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

November 20, 2023

2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS23-RER-04

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: Matthew Virgile *MV*
Chin-Fang Weng *CFW*
Center for Behavioral Science Methods

Gregory J. Mills *GJM*
American Community Survey Office

Samantha Spiers *SS*
Decennial Statistical Studies Division

Subject: 2022 American Community Survey Content Test Evaluation Report: Respondent Burden

Attached is the 2022 American Community Survey (ACS) Content Test report for the respondent burden analysis. This report discusses the impact the new and revised questions had on respondent burden.

If you have any questions about this report, please contact Matthew Virgile at 301-763-4745.

Attachment

DRB Clearance Number – CBDRB-FY23-ACSO003-B0063

cc:

acso.re.workgroup.list@census.gov	
Luke Larsen	CBSM
Angela C O'Brien	
Lauren Contard	DSSD
Lindsay Longsine	
Peter Massarone	
Kelly Mathews	
Broderick Oliver	
Michael Risley	
Hyon Shin	SEHSD
Sharon Stern	

This page intentionally left blank

DRB Clearance Number – CBDRB-FY23-ACSO003-B0063

American Community Survey Research and Evaluation Program

November 20, 2023

2022 American Community Survey Content Test Evaluation Report: Respondent Burden

FINAL REPORT

Matthew Virgile & Chin-Fang Weng
Center for Behavioral Science Methods

Greg Mills
American Community Survey Office

Samantha Spiers
Decennial Statistical Studies Division

United States™ Census Bureau

This page intentionally left blank

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

Overview

From September to December of 2022, the Census Bureau conducted the 2022 American Community Survey (ACS) Content Test, a field test of new and revised content. The 2022 ACS Content Test tested the wording, format, and placement of both new and revised questions for potential inclusion in the ACS data collection instruments. This report discusses the impact the new and revised questions had on respondent burden.

In preparation for the 2022 ACS 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 (referred to as Test and Roster).

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 followup phase, conducted via Computer-Assisted Personal Interviewing (CAPI). The CAPI operation lasted about one month.

Respondent Burden

We examined respondent burden in the 2022 ACS Content Test to ensure that any changes to ACS content will not cause undue burden on future ACS respondents. Our measures of burden included comparisons between treatments for median survey completion time, median time spent on particular topics, rates of clicking a “help” access screen within these topics, overall and topic-level breakoff rates, and form completeness rates.

Based on nominal comparisons of median survey completion times (without tests for statistical significance), the Test treatment took respondents longer to complete than either the Control or Roster treatment. The median completion times for all treatments fell between 32-33 minutes for internet and between 18-20 minutes for CAPI, with the Test treatment taking the longest overall.

For median time spent on particular topics, most topics had differences of less than five seconds between treatments. However, three topics had larger differences:

- The Disability topic had the largest overall difference in time spent between treatments. The median time spent on the Disability questions in the Test treatment was 22 seconds longer in the internet mode and about 30 seconds longer in the CAPI mode than in either of the other treatments. This difference is most likely because the question set in the Test treatment had more response options and one extra question compared to the question sets in the other treatments.

- For the Health Insurance Coverage topic, there were differences in median completion times between treatments of 13 seconds for the internet mode and between three to nine seconds for the CAPI mode, with the median time spent being lowest in the Roster treatment. The Roster treatment had a mark-all-that-apply question design unlike the other treatments which had individual yes/no questions.

- For the Labor Force topic, there were differences in median completion times between treatments of seven seconds for the internet mode and of about 10 seconds for the CAPI mode, with the median time spent being lowest in the Control treatment. The Test and Roster treatments had one extra question in the Labor Force topic than the Control treatment.

For help screen access, four topics had rates that were significantly different between treatments:

- The Income topic had significant differences between each pair of treatments, with the highest rate for the Test treatment, followed by the Roster treatment, and then the Control treatment.

- The Household Roster topic had a significantly higher help screen access rate in the Control treatment than in the Roster treatment.

- The Health Insurance Coverage topic had a significantly lower help screen access rate in the Roster treatment compared to the Control and Test treatments.

- The Supplemental Nutrition Assistance Program (SNAP) topic had a significantly higher help screen access rate in the Test treatment than in the Control treatment.

For breakoff rates, the Test and Roster treatments each had significantly higher overall breakoff rates than the Control treatment. However, only the Income topic had a significant difference in breakoff rates between any treatments. The Test treatment had significantly more households break off at the Income topic than did the Control treatment.

For form completeness, two of the response modes had significant differences between treatments. For internet response, the Control treatment had the highest form completeness

rate overall and for the Detailed Person section, followed by the Roster treatment, and then the Test treatment. For CAPI response, the overall completeness rate in the Control treatment was significantly higher than in the Test treatment. There were no significant differences between treatments for other sections of the questionnaire or for the mail mode.

Conclusions

Overall, the Test treatment had the highest respondent burden among the three treatments. The Test treatment had the longest median survey completion time and had significantly lower internet form completeness rates, overall and for the Detailed Person section. When considering help screen access rates and breakoff rates, the results are mixed in regard to which treatment was most burdensome. The help screen access rates were not significantly different among treatments for most topics, and for the four where there were differences, the treatment with the highest rate varied by topic. For breakoff rates, both the Test and Roster treatments had significantly higher overall breakoff rates than the Control treatment, and the only significant topic-level difference was in Income between the Control and Test treatments, where more households broke off in the Test treatment than the Control treatment. Decisions for future ACS questionnaire design should consider the results from this analysis as well as the results from the topic-level reports on response data from each treatment.

1. INTRODUCTION

From September to December of 2022, the Census Bureau conducted the 2022 American Community Survey (ACS) Content Test, a field test of new and revised content. The goal of the test was to improve the content of the ACS data collection instruments by testing the wording, format, and placement of both new and revised questions. The results will help determine future ACS content and assess the expected data quality of revised questions and new questions added to the ACS.

The 2022 ACS Content Test included the following topics:

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

This report discusses the impact the new and revised questions had on respondent burden. Separate reports discuss the methods and results of testing the individual question topics included in the Content Test.

2. BACKGROUND

2.1 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 revisions to existing questions, the proposals contained a justification for each change and described previous testing of question wording, the expected impact of revisions to ACS estimates, and the estimated net impact on respondent burden.

For new questions, the proposals provided an explanation of the need for the new data, why other data sources that provide similar information are not sufficient, how policy or data needs would be addressed through the new questions, and an explanation of why the data were needed with the geographic precision and frequency provided by the ACS. 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 existing 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.

2.2 Cognitive Testing

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 conducts cognitive interviewing to pretest survey questions prior to field testing or implementing the questions in production. In preparation for the 2022 ACS Content Test, the Census Bureau contracted with Research Triangle Institute (RTI) International to conduct three rounds of cognitive testing. The first two rounds focused on the proposed content for the Content Test field test, with cognitive interviews from participants nationwide (excluding Puerto Rico) about their housing unit. The third round was conducted with cognitive interviews from participants in Puerto Rico and in Group Quarters (GQ), but the 2022 ACS Content Test ultimately did not include field testing in these areas, as survey data collection for the Content Test would be limited to housing units in the statewide US. For more information on the cognitive testing procedures and results, see RTI International (2022a) and RTI International (2022b).

