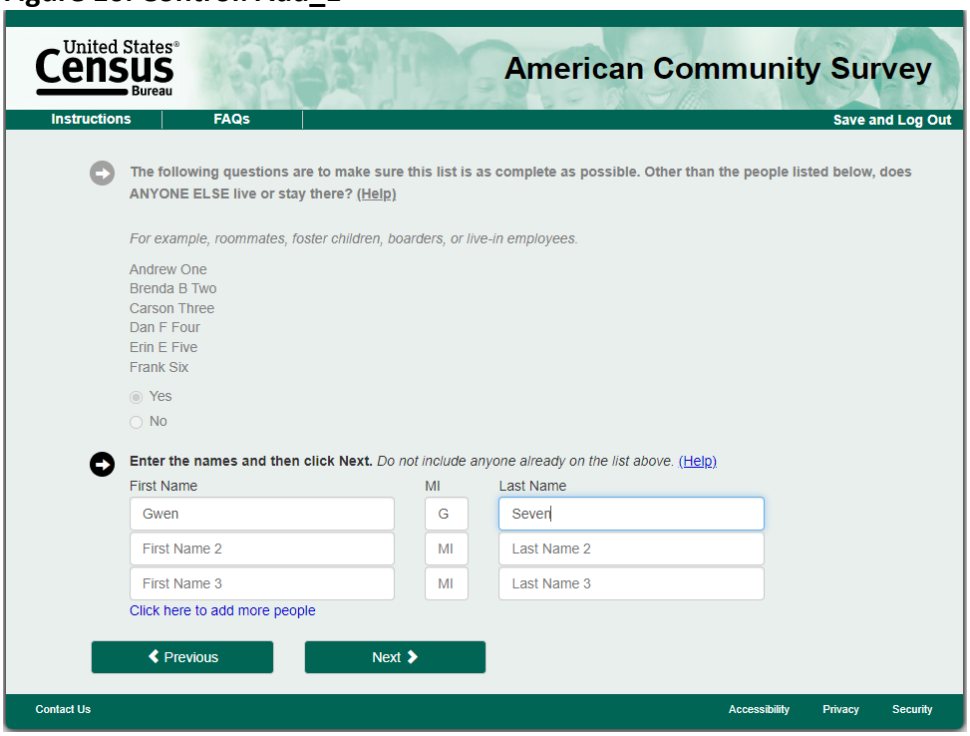


Figure 11. Roster Test: Add_1

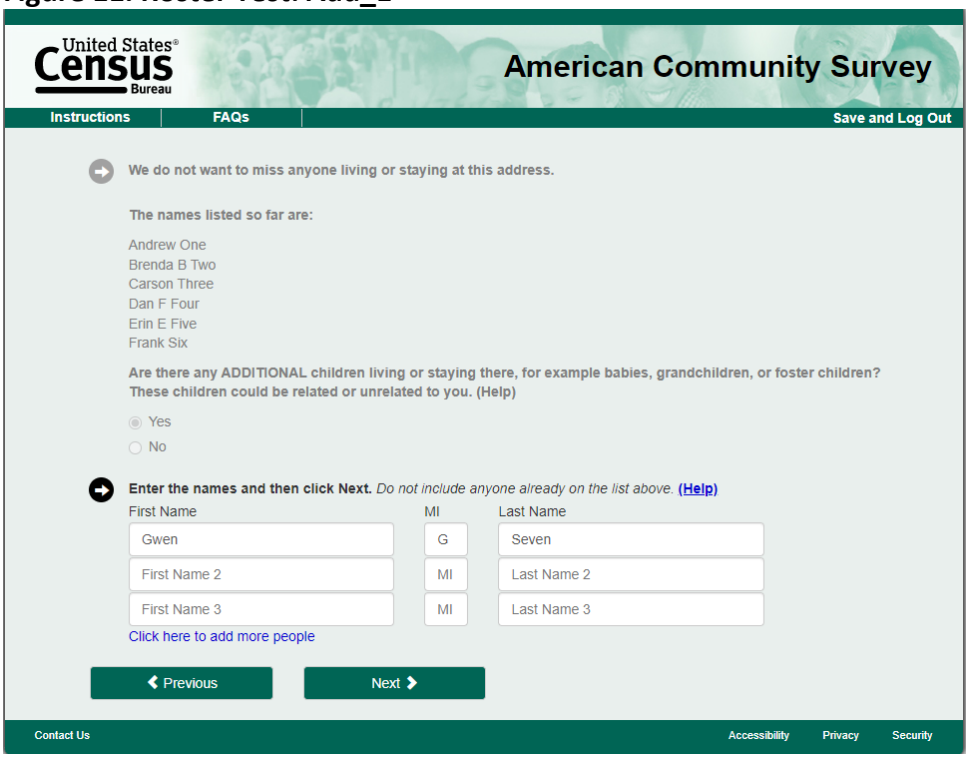


Figure 12. Control: Roster_C

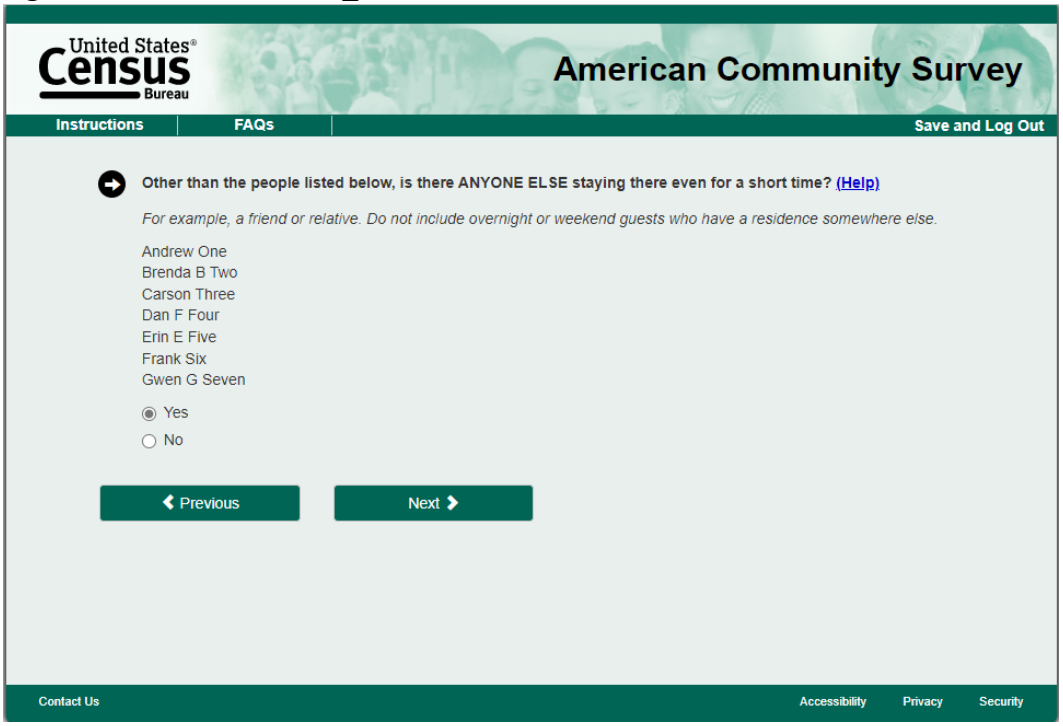


Figure 13. Roster Test: Roster_C