2.3 Motivation for Respondent Burden Analysis

As the premier source for detailed population and housing information about the U.S., the ACS is designed to collect data that meet the needs of federal government agencies as well as provide local area information to state, local, and tribal governments. However, the Census Bureau must balance the need for detailed information with respect for people's privacy and time. We examined respondent burden in the Content Test to ensure that any changes to ACS content will not cause undue burden on respondents due to the mandatory reporting requirement.

2.4 Research Questions

The research questions for this analysis were as follows:

1. *Is there a difference between completion times for the test questionnaires versus the Control questionnaire? (internet and CAPI)*

2. *Is there a difference between completion times for the individual tested topics? (internet and CAPI)*

3. *Is there a difference in the number of times respondents access the help screens for the tested questions for the test treatments versus the Control treatment? (internet only)*

4. *Is there a difference in breakoff rates for the tested questions? (internet only)*

5. *Is there a difference in the form completeness rates in the original interview for the test treatments versus the Control treatment? (internet, mail, and CAPI)*

3. METHODOLOGY

3.1 Experimental Design

The 2022 ACS Content Test had three treatments: a Control treatment, a Test treatment, and a Roster Test treatment. Table 1 shows which question version was in each treatment for all the Content Test topics.

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 version was asked in the Control treatment and Roster Test treatment, and the other version was asked 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 number and types of people included in the household could impact the results of person-level topics. Three topics – Health Insurance Coverage, Labor Force, and Income – had two test versions in addition to the Control. Therefore, the analyses for Test Version 2 of these questions could have been impacted by changes in the Household Roster questions. 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.[1]

[1] 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.

Table 1. Questions by Treatment

Topic	Control Treatment	Test Treatment	Roster Test Treatment
Household Roster	Production	Production	Test Version
Sewer	Test Version	Test Version	Test Version
Electric Vehicles	Test Version 1	Test Version 2	Test Version 1
Solar Panels	Test Version	Test Version	Test Version
SNAP	Production	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
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.

3.2 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.[2]

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 total, three samples were selected, one for the Control treatment and two for the two test treatments.

[2] 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).

3.3 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 (CAPI) operation, where Census Bureau field representatives (FRs) attempt to complete a survey for a subsample of the remaining nonresponding addresses. During CAPI, FRs conducted interviews in person and over the phone. Self-response via internet or paper was accepted throughout the three-month data collection period.

Respondents completing the ACS through Telephone Questionnaire Assistance (TQA), a toll-free phone number people can call to ask questions about the ACS or respond to it, were excluded from the Content Test analyses. If respondents called TQA and opted to complete the survey over the phone, the interviewers conducted the survey using the production ACS questionnaire.[3]

Content Test data collection also included a Content Follow-Up (CFU) reinterview to measure response error. A CFU reinterview was attempted with every household with an original Content Test interview that met the CFU eligibility requirements. All CFU reinterviews were conducted by telephone, and cases were 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.[4] We did not examine CFU response data for this report.

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

3.4 Data Sources

3.4.1 Response Data

The 2022 ACS Content Test response data included questionnaire responses from the original interview and CFU reinterview, as well as some variables related to data processing. Because there are three response modes for the original interview, a sample address could have

[3] 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.

[4] For more information on the CFU operations, see Spiers et al. (2023).

multiple returns. We selected which return to use for analysis using a Primary Selection Algorithm (PSA). The PSA chose which returns to keep using the following steps:

1. If a sample address has only one return, keep that return.
2. If there was a CFU reinterview attempted, select the original Content Test return that triggered the CFU attempt.
3. If there was no CFU reinterview attempted, select the return with the earliest return date. If two returns have the same return date, select the one with the mode that occurs first in this list: (1) internet, (2) mail, and (3) CAPI.

The result of the process was analysis files with data from at most one return per sample address.

We did not alter response data using typical production ACS edits because the primary concern of this test was how changes to existing questions and additions of new questions affected the unaltered responses provided directly by respondents. For this reason, responses were not imputed either. We applied a few edits to the non-topic data, such as calculating a person's age based on their date of birth, but such edits were minimal.[5]

3.4.2 Internet and CAPI Paradata

Paradata are the data collected about the survey process beyond the responses themselves. Paradata can be used to measure respondent burden by assessing how a respondent interacts with a survey instrument. It can also indicate what screens were seen by respondents, which can help identify breakoff points.

For the ACS, paradata from internet respondents are collected in the internet instrument. Unlike survey response datasets, which typically contain one row for each unique household or person, the web survey paradata typically contain hundreds of rows for each unique household. These include data at the household ID level upon every login, with an observation for every action by the respondent (such as clicking a navigation button, clicking a response option, or entering a text response), the screens on which these actions took place, and timestamps for every action during a particular login session. These data can be used to compute various measures of respondent burden, such as survey completion time for each household, time spent on specific screens or topics, whether a "help" hyperlink was ever clicked by a respondent and at what points, and whether someone began responding to the ACS but failed to complete it and submit their final responses (i.e., breakoffs).

For CAPI, paradata are available based on actions by FRs in the CAPI data collection instrument. These paradata provide much of the same information as that from internet paradata and can

[5] 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.

be used to compute various metrics such as survey completion time or time spent on specific topics during an interview. The CAPI paradata also includes information about tasks completed by the FR that fall outside the administration of the survey questionnaire; we exclude these actions from our analysis.

3.5 Analysis Metrics

The purpose of the 2022 ACS Content Test was to field test new and revised questions in the ACS data collection instruments. For the first two research questions (on completion times), we assessed the impact the new and revised questions had on respondent burden by examining paradata from the internet and CAPI instruments. For the next two research questions (on accessing help screens and breakoff rates), we examined paradata from the internet instrument only. For the final research question (on form completeness), we examined Content Test response data from each of the three survey response modes (mail, internet, and CAPI). Analysis metrics using internet or CAPI paradata were unweighted; form completeness rates using response data were weighted using final Content Test weights (see Section 3.6 for more information). For help screen access rates, breakoff rates, and form completeness rates, we compared treatments using two-tailed t-tests, adjusted for multiple comparisons using the Hochberg method (Hochberg, 1988).

3.5.1 Completion Times

Adding new questions to the data collection instruments may increase the median time it takes to complete the ACS. The same is true of the modified questions, some of which add details to improve question clarity. The time to complete a survey is an important metric for assessing respondent burden.

We calculated the survey completion time for each household with an internet or CAPI response as the difference between the first and last action, based on timestamp. We removed particular gaps in time due to exiting a survey before completion and logging in again later. We calculated the median completion time across all households.[6]

We limited completion time calculations to only occupied households (excluding vacant households) with "complete" responses. The Content Test response data include whether a household is occupied or vacant based on a variable indicating the number of people living at the sampled address. For internet, "complete" responses are those from households which have a unique ID in the internet paradata, and where the respondent clicked the "submit"

[6] For completion times, we followed traditional ACS practices by reporting median values, without standard errors. Medians are treated as a preferable metric to means, since mean values for completion times (and their associated standard errors) may be skewed by respondents or FRs who took excessive time due to distractions or multi-tasking. We acknowledge, however, that this type of skewness would not be an argument against presenting standard errors of *medians*. Thus, the reporting of standard errors for *median* completion times may be included in future ACS reports, depending on policy discussions.

button at the end of the survey to submit their final responses. For CAPI, "complete" responses are those from households which have a unique ID in the CAPI paradata along with an outcome code indicating a complete interview. Also, our calculations for CAPI omitted a subset of households which had incomplete paradata files.

We calculated the unweighted median survey completion times for each treatment by internet and CAPI modes, as well as by household size in each of these modes (1-person, 2-person, and 3+ person households). We also calculated median completion times for each of the ten individual test topics in these modes. Finally, for the Household Roster topic, and for test topics which are normally administered about individual household members (Educational Attainment, Health Insurance Coverage, Disability, Labor Force, and Income), we calculated median completion times by household size for each mode.

It is important to note that while some topics may have larger differences in median completion time between treatments than others, we cannot say in this report whether any differences in completion time are statistically significant.

3.5.2 Accessing Help Screens

Accessing the help screen could indicate that the respondent did not understand the question being asked and needed further instructions. We calculated and compared how frequently the help screens were accessed among the treatments for each tested topic using the following unweighted formula[7]:

$$\text{Help Access Rate by topic} = \frac{\text{Number of households where "help" was clicked on a screen within the topic}}{\text{Total number of households where a screen within the topic was accessed}} \times 100$$

This calculation was done for internet only. We excluded CAPI responses because we were only interested in when the respondent accessed the help screen, not the interviewer. Also, we included households regardless of survey completion status and included both occupied and vacant households, since some test topics would still be administered to vacant households. The numerator in the formula above was the number of households whether "help" was ever clicked on a screen within a topic; it was not relevant in this formula whether "help" was clicked by a household more than once within a topic, or on which particular screen(s) within the topic.

[7] For this report, we followed traditional ACS practices by reporting unweighted values for metrics based on paradata such as survey completion time, help access rates, and breakoff rates. However, the reporting of weighted values for these metrics may be included in future ACS reports, depending on policy discussions.

For the individual topics, we made statistical comparisons between treatments. Which treatments were included in the comparison were determined based on question differences between the treatments and whether comparisons were used for the topic-level report. For example, there were no differences in question design for Sewer and Solar Panels, thus we did not compare treatments for these topics. See Table 1 for which question version was in each treatment for all of the Content Test topics.

3.5.3 Breakoff Rates

A breakoff occurs when the respondent leaves the instrument without ever completing the survey. A breakoff may indicate potential problems with an individual question or topic. If a breakoff occurs, we can see which screen the respondent was on and whether it was within one of the topics being tested.

We defined a breakoff occurring if the respondent accessed the internet instrument but did not ever reach the "submit" screen or action in any session. We calculated the overall unweighted breakoff rate for each treatment as the proportion of breakoffs among all households that ever logged in to begin the internet survey within that treatment. We also calculated unweighted breakoff rates for each test topic, within treatment, using the following formula:

$$\text{Breakoff Rate by topic} = \frac{\text{Number of internet breakoffs within the topic}}{\text{Total number of households with an internet response started}} \times 100$$

The denominator in the formula above does not vary by topic, but always uses the number of households where an internet response was started within a treatment group. Also, these calculations include both occupied and vacant households. The numerator in the formula above was the number of households with a final recorded action on a screen within a topic, and that never completed the survey; it was not relevant in this formula whether a household logged out before completion if they logged in again later and completed the survey, or if they logged out more than once.

We calculated breakoff rates for internet only. We excluded CAPI responses because it is less clear how to define breakoffs for this mode. The CAPI mode historically has fewer breakoffs than internet, and among the CAPI breakoffs that do occur, there are more potential reasons for a breakoff besides respondent burden, so the results are more difficult to interpret.

As with help access rates, we made statistical comparisons between treatments for the individual test topics based on any question differences between the treatments and whether

comparisons were used for the topic-level report. See Table 1 for which question version was in each treatment for all of the Content Test topics.

3.5.4 Form Completeness Rates

Form completeness provides a way to assess how much of the survey a respondent completed; it measures the number of questions on the form that were answered among those that should have been answered.[8]

The ACS has three question sections: a basic person-level demographic section, a housing section, and a detailed person-level section. We calculated weighted completeness rates for each section of the questionnaire, as well as overall. Calculations were made using the following formulas:

$$\text{Overall Form Completeness Rate} = \frac{\text{Number of questions answered}}{\text{Number of questions that should have been answered}} \times 100$$

$$\text{Section Completeness Rate} = \frac{\text{Number of questions answered in a specific section}}{\text{Number of questions in the specific section that should have been answered}} \times 100$$

We compared the form completeness rates among the treatments for the internet, mail, and CAPI modes. All form completeness rates were weighted using replicate base weights adjusted for CAPI sub-sampling.

3.6 Weighting and Standard Error Calculations

For help access rates, breakoff rates, and form completeness rates, 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:

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

[8] The number of questions that should have been answered is determined based on questionnaire skip patterns and respondent answers. If it is not clear if a question should have been answered (because a prior question was left blank), it is excluded from the calculation.

where:

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

The final 2022 ACS Content Test weights account for the initial probability of selection (the base weight) and CAPI sub-sampling.[9] For help screen access rates and breakoff rates, the respondent burden metrics are unweighted; we divided the replicate base weights by the base weight to allow for the weight of each observation to equal one, but still used the SDR method with the replicate factors to calculate the variance. Analysis involving internet and CAPI paradata were not weighted because, due to specificity of the data to different response modes, we cannot make generalizable conclusions to the population. Calculating and comparing unweighted metrics instead allows us to examine how the specific Content Test respondents interacted with the test topics.

For form completeness rates, the metrics are weighted; we applied the CAPI sub-sampling weights to each observation when computing rates for each treatment and used the SDR method with replicate weights to calculate the variance.

4. ASSUMPTIONS AND LIMITATIONS

4.1 Assumptions

1. The sample addresses for the Control and 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 Spiers et al. (2023) for more information.

2. 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).

4.2 Limitations

1. GQs were not included in the sample for the 2022 ACS Content Test. The results of the test may not extend to GQ populations.

[9] The Content Test weight creation process does not include all the steps followed in the ACS (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).

2. Housing unit addresses from Alaska, Hawaii, and Puerto Rico were not included in the sample for the test. The results may not extend to the housing unit population in these areas.

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

4. 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).

5. Due to 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.

6. TQA responses were excluded from the respondent burden analysis because survey responses completed via the TQA operation were only conducted using the ACS production data collection instrument.

7. We did not examine all respondent burden metrics for all modes. It is difficult to assess burden from paper questionnaire responses because it is unknown how someone interacted with the paper questionnaire without viewing their behavior. For CAPI, interviewers are trained to prevent breakoffs and can often provide help to respondents without accessing help screens.

5. RESEARCH QUESTIONS AND RESULTS

This section presents the results from the respondent burden analyses of the 2022 ACS Content Test field test.

5.1 Completion Times, Overall[10]

Is there a difference between completion times for the test questionnaire versus the Control questionnaire? (internet and CAPI)

[10] Note that while some treatments may have larger differences in median completion time than others, we cannot say in this report whether any differences in completion time are statistically significant.