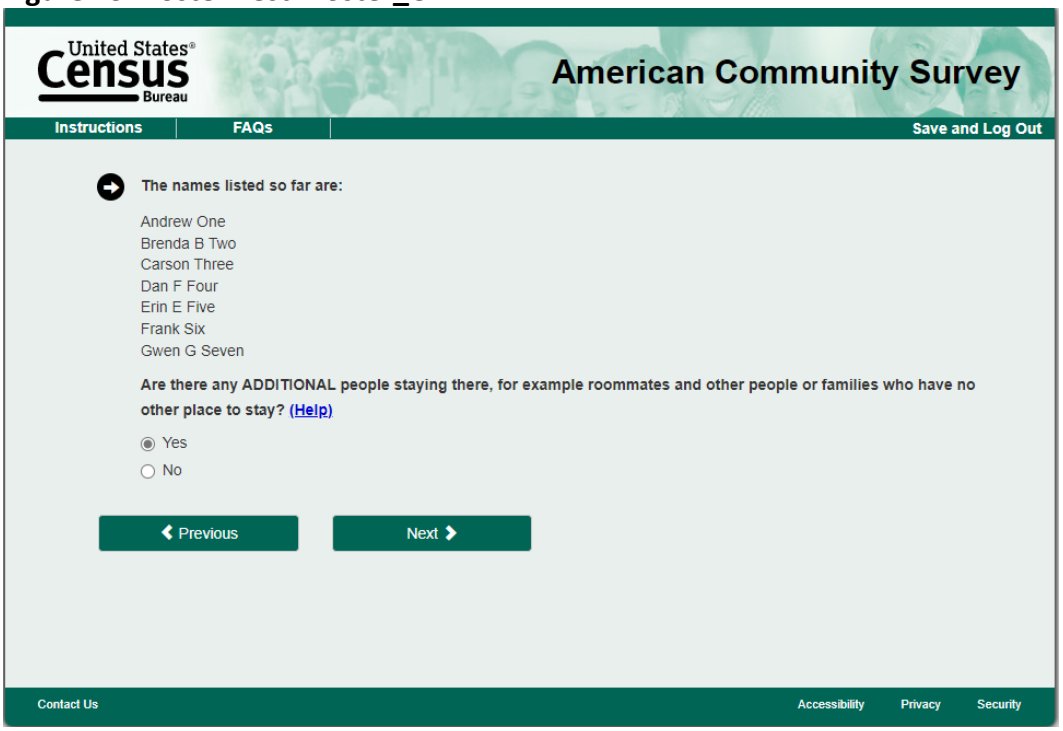


Figure 14. Control: Add_2

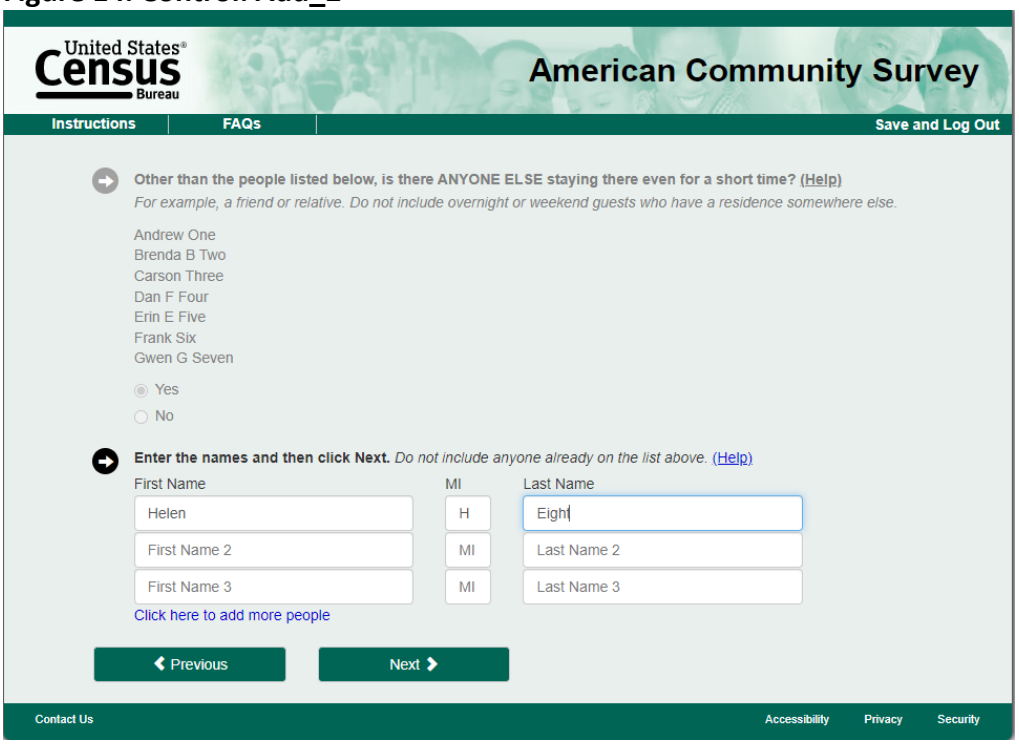


Figure 15. Roster Test: Add_2

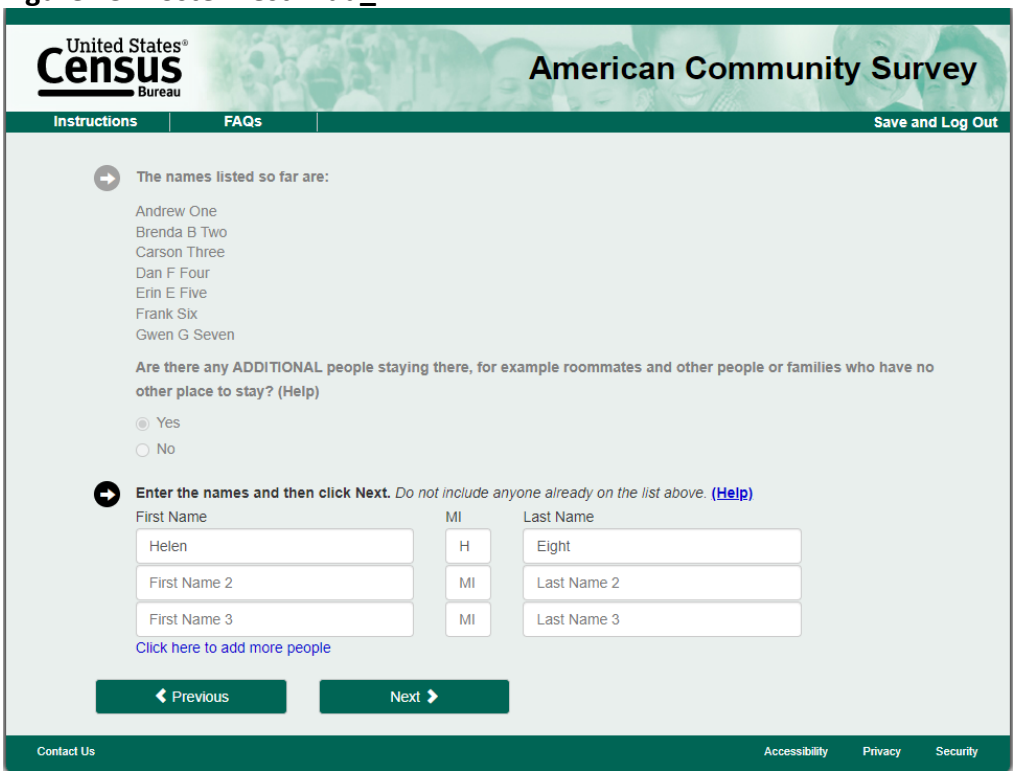


Figure 16. Control: Away_Now

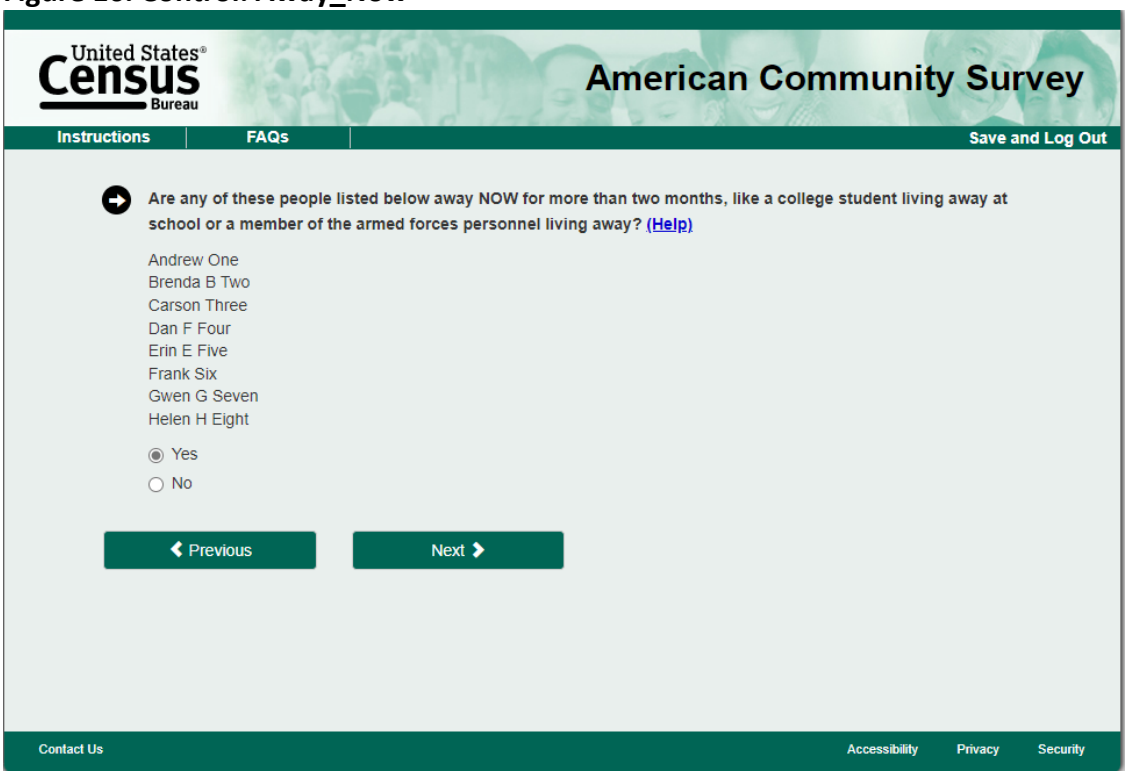


Figure 17. Roster Test: Away_Now

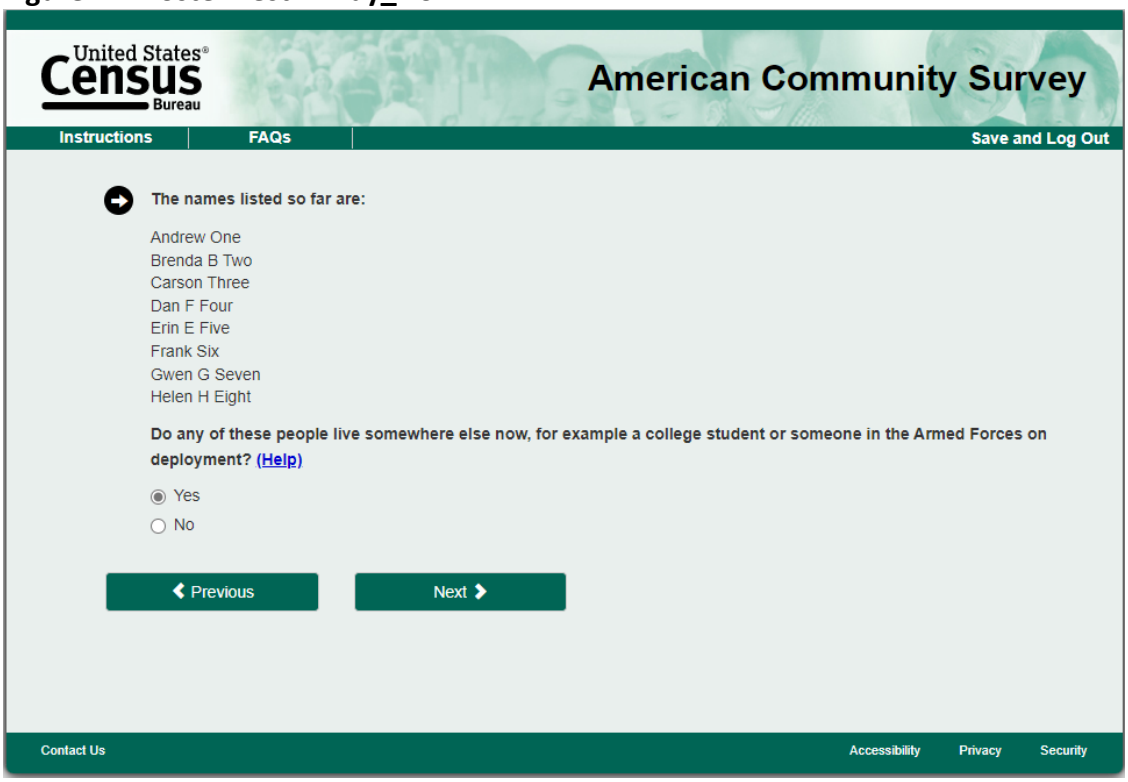


Figure 18. Control: Remove_One

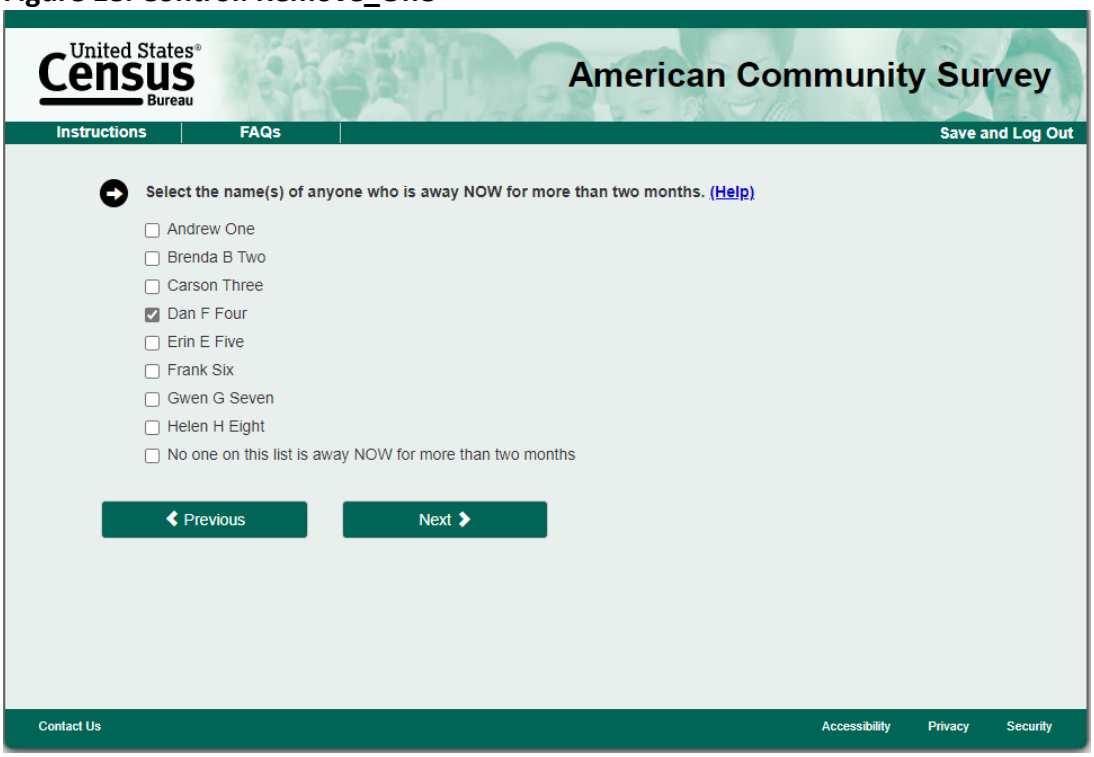