Table 2. Median Survey Completion Time – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	32:42	32:58	32:30	-0:16	0:12	0:28
	1	22:48	23:09	22:49	-0:21	-0:01	0:20
	2	34:02	34:44	33:46	-0:42	0:16	0:58
	3+	40:53	41:05	40:15	-0:12	0:38	0:50
CAPI	All	18:45	19:31	18:52	-0:46	-0:07	0:39
	1	13:16	13:06	12:30	0:10	0:46	0:36
	2	18:41	20:00	18:49	-1:19	-0:08	1:11
	3+	23:38	25:48	25:26	-2:10	-1:48	0:22

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 2 provides the median survey completion time in each mode and treatment, as well as by household size (1-person homes, 2-person homes, and 3+ person homes). Note that completion times are longer for internet than for CAPI. Based on internet paradata, the median completion time in all three treatments fell between 32 and 33 minutes, and the time increased for households with more occupants. Based on CAPI paradata, the median completion time in all three treatments fell between 18 and 20 minutes, and the time also increased for households with more occupants.

The median time for the Test treatment was the highest among the three treatments in either mode (with 32:58 in internet, and 19:31 in CAPI). Also, in either mode the difference in overall median time between the Test treatment and either of the other two treatments (Control or Roster) is higher than the difference between the Control and Roster treatments. These findings may be due in part to the experimental design; for the experimental topics in which two treatments had the same design and one had a different design, the Control and Roster treatments usually had the same design, while the Test treatment had a different design.

5.2 Completion Times, By Topic[11]

Is there a difference between completion times for the individual test topics? (internet and CAPI)

Table 3. Median Completion Time for Household Roster Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	1:56	1:56	1:52	0:00	0:04	0:04
	1	1:37	1:38	1:37	-0:01	0:00	0:01
	2	1:55	1:56	1:49	-0:01	0:06	0:07
	3+	2:10	2:11	2:07	-0:01	0:03	0:04
CAPI	All	1:12	1:13	1:18	-0:01	-0:06	-0:05
	1	0:43	0:44	0:48	-0:01	-0:05	-0:04
	2	1:08	1:05	1:12	0:03	-0:04	-0:07
	3+	1:45	1:52	1:54	-0:07	-0:09	-0:02

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Median times in minutes and seconds (MM:SS).

Table 3 provides the median time spent on the Household Roster topic, by mode, household size, and treatment. The design for the Household Roster topic was the same for the Control and Test treatments, and different for the Roster treatment. For internet, the median time spent in the Roster treatment (one minute and 52 seconds) was four seconds less than in either of the other treatments (one minute and 56 seconds each). However, for CAPI, the median time spent in this treatment (one minute and 18 seconds) was five or six seconds longer than in the Control and Test treatments (one minute and 12 seconds, and one minute and 13 seconds, respectively). When examining the results by household size within mode, the same trends hold, with the Roster treatment taking a few seconds less than the Control or Test treatments in internet, but a few seconds longer than these treatments in CAPI.

Table 4. Median Completion Time for Sewer Topic – By Mode and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:06	0:06	0:06	0:00	0:00	0:00
CAPI	All	0:06	0:06	0:06	0:00	0:00	0:00

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Median times in minutes and seconds (MM:SS).

[11] Note that while some topics may have larger differences in median completion time between treatments than others, we cannot say in this report whether any differences in completion time are statistically significant.

Table 4 provides the median time spent on the Sewer topic, by mode and treatment. This topic had no difference in design between treatments. In either mode or in any treatment, the median time spent on this topic was six seconds. Since this topic was administered at the household level and not the person level, we did not provide results here by household size, nor did we provide this for the next three topics which are at the household level (Electric Vehicles, Solar Panels, and SNAP).

Table 5. Median Completion Time for Electric Vehicles Topic – By Mode and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:08	0:06	0:08	0:02	0:00	-0:02
CAPI	All	0:10	0:07	0:10	0:03	0:00	-0:03

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 5 provides the median time spent on the Electric Vehicles topic, by mode and treatment. In either mode, the median time spent on this topic was the same in the Control and Roster treatments, with eight seconds in internet and 10 seconds in CAPI. The Test treatment, which differed in design from the other treatments, took two seconds less than these treatments in internet (median time of six seconds), and three seconds less than these treatments in CAPI (median time of seven seconds).

Table 6. Median Completion Time for Solar Panels Topic – By Mode and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:04	0:04	0:04	0:00	0:00	0:00
CAPI	All	0:05	0:05	0:05	0:00	0:00	0:00

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 6 provides the median time spent on the Solar Panels topic, by mode and treatment. Like the Sewer topic, this topic had no difference in design between treatments. We found the same median times in each treatment, with four seconds in internet and five seconds in CAPI.

Table 7. Median Completion Time for SNAP Topic – By Mode and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:06	0:07	0:07	-0:01	-0:01	0:00
CAPI	All	0:06	0:06	0:06	0:00	0:00	0:00

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 7 provides the median time spent on the SNAP topic, by mode and treatment. This topic had the same design in the Test and Roster groups, but a different reference period for the question in the Control group. The median time spent on this topic for internet in the Control treatment was six seconds, one second less than either the Test or Roster treatments (at seven seconds each). In CAPI, the median time spent was the same in all three treatments, at six seconds.

Table 8. Median Completion Time for Educational Attainment Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:31	0:29	0:31	0:02	0:00	-0:02
	1	0:17	0:16	0:17	0:01	0:00	-0:01
	2	0:30	0:27	0:30	0:03	0:00	-0:03
	3+	0:47	0:44	0:47	0:03	0:00	-0:03
CAPI	All	0:18	0:19	0:18	-0:01	0:00	0:01
	1	0:09	0:09	0:09	0:00	0:00	0:00
	2	0:17	0:18	0:17	-0:01	0:00	0:01
	3+	0:30	0:31	0:31	-0:01	-0:01	0:00

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 8 provides the median time spent on the Educational Attainment topic, by mode, household size, and treatment. This topic had the same design in the Control and Roster treatments, but a different design in the Test treatment. The median time spent on this topic in the Test treatment for internet was 29 seconds, two fewer than either the Control or Roster treatments (with 31 seconds each). In CAPI, however, the median time spent on this topic in the Test treatment was 19 seconds, one more than either of the other two treatments (with 18 seconds each). This trend held by household size as well, with the Test treatment taking the shortest time of all three treatments in internet, while taking nearly the same time as the Control and Roster treatments in CAPI.

Table 9. Median Completion Time for Health Insurance Coverage Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:44	0:44	0:31	0:00	0:13	0:13
	1	0:26	0:27	0:19	-0:01	0:07	0:08
	2	0:45	0:45	0:32	0:00	0:13	0:13
	3+	1:01	1:01	0:41	0:00	0:20	0:20
CAPI	All	0:37	0:43	0:34	-0:06	0:03	0:09
	1	0:21	0:25	0:22	-0:04	-0:01	0:03
	2	0:36	0:43	0:34	-0:07	0:02	0:09
	3+	0:49	1:00	0:46	-0:11	0:03	0:14

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 9 provides the median time spent on the Health Insurance Coverage topic, by mode, household size and treatment. This topic had a different design in each of the three treatments; the Control and Test versions asked respondents to mark “Yes” or “No” for each of several insurance types, while the Roster version asked respondents to choose all that apply from the list of insurance types.[12]

For internet, the median time spent on this topic in the Roster treatment was the shortest at 31 seconds, 13 seconds less than either the Control or Test treatments (with 44 seconds each). In CAPI, the Roster treatment also had the shortest median time spent on this topic, but with smaller differences between treatments than that found in internet; the median CAPI time for this topic in the Roster treatment was 34 seconds, which was three seconds less than the Control treatment (37 seconds) and nine seconds less than the Test treatment (43 seconds). These trends mostly held among the different household sizes, except for CAPI 1-person households where the median time for this topic in the Control treatment was one second less than that in the Roster treatment.