Figure 19. Roster Test: Remove_One

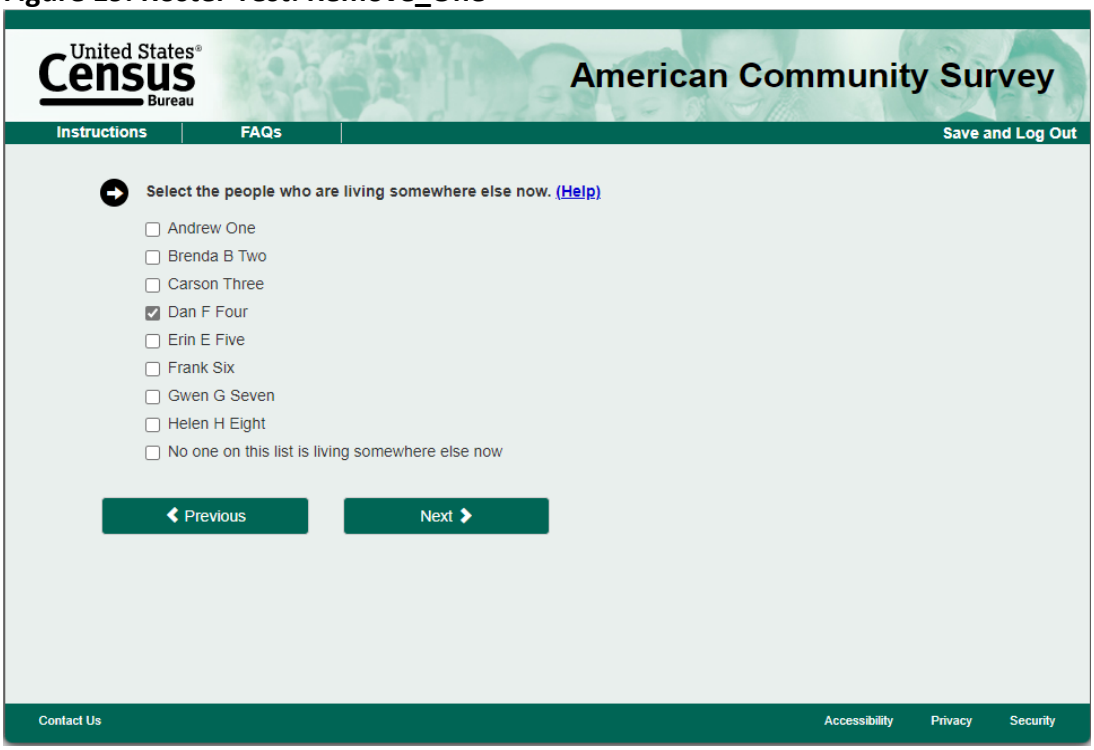


Figure 20. Control: Another_Home

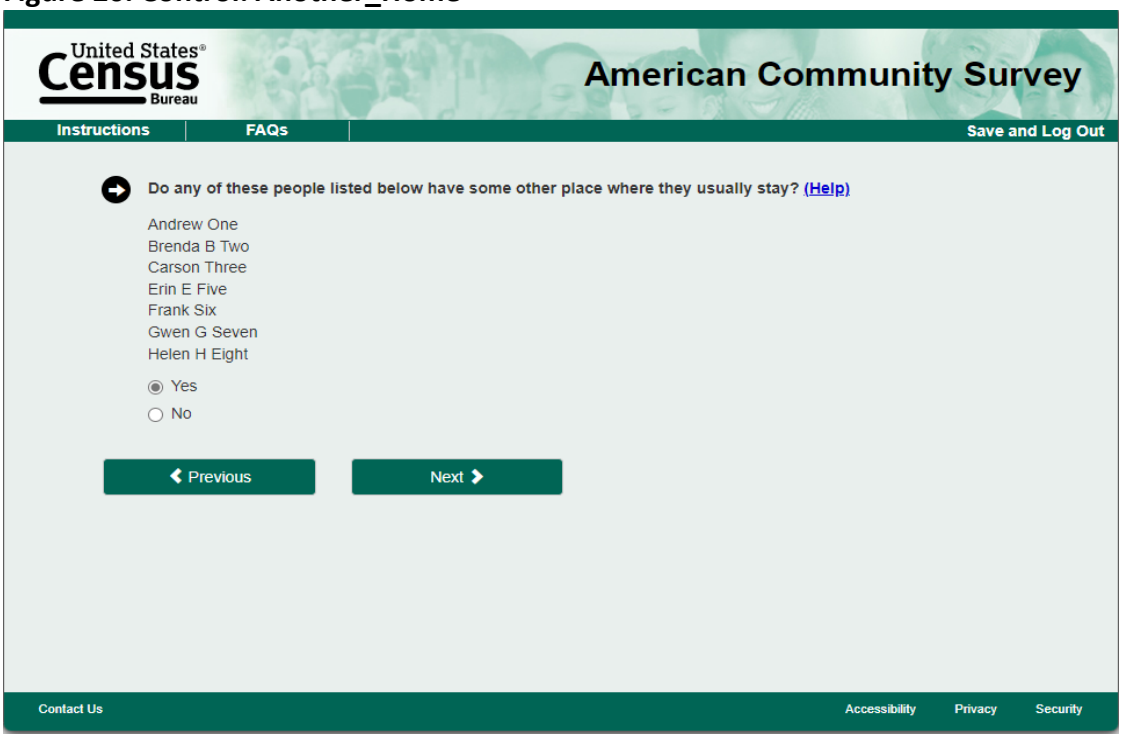


Figure 21. Roster Test: Another_Home