[12] For internet, respondents viewed a list of insurance options and could choose any that applied. For CAPI, respondents were shown a flashcard for in-person interviews, or were read aloud a list of insurance options for phone interviews, and asked which insurance options applied.

Table 10. Median Completion Time for Disability Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:57	1:19	0:57	-0:22	0:00	0:22
	1	0:33	0:50	0:34	-0:17	-0:01	0:16
	2	1:00	1:24	1:01	-0:24	-0:01	0:23
	3+	1:20	1:46	1:18	-0:26	0:02	0:28
CAPI	All	0:40	1:11	0:42	-0:31	-0:02	0:29
	1	0:29	0:51	0:29	-0:22	0:00	0:22
	2	0:42	1:14	0:45	-0:32	-0:03	0:29
	3+	0:51	1:29	0:54	-0:38	-0:03	0:35

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 10 provides the median time spent on the Disability topic, by mode, household size, and treatment. This topic had the same design in the Control and Roster treatments, containing six questions answered with either “Yes” or “No,” while the Test treatment had a different design, containing seven questions answered on a four-point scale.

Of all test topics, the Disability topic had the largest median difference in overall completion times between treatments. For internet, the Test treatment had a median time of one minute and 19 seconds for this topic; this time was 22 seconds longer than either the Control or Roster treatments (at 57 seconds each). For CAPI, the Test treatment had a median time of one minute and 11 seconds for this topic; this time was 31 seconds longer than the Control treatment (at 40 seconds), and 29 seconds longer than the Roster treatment (at 42 seconds). This trend held by household size as well, with the Test treatment taking the longest time of all three treatments, while the Control and Roster treatments took nearly the same amount of time.

Table 11. Median Completion Time for Labor Force Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	0:34	0:41	0:41	-0:07	-0:07	0:00
	1	0:17	0:22	0:22	-0:05	-0:05	0:00
	2	0:35	0:43	0:43	-0:08	-0:08	0:00
	3+	0:53	0:59	1:00	-0:06	-0:07	-0:01
CAPI	All	0:23	0:33	0:32	-0:10	-0:09	0:01
	1	0:11	0:14	0:13	-0:03	-0:02	0:01
	2	0:22	0:34	0:33	-0:12	-0:11	0:01
	3+	0:38	0:50	0:52	-0:12	-0:14	-0:02

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 11 provides the median time spent on the Labor Force topic, by mode, household size and treatment. This topic had the same design in the Test and Roster treatments, but a different design in the Control treatment.[13] Both the Test and Roster treatments contained one item more than the Control treatment.[14]

The Control treatment time for this topic was shorter than the other treatments, in both internet and CAPI. For internet, the median time in the Control treatment was 34 seconds, seven seconds less than the other treatments (at 41 seconds each). For CAPI, the median time in the Control treatment was 23 seconds, 10 fewer than the Test treatment (at 33 seconds) and nine fewer than the Roster treatment (at 32 seconds). This trend held by household size as well, with the Control treatment taking the shortest time of all three treatments, while the Test and Roster treatments took nearly the same amount of time.

[13] The Test and Roster treatments had different question designs for the paper instrument, but the same design for the automated instruments (internet and CAPI).

[14] For all treatments, the Labor Force topic included the following questions: when last worked, worked every week, number of weeks worked, and number of hours worked. The Test and Roster treatments also included the question worked even for a few days.

Table 12. Median Completion Time for Income Topic – By Mode, Household Size, and Treatment

Mode	Household Size	Control Median Time	Test Median Time	Roster Median Time	Control – Test	Control – Roster	Test – Roster
Internet	All	3:36	3:45	3:31	-0:09	0:05	0:14
	1	2:19	2:33	2:20	-0:14	-0:01	0:13
	2	4:15	4:20	4:10	-0:05	0:05	0:10
	3+	4:05	4:13	3:51	-0:08	0:14	0:22
CAPI	All	1:39	1:38	1:34	0:01	0:05	0:04
	1	1:16	1:14	1:09	0:02	0:07	0:05
	2	1:49	1:55	1:46	-0:06	0:03	0:09
	3+	1:54	1:52	1:50	0:02	0:04	0:02

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Median times in minutes and seconds (MM:SS).

Table 12 provides the median time spent on the Income topic, by mode, household size, and treatment. This topic had a different design in each of the three treatments, however, the Control and Roster treatments only differed in reference period. Note that this topic had the most items of all test topics, and therefore took the most time compared to other topics.

Overall, the Roster treatment had the shortest median time for this topic in both modes. For internet, the median time was three minutes and 31 seconds; this time was five seconds less than the Control treatment (at three minutes and 36 seconds), and 14 seconds less than the Test treatment (at three minutes and 45 seconds). For CAPI, the median Roster treatment time was one minute and 34 seconds; this time was five seconds less than the Control treatment (at one minute and 39 seconds), and four seconds less than the Test treatment (at one minute and 38 seconds). This trend mostly held by household size as well, except for one-person households in internet where the Control treatment was one second less than the Roster treatment.

5.3 Use of Help Screens

Is there a difference in the number of times respondents access the help screens for the tested topics in each treatment? (internet only)

Table 13. Help Screen Access Rates for All Test Topics, by Treatment

Topic	Control Percent	Test Percent	Roster Percent
Household Roster	3.8 (0.2)	3.7 (0.2)	3.1 (0.2)
Sewer	1.7 (0.1)	1.8 (0.1)	1.6 (0.1)
Electric Vehicles	0.3 (0.1)	0.2 (0.1)	0.2 (<0.1)
Solar Panels	0.1 (<0.1)	0.2 (<0.1)	0.2 (<0.1)
SNAP	0.1 (<0.1)	0.3 (<0.1)	0.2 (<0.1)
Educational Attainment	0.4 (0.1)	0.4 (0.1)	0.4 (0.1)
Health Insurance Coverage	0.5 (0.1)	0.4 (0.1)	0.2 (<0.1)
Disability	2.9 (0.2)	3.1 (0.2)	2.6 (0.2)
Labor Force	2.1 (0.2)	2.2 (0.1)	2.2 (0.2)
Income	11.5 (0.3)	14.8 (0.3)	12.9 (0.4)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Standard errors are in parentheses.

Table 13 provides the help screen access rate for each topic, by treatment group. As shown by these data, the income topic had the highest help screen access rate among all topics, with values that fell between 11 and 15 percent.[15] The next highest rates were among the Household Roster topic, which fell between three and four percent. After this topic, the Sewer, Disability, and Labor Force topics had help screen access rates that usually fell between one and three percent; all other topics had rates below one percent.

Significance testing for each topic is shown in the tables below. Note that for statistical tests within a topic, we only compared pairs of treatments in which there was a difference in design between treatments with regard to that topic. Therefore, we omit the Sewer and Solar Panels topics since these had the same design in each treatment. If the Control treatment had the same topic design as one other treatment (either Test or Roster) but not the other, we tested for statistical significance between only the Control treatment and the differing treatment for that topic. For the SNAP topic, since the Test and Roster treatments are the same, we tested for statistical significance between only the Control and Test treatments. Finally, note that in each table, the "Rate 1 Percent" pertains to the first of the two treatments listed in the corresponding "Comparison" cell, and the "Rate 2 Percent" pertains to the second of the two treatments listed in this cell.