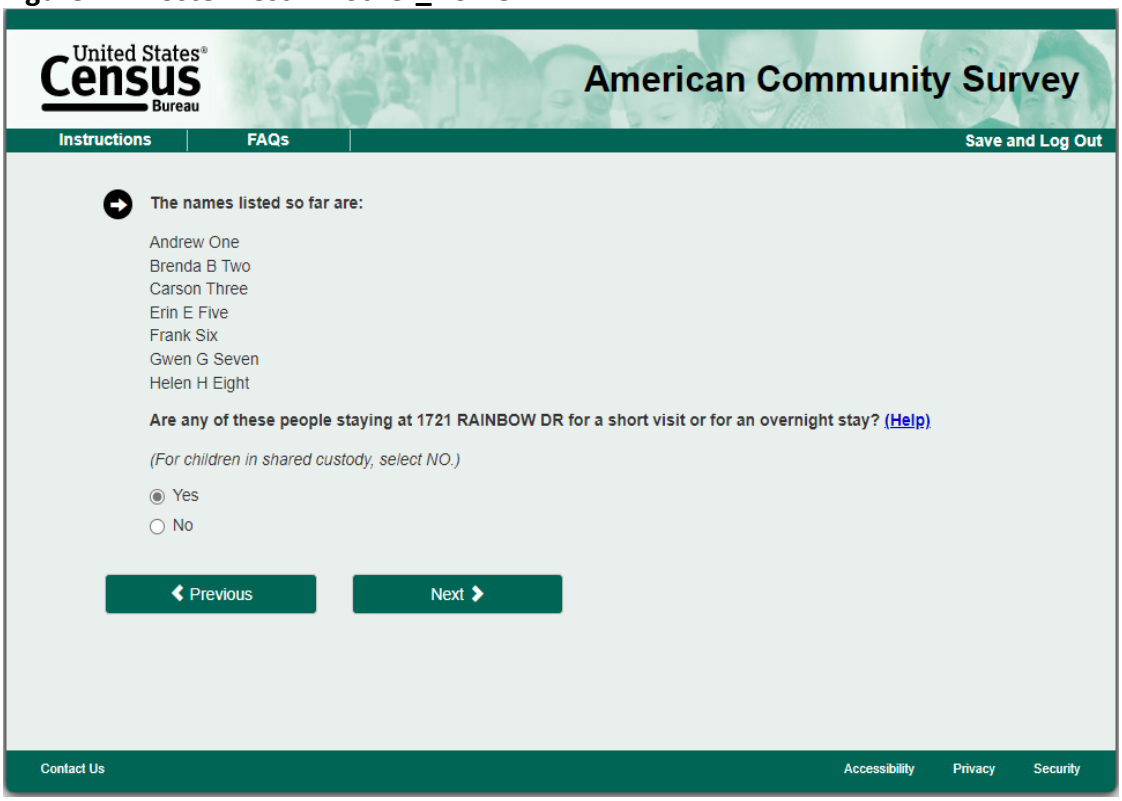


Figure 22. Control: Another_Home_Who

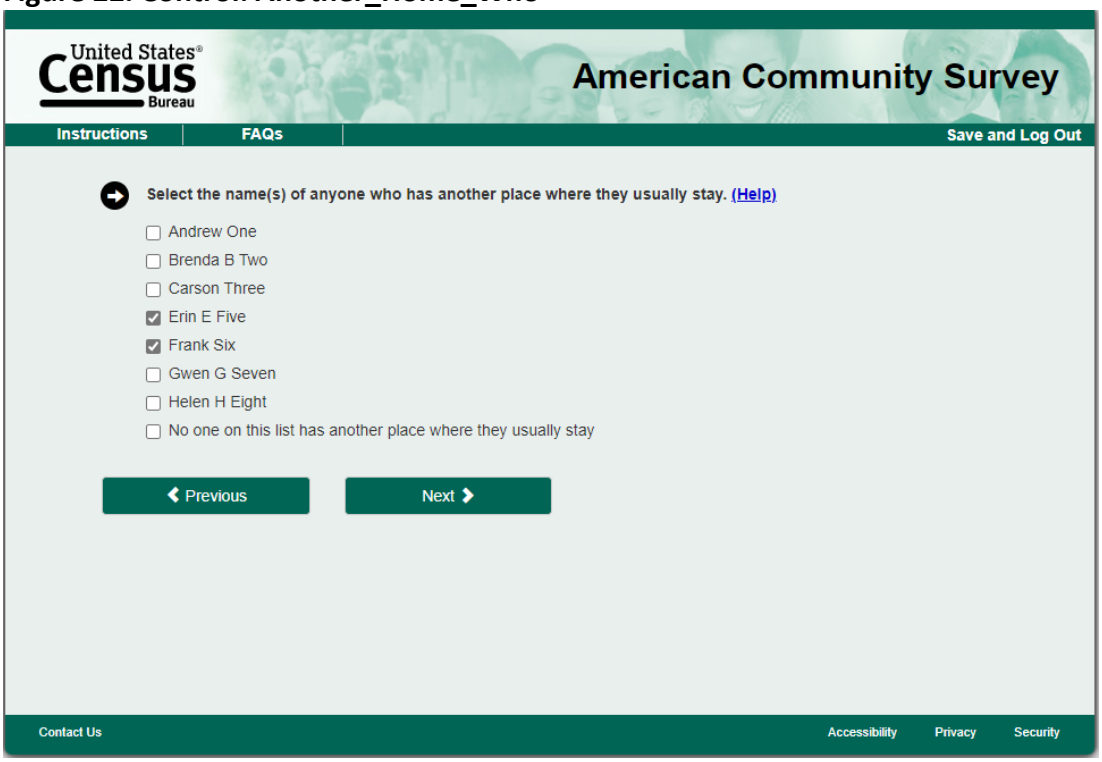


Figure 23. Roster Test: Another_Home_Who

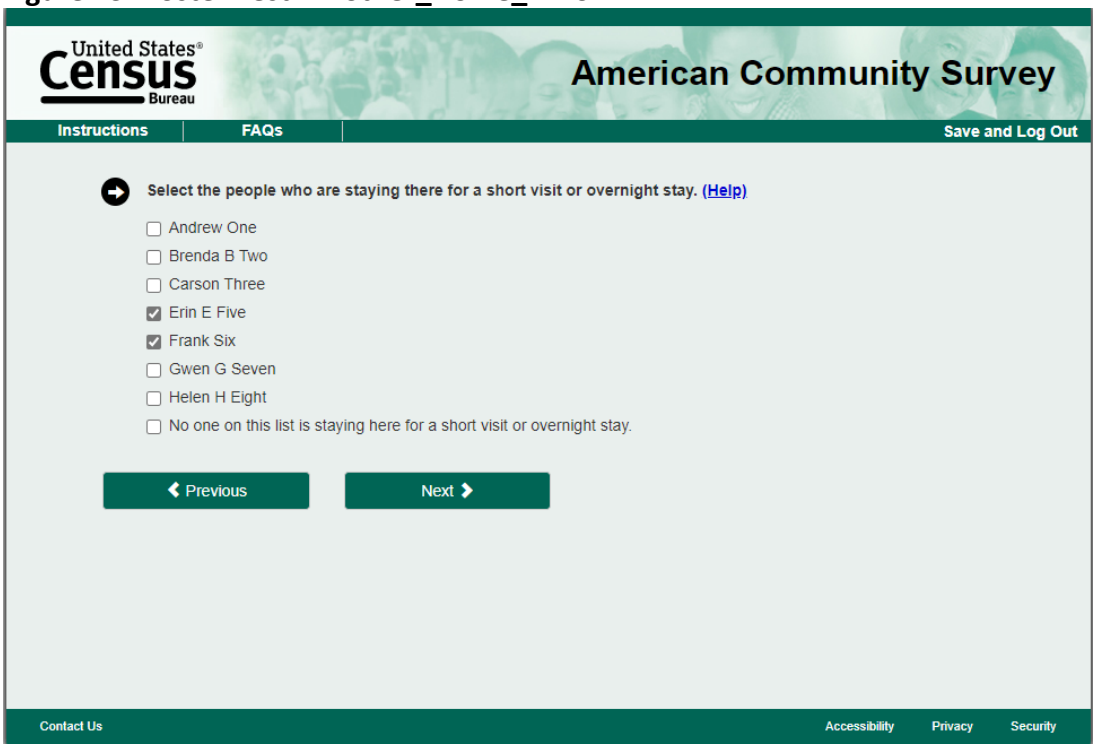


Figure 24. Control: More_Than_Two_2/4

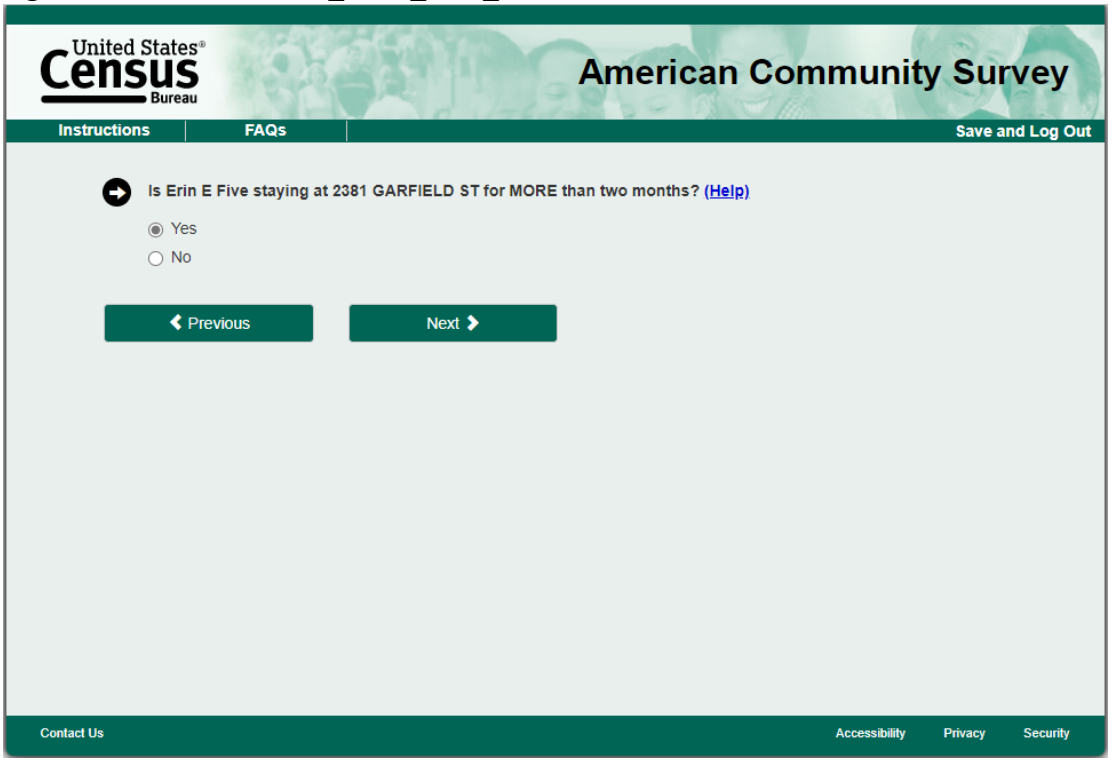


Figure 25. Roster Test: More_Than_Two_2/4

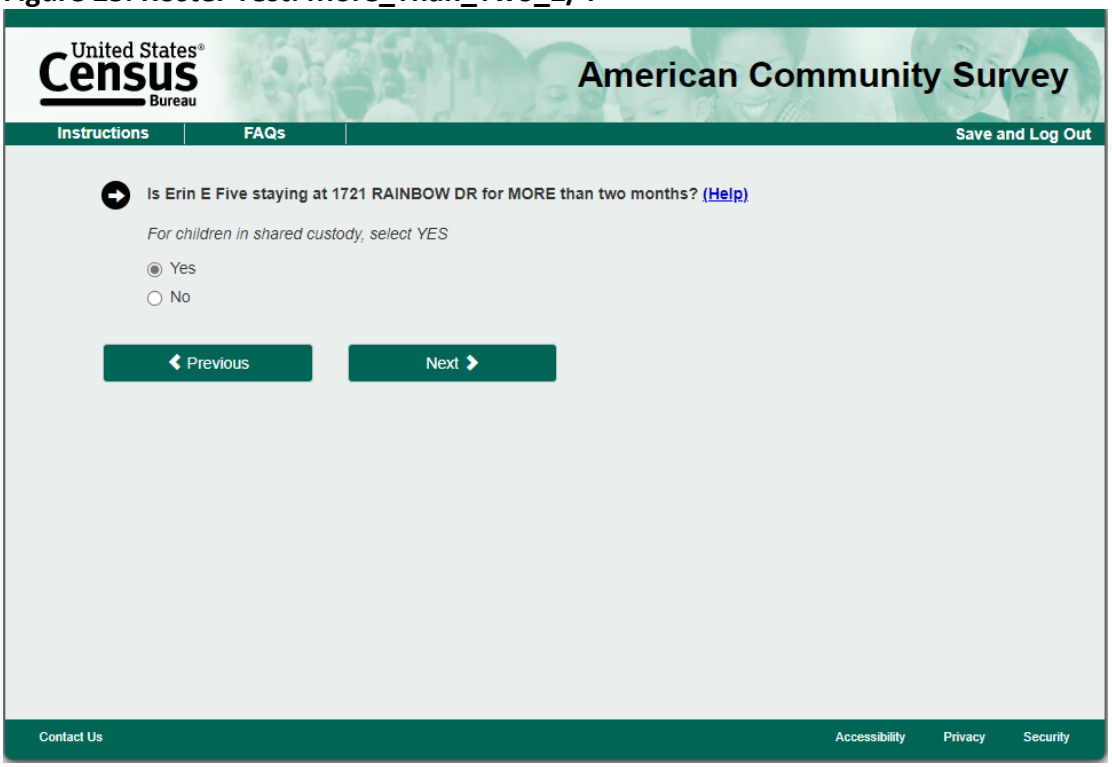


Figure 26. Control: More_Than_Two_2/5