[15] All comparative statements in this paragraph are nominal; we did not statistically test the ranking of topics in regard to help screen access rates.

Table 14. Difference in Help Screen Access Rates for Household Roster Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Roster	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-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 14 provides the help screen access rates for the Household Roster topic, with statistical comparisons between the Control and Roster treatments. The help screen access rate in the Control treatment was significantly higher than the rate for the Roster treatment.

Table 15. Difference in Help Screen Access Rates for Electric Vehicles Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	0.3 (0.1)	0.2 (0.1)	<0.1 (0.1)	0.69

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 15 provides the help screen access rates for the Electric Vehicles topic, with statistical comparisons between the Control and Test treatments. The help access rate for the Control treatment was not statistically significantly different from the Test treatment.

Table 16. Difference in Help Screen Access Rates for SNAP Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	0.1 (<0.1)	0.3 (<0.1)	-0.1 (0.1)	**0.05***

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 16 presents the help screen access rates for the SNAP topic, with statistical comparisons between the Control and Test treatments. The help access rate for the Control treatment was significantly lower than the Test treatment help access rate.

Table 17. Difference in Help Screen Access Rates for Educational Attainment Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	0.4 (0.1)	0.4 (0.1)	<0.1 (0.1)	0.76

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 17 provides the help screen access rates for the Educational Attainment topic, with statistical comparisons between the Control and Test treatments. The help access rates for the Control and Test treatments were not statistically significantly different.

Table 18. Difference in Help Screen Access Rates for Health Insurance Coverage Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	0.5 (0.1)	0.4 (0.1)	0.1 (0.1)	0.15
Control vs Roster	0.5 (0.1)	0.2 (<0.1)	0.3 (0.1)	**<0.01***
Test vs Roster	0.4 (0.1)	0.2 (<0.1)	0.2 (0.1)	**0.02***

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Table 18 provides the help screen access rates for the Health Insurance Coverage topic, with statistical comparisons between all three pairs of treatments. There was no statistically significant difference between the Control and Test treatments. However, the help screen access rates for the Control and Test treatments were significantly higher than the rate for the Roster treatment.

Table 19. Difference in Help Screen Access Rates for Disability Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	2.9 (0.2)	3.1 (0.2)	-0.2 (0.3)	0.51

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 19 provides the help screen access rates for the Disability topic, with statistical comparisons between the Control and Test treatments. There was no statistically significant difference between the Control treatment and the Test treatment.

Table 20. Difference in Help Screen Access Rates for Labor Force Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	2.1 (0.2)	2.2 (0.1)	-0.1 (0.2)	0.79
Control vs Roster	2.1 (0.2)	2.2 (0.2)	-0.1 (0.2)	0.79
Test vs Roster	2.2 (0.1)	2.2 (0.2)	0.1 (0.2)	0.79

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Table 20 provides the help screen access rates for the Labor Force topic, with statistical comparisons between all three pairs of treatments. There were no statistically significant differences between any of the treatments.

Table 21. Difference in Help Screen Access Rates for Income Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	11.5 (0.3)	14.8 (0.3)	-3.3 (0.5)	**<0.01***
Control vs Roster	11.5 (0.3)	12.9 (0.4)	-1.4 (0.5)	**<0.01***
Test vs Roster	14.8 (0.3)	12.9 (0.4)	1.9 (0.5)	**<0.01***

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Finally, Table 21 provides the help screen access rates for the Income topic, with statistical comparisons between all three pairs of treatments. All three treatment comparisons had significant differences; the help access rate in the Test treatment was significantly higher than that in the Control treatment and in the Roster treatment. Additionally, the rate in the Roster treatment was significantly higher than that in the Control treatment.

5.4 Breakoff Rates

Is there a difference in breakoff rates, overall or for the tested topics in each treatment? (internet only)

Table 22. Survey Breakoff Rate Overall and for Test Topics, by Treatment

Topic	Control Percent	Test Percent	Roster Percent
Overall	19.4 (0.4)	21.3 (0.4)	21.4 (0.4)
Household Roster	1.1 (0.1)	1.2 (0.1)	1.0 (0.1)
Sewer	0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
Electric Vehicles	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
Solar Panels	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
SNAP	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
Educational Attainment	0.1 (<0.1)	0.1 (<0.1)	0.2 (<0.1)
Health Insurance Coverage	0.2 (<0.1)	0.2 (<0.1)	0.1 (<0.1)
Disability	0.4 (0.1)	0.3 (0.1)	0.5 (0.1)
Labor Force	0.4 (0.1)	0.4 (0.1)	0.4 (0.1)
Income	2.5 (0.2)	3.2 (0.1)	2.9 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Standard errors are in parentheses.

Table 22 provides overall breakoff rates for all three treatments, as well as rates of breakoffs for each of the ten test topics. Note the topic level breakoff rates do not sum to the overall breakoff rates, since breakoffs may occur at many other points in the survey outside of the ten test topics. Recall that we define breakoffs as instances in which a household began but did not complete the survey; we do not consider households with multiple logins and logouts to be breakoffs as long as they completed the survey.

The Income topic had the highest rate of breakoffs among the ten test topics; between two and four percent of households broke off somewhere in this topic in each treatment. This was followed by the Household Roster topic; about one percent of households broke off somewhere in this topic in each treatment. For the other topics, no more than 0.5 percent broke off at that topic location.[16]

Significance testing for each topic is provided in the following tables, overall and for each topic. Note that for statistical tests within a topic, we only compared pairs of treatments in which there was a difference in design between treatments. Those topic-level treatment comparisons were the same as those for help screen access rates in Section 5.3. Like the tables in the previous section, the "Rate 1 Percent" pertains to the first of the two treatments listed in the corresponding "Comparison" cell, and the "Rate 2 Percent" pertains to the second of the two treatments listed in this cell.

[16] All comparative statements in this paragraph are nominal; we did not statistically test the ranking of topics compared to each other in regard to survey breakoff rates.

Table 23. Difference in Overall Survey Breakoff Rates for each Treatment

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	19.4 (0.4)	21.3 (0.4)	-1.9 (0.5)	**<0.01***
Control vs Roster	19.4 (0.4)	21.4 (0.4)	-1.9 (0.5)	**<0.01***
Test vs Roster	21.3 (0.4)	21.4 (0.4)	-0.1 (0.5)	0.90

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Table 23 shows the overall breakoff rates for the survey, including statistical comparisons across treatments. The breakoff rate for the Control treatment (19.4 percent) was significantly lower than that for the Test treatment (21.3 percent) and that for the Roster treatment (21.4 percent). There was no statistically significant difference in the breakoff rates between the Test and Roster treatments.

Table 24. Difference in Breakoff Rates for Household Roster Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Roster	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-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level..

Table 24 provides breakoff rates for the Household Roster topic, including a statistical comparison between the Control and Roster treatments. There was no statistically significant difference in breakoff rates at this topic between treatments.

Table 25. Difference in Breakoff Rates for Electric Vehicles Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	0.44

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 25 provides breakoff rates for the Electric Vehicles topic, including a statistical comparison between the Control and Test treatments. There was no statistically significant difference between treatments.

Table 26. Difference in Breakoff Rates for SNAP Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	0.64

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 26 provides breakoff rates for the SNAP topic, including a statistical comparison between the Control and Test treatments. There was no significant difference between treatments.

Table 27. Difference in Breakoff Rates for Educational Attainment Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	0.1 (<0.1)	0.1 (<0.1)	<0.1 (<0.1)	0.71

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level..

Table 27 provides breakoff rates for the Educational Attainment topic, including a statistical comparison between the Control and Test treatments. There was no statistically significant difference between treatments.

Table 28. Difference in Breakoff Rates for Health Insurance Coverage Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	0.2 (<0.1)	0.2 (<0.1)	<0.1 (<0.1)	0.52
Control vs Roster	0.2 (<0.1)	0.1 (<0.1)	0.1 (<0.1)	0.29
Test vs Roster	0.2 (<0.1)	0.1 (<0.1)	0.1 (<0.1)	0.16

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Table 28 provides breakoff rates for the Health Insurance Coverage topic, including statistical comparisons between all three pairs of treatments. None of the breakoff rates were statistically significantly different between treatments.

Table 29. Difference in Breakoff Rates for Disability Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	P-Value
Control vs Test	0.4 (0.1)	0.3 (0.1)	<0.1 (0.1)	0.54

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level.

Table 29 provides breakoff rates for the Disability topic, including a statistical comparison between the Control and Test treatments. There was no statistically significant difference between treatments.

Table 30. Difference in Breakoff Rates for Labor Force Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	0.4 (0.1)	0.4 (0.1)	<0.1 (0.1)	0.99
Control vs Roster	0.4 (0.1)	0.4 (0.1)	<0.1 (0.1)	0.99
Test vs Roster	0.4 (0.1)	0.4 (0.1)	<0.1 (0.1)	0.99

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Table 30 provides breakoff rates for the Labor Force topic, including statistical comparisons between all three pairs of treatments. None of the breakoff rates were statistically significantly different between treatments.

Table 31. Difference in Breakoff Rates for Income Topic

Comparison	Rate 1 Percent	Rate 2 Percent	Difference	Adjusted P-Value
Control vs Test	2.5 (0.2)	3.2 (0.1)	-0.6 (0.2)	**<0.01***
Control vs Roster	2.5 (0.2)	2.9 (0.1)	-0.4 (0.2)	0.14
Test vs Roster	3.2 (0.1)	2.9 (0.1)	0.3 (0.2)	0.14

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method.

Finally, Table 31 provides breakoff rates for the Income topic, including statistical comparisons between all three pairs of treatments. Among the ten test topics, Income was the only one which had any significant difference in breakoff rates between treatments. For the Income

topic, the breakoff rate in the Control treatment was significantly lower than the rate in the Test treatment. The breakoff rate for the Roster treatment was not statistically significantly different from either the Control or Test treatment rates.

5.5 Form Completeness Rates

Is there a difference in the form completeness rates in the original interview for the test treatments versus the Control treatment? (internet, mail, and CAPI)?

Table 32 through Table 34 show the form completeness rates for the mail, internet, and CAPI modes. We examined rates by mode because of known differences in form completeness between them. For example, the involvement of an interviewer in CAPI can lead to higher form completeness than is seen in self-response modes. We also examined rates within questionnaire section because sections that appear first or have less questions can have higher form completeness.

Table 32. Mail Form Completeness Rates, Overall and Within Section, by Control (C), Test (T), and Roster (R) Treatments

Section	C Percent	T Percent	R Percent	C – T	C – T P-Value	C – R	C – R P-Value	T – R	T – R P-Value
Overall	88.1 (0.4)	88.0 (0.4)	88.2 (0.4)	0.1 (0.5)	0.81	-0.1 (0.5)	0.81	-0.2 (0.5)	0.81
Basic Person	96.8 (0.2)	97.4 (0.2)	97.2 (0.2)	-0.6 (0.3)	0.19	-0.4 (0.3)	0.77	0.2 (0.3)	0.77
Housing	91.2 (0.4)	92.3 (0.4)	92.4 (0.3)	-1.0 (0.5)	0.33	-1.2 (0.5)	0.12	-0.1 (0.5)	0.77
Detailed Person	85.3 (0.5)	84.8 (0.5)	85.0 (0.5)	0.5 (0.7)	0.77	0.3 (0.7)	0.77	-0.3 (0.7)	0.77

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method; adjustments were based on nine comparisons by section completeness (Basic Person, Housing, Detailed Person) and three comparisons by overall completeness.

Table 32 provides the mail form completeness rates, overall and within section, by treatment. The overall mail completeness rate is approximately 88 percent in each treatment; the rates were not statistically significantly different between treatments. When examining within questionnaire section, none of the rates were statistically significantly different between treatments.

Table 33. Internet Form Completeness Rates, Overall and Within Section, by Control (C), Test (T), and Roster (R) Treatments

Section	C Percent	T Percent	R Percent	C – T	C – T P-Value	C – R	C – R P-Value	T – R	T – R P-Value
Overall	90.7 (0.2)	88.6 (0.2)	89.7 (0.3)	2.1 (0.3)	**<0.01***	1.0 (0.3)	**<0.01***	-1.2 (0.3)	**<0.01***
Basic Person	99.1 (0.1)	99.2 (0.1)	99.1 (0.1)	<0.1 (0.1)	0.93	0.1 (0.1)	0.93	0.1 (0.2)	0.93
Housing	96.9 (0.1)	96.9 (0.1)	96.7 (0.2)	<0.1 (0.2)	0.93	0.2 (0.2)	0.93	0.2 (0.2)	0.93
Detailed Person	87.3 (0.3)	84.2 (0.3)	85.9 (0.4)	3.1 (0.5)	**<0.01***	1.3 (0.5)	**0.03***	-1.7 (0.4)	**<0.01***

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method; adjustments were based on nine comparisons by section completeness (Basic Person, Housing, Detailed Person) and three comparisons by overall completeness.

Table 33 provides the internet form completeness rates, overall and within section, by treatment. Overall, we found significant differences between all pairs of treatments in internet form completeness. The Control treatment rate was significantly higher than the rate for the Test treatment and the rate for the Roster treatment. In addition, the overall rate for the Test treatment was significantly lower than that for the Roster treatment. The same findings hold for the Detailed Person section form completeness rates. The form completeness rate in this section was highest in the Control treatment, followed by the Roster treatment, followed by the Test treatment, with significant differences between each pair of treatments. We did not find any statistically significant differences between treatments for either the Basic Person section or the Housing section.

Table 34. CAPI Form Completeness Rates, Overall and Within Section, by Control (C), Test (T), and Roster (R) Treatments

Section	C Percent	T Percent	R Percent	C – T	C – T P-Value	C – R	C – R P-Value	T – R	T – R P-Value
Overall	93.7 (0.3)	92.6 (0.4)	93.5 (0.3)	1.1 (0.5)	**0.06***	0.2 (0.3)	0.50	-0.9 (0.5)	0.12
Basic Person	99.6 (0.1)	99.4 (0.1)	99.7 (0.1)	0.2 (0.2)	0.69	<0.1 (0.1)	0.93	-0.2 (0.1)	0.69
Housing	92.6 (0.2)	92.1 (0.2)	92.6 (0.2)	0.5 (0.3)	0.69	-0.1 (0.3)	0.93	-0.6 (0.3)	0.57
Detailed Person	93.2 (0.4)	91.8 (0.5)	92.8 (0.4)	1.4 (0.7)	0.27	0.4 (0.5)	0.93	-1.1 (0.7)	0.69

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0063

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result. Significance was tested based on a two tailed t-test at the α=0.1 level. P-values were adjusted for multiple comparisons using the Hochberg method; adjustments were based on nine comparisons by section completeness (Basic Person, Housing, Detailed Person) and three comparisons by overall completeness.