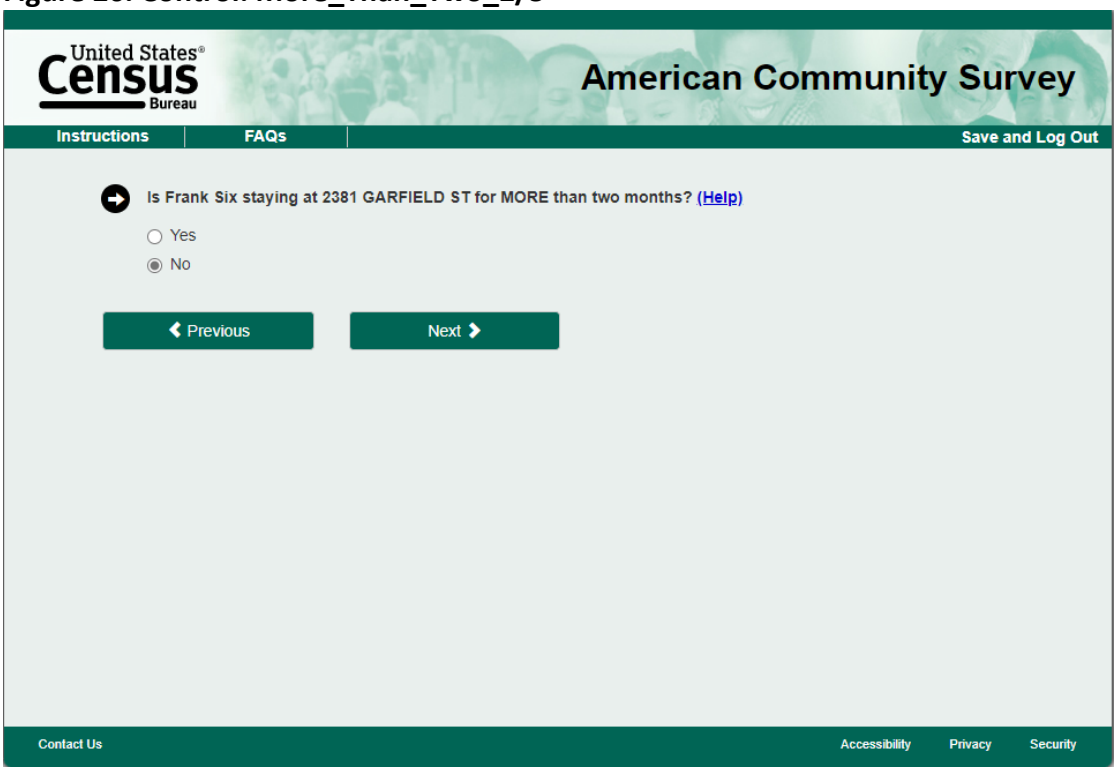


Figure 27. Roster Test: More_Than_Two_2/5

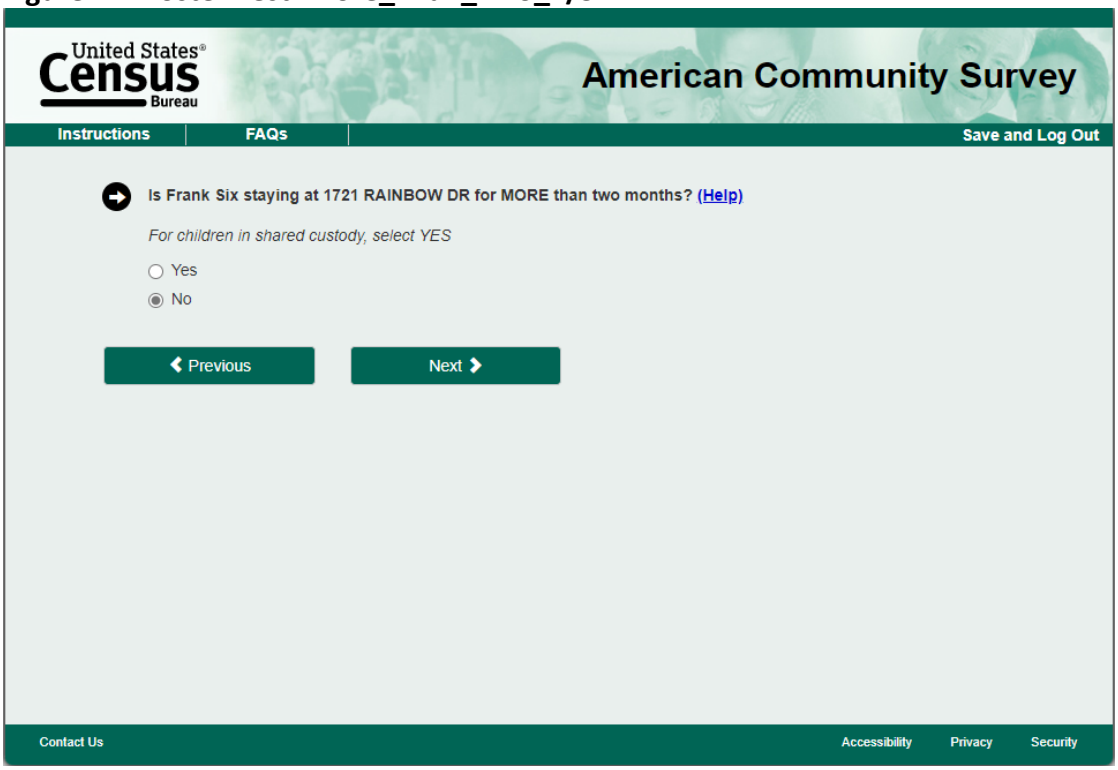


Figure 28. Control: Roster_Check

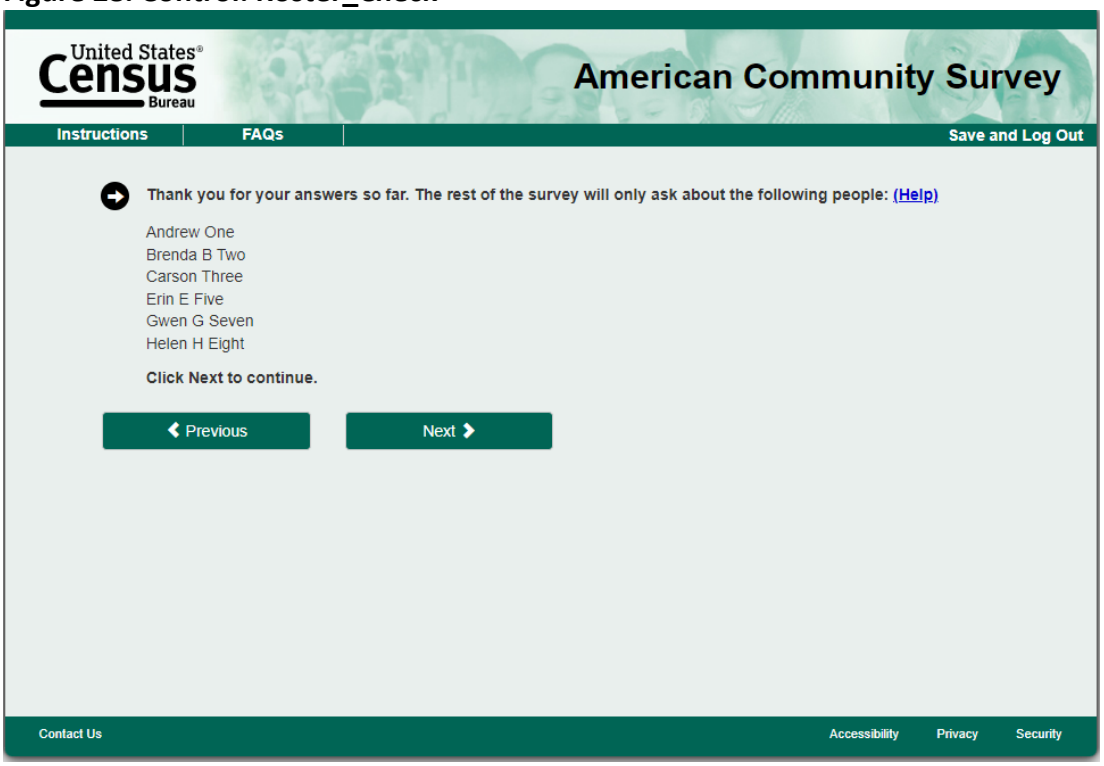


Figure 29. Roster Test: Roster_Stay/5 (new screen in Test version only)

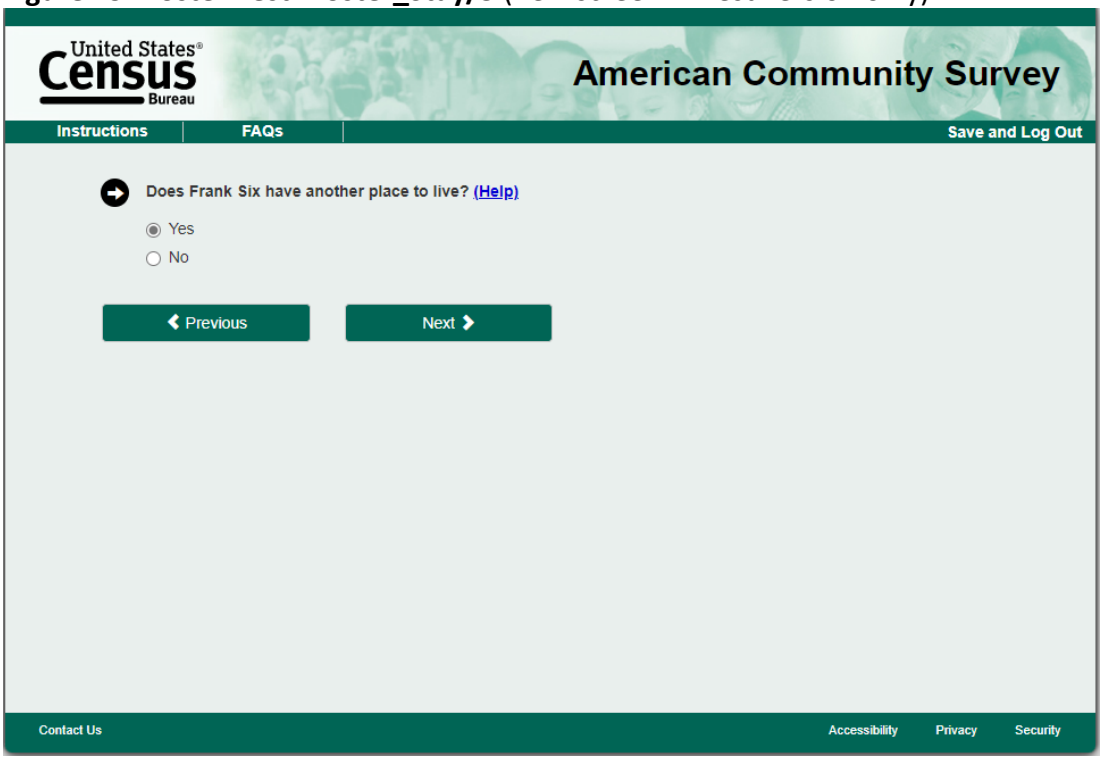


Figure 30. Roster Test: Roster_Check

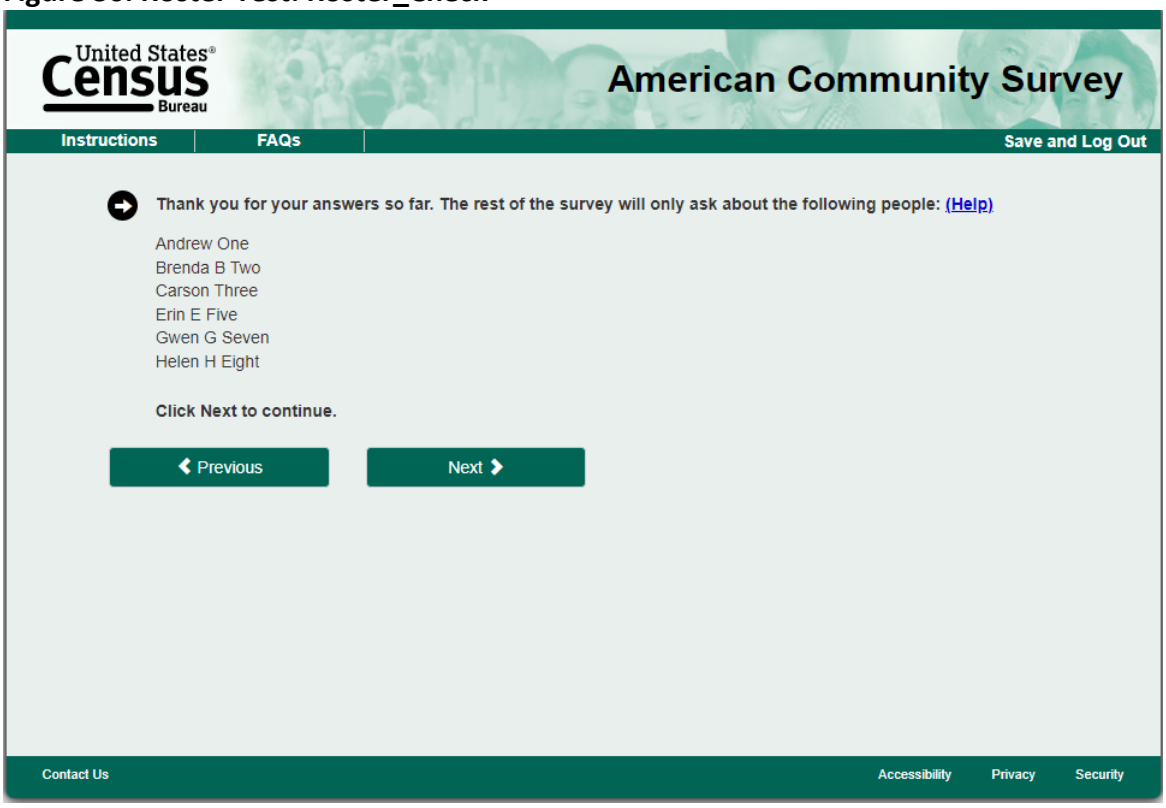


Figure 31. Full Screen Path