Finally, Table 34 provides the CAPI form completeness rates, overall and within section, by treatment. We did not find any statistically significant differences between treatments for any of the three sections. However, for overall CAPI form completeness, there was a significant

difference between the Control and Test treatments, with the Control treatment rate being significantly higher than the Test treatment rate. There were no statistically significant differences between the Roster treatment rate and the Control and Test treatment rates.

6. CONCLUSIONS

We examined respondent burden in the 2022 ACS Content Test to ensure that any changes to ACS content will not cause undue burden on future ACS respondents. Our first analyses of respondent burden were based on internet and CAPI paradata, including median survey completion time and median time spent on specific topics, for each of the three treatments. Our next analyses of respondent burden were based on internet paradata only, including rates of clicking a "help" access screen and breakoff rates at specific survey topics, with tests for statistically significant differences between treatments. Our final analysis of respondent burden was based on response data from three survey modes (mail, internet, and CAPI), comparing form completeness rates overall and by section for each mode, with tests for statistically significant differences between treatments.

Overall, the Test treatment had the highest respondent burden among the three treatments. The Test treatment had the longest median survey completion time and had significantly lower internet form completeness rates, overall and for the Detailed Person section. When considering help screen access rates and breakoff rates, the results are mixed in regard to which treatment was most burdensome. Decisions for future ACS questionnaire design should take into account the results from this analysis and well as the results from the topic-level reports on response data from each treatment. The following summarizes the findings for each metric.

Median Completion Time

We did not include standard error values due to inherent skewness and thus did not statistically test differences in median survey completion times. However, based on nominal comparisons, the Test treatment took respondents longer to complete than either the Control or Roster treatments. For internet, the median completion time in the Test treatment was nearly 33 minutes; the other treatments took less time but fell within 30 seconds of the Test median. For CAPI, the median completion time in the Test treatment was between 19 and 20 minutes; the other treatments took less time but fell within one minute of the Test median.

The Test treatment taking longer to complete is mostly driven by the Disability topic, where the question set in the Test treatment had more response options and one extra question compared to the question sets in the Control and Roster treatments. For internet, the median time spent on this topic in the Test treatment was one minute and 19 seconds; this was 22 seconds longer than in either the Control or Roster treatments. For CAPI, the median time

spent on this topic in the Test treatment was one minute and 11 seconds; this was 31 seconds longer than in the Control treatment, and 29 seconds longer than in the Roster treatment.

Health Insurance Coverage and Labor Force also contributed to treatment differences. Median time spent on the Health Insurance Coverage topic in the Roster treatment was 13 seconds fewer for internet and less than 10 seconds fewer for CAPI compared to the Control and Test treatments. This difference is because the Roster treatment version was a mark-all-that-apply question as opposed to a series of yes/no questions in the other treatments. Median time spent on the Labor Force topic in the Control treatment was seven seconds fewer for internet and about 10 seconds fewer for CAPI compared to the Test and Roster treatments. This difference is again due to design differences since the Test and Roster treatments had an additional yes/no question compared to the Control treatment.

For any treatment, as expected, households with more occupants took longer to complete the survey than households with fewer occupants. The household-level topics (Sewer, Electric Vehicles, Solar Panels, and SNAP) took the least time, with median times of ten seconds or less in any treatment or mode. The person-level topics (Household Roster, Educational Attainment, Health Insurance Coverage, Disability, Labor Force, and Income) took longer since these were administered about multiple household members, and some contained multiple screens. Of these, the Income topic took the longest, with median internet time spent falling between three and a half and four minutes in any treatment, and median CAPI time spent slightly exceeding one and a half minutes in any treatment.

Help Screen Access

Higher help screen access rates could indicate that the respondent did not understand the question being asked and needed further instructions. Based on internet paradata, four topics had help screen access rates that were significantly different between treatments. For Household Roster, the Roster treatment rate was significantly lower than the Control treatment rate. For SNAP, the Test treatment rate was significantly higher than the Control treatment rate. For Health Insurance Coverage, the Roster treatment rate was significantly lower than the Control treatment rate and the Test treatment rate. And for Income, the Test treatment had the highest rate, followed by the Roster treatment, and then by the Control treatment.

Breakoff Rates

A large number of breakoffs may indicate potential problems with an individual question or topic. Considering the entirety of the survey, the Control treatment had significantly fewer breakoffs (19.4 percent) than either the Test treatment (21.3 percent) or Roster treatment (21.4 percent). Among topics, only the Income topic had a significant difference in breakoff rates between any treatments. The Control treatment had significantly fewer households break off at the Income topic (2.5 percent) than did the Test treatment (3.2 percent).

Form Completeness

Form completeness provides a way to assess how much of the survey a respondent completed. High form completeness can improve data quality and indicate less respondent burden. Based on Content Test response data, two of the response modes (internet and CAPI) had significant differences in form completeness between treatments. For mail form completeness, no treatments had any statistically significant difference, by section or overall. For internet form completeness, each pair of treatments had a significant difference in the Detailed Person section and overall; the Control treatment had the highest form completeness rate for this section and overall, followed by the Roster treatment, and finally by the Test treatment. For CAPI form completeness, there were no statistically significant differences in any section by treatment, but for overall form completeness, the rate in the Control treatment (93.7 percent) was significantly higher than in the Test treatment (92.6 percent). There is some evidence that differences in the test questionnaires led respondents to provide less information on their ACS responses than respondents from the Control treatment, but the impact varied by mode and questionnaire section.

7. ACKNOWLEDGEMENTS

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 (CBSM), 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.

We would also like to thank Luke Larsen from CBSM for his disclosure review of the report, as well as Angie O'Brien from CBSM for her assistance with summarizing findings in our analysis.

8. REFERENCES

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.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4), 800-802. https://doi.org/10.2307/2336325

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 September 16, 2022, from https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

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). *Cognitive testing for the 2022 ACS Content Test, round 1 and 2 combined briefing report.* Retrieved September 16, 2022, from https://www.census.gov/library/working-papers/2022/acs/2022_Wilson_01.html

RTI International. (2022b). *Cognitive testing for the 2022 ACS Content Test, round 3 briefing and recommendations report.* Retrieved September 16, 2022, from https://www.census.gov/library/working-papers/2022/acs/2022_Wilson_02.html

Spiers, S., Baumgardner, S., & Mojica, A. (2023). *2022 American Community Survey Content Test Evaluation Report: Unit-Level Response*. Retrieved November 20, 2023 from https://www.census.gov/library/working-papers/2023/acs/2023_Spiers_01.html

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

U.S. Census Bureau. (2022b). *American Community Survey and Puerto Rico Community Survey Design and Methodology, Version 3.0.* Retrieved January 6, 2023, from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html