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-11

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

From: David Waddington *David Waddington*
Chief, Social, Economic, and Housing Statistics Division

Prepared by: Adriana Hernandez-Viver *AH*
Katherine Keisler-Starkey *KKS*
Sharon M. Stern *SMS*
Social, Economic, and Housing Statistics Division

Samantha Spiers *SS*
Peter Massarone *PM*
Decennial Statistical Studies Division

Subject: 2022 American Community Survey Content Test Evaluation Report: Health Insurance Coverage

Attached is the 2022 American Community Survey (ACS) Content Test report for Health Insurance Coverage. This report presents the testing methods and results of revised versions of the Health Insurance Coverage question.

If you have any questions about this report, please contact Katherine Keisler-Starkey at 301-763-3647.

Attachment
cc:

acso.re.workgroup.list@census.gov		Broderick Oliver	
Joanne Pascale	CBSM	Michael Risley	
Lauren Contard	DSSD	Kelly Holder	SEHSD
Lindsay Longsine		Hyon Shin	

This page is intentionally blank

DRB Clearance Number – CBDRB-FY23-ACSO003-B0067

American Community Survey Research and Evaluation Program

November 20, 2023

2022 American Community Survey Content Test Evaluation Report: Health Insurance Coverage

FINAL REPORT

Adriana Hernandez-Viver, Katherine Keisler-Starkey, and Sharon M. Stern
Social, Economic, and Housing Statistics Division

Samantha Spiers and Peter Massarone
Decennial Statistical Studies Division

This page is intentionally blank

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Overview

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 Health Insurance Coverage.

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 (CFU) telephone reinterview was conducted to measure response error.

Health Insurance Coverage Question

The purpose of testing the revised health insurance question was to enhance question reliability and validity. Since implementation in 2008, analyses have revealed some limitations in the current measure. Testing as part of the 2016 ACS Content Test showed mixed results and was considered insufficient evidence to implement the set of proposed revisions. Berchick et al. (2017) concluded that some of the question changes should be tested again in the future. The revisions to the health insurance coverage question in the 2022 test incorporated some of those previously tested changes.

The primary objectives of revising the health insurance coverage question were:

1. to improve measurement of public coverage and accuracy of direct purchase coverage,

2. to reduce the reporting of single-service, non-comprehensive insurance plans, and
3. to reduce erroneous reports of multiple coverage.

The 2022 Content Test used several strategies to further these objectives. For example, the test treatments introduced an instruction to exclude single-service, non-comprehensive insurance plans in the paper and internet forms. Additionally, the test treatments re-ordered the options moving direct purchase down the list, after the more common types of coverage. Furthermore, the text for the options reflected changes in the health insurance landscape. These include adding recognizable terms like "Marketplace" and "HealthCare.gov." Finally, the Content Test included two test versions of the Health Insurance Coverage questions. Test Version 1 included proposed improvements as described above, while leaving the question format essentially unchanged: a series of yes/no items. Test Version 2 included almost identical wording to Test Version 1, but changed the style to a check-all-that-apply format and included a "no, uninsured" option.

Research Questions and Results

Control vs Test Version 1

The first part of the test was to determine whether the changes to question order and the text of the question represented an improvement over the Control. The decision criteria for these changes focused on the following: item missing data rates, write-in response distributions, and the frequency of multiple coverage reporting. Given the goals of testing the health insurance coverage question, the prevalence of coverage types was not considered a decision criterion in this analysis. It was only to be used to evaluate versions if there were not statistical differences in the higher priority items. Each research question was examined using comparisons between the Control Treatment and Test Version 1.

Although there were no statistically significant differences between treatments in any of the item missing data rates for complete missingness, the proportion of partial responses was significantly lower in Test Version 1 than Control (38.5 percent vs 40.7 percent, respectively). This difference was driven by the internet mode, where the proportions of partial responses in internet alone and self-response (mail and internet combined) were significantly lower in Test Version 1 than Control. The item missing data rates for Medicare, Medicaid, VA, TRICARE, Medicare combined with Medicaid, and Medicare combined with direct purchase were also significantly lower in Test Version 1 compared to Control.

Health insurance coverage write-in rates were another decision criterion. Since respondents may write in a health insurance coverage when they misunderstand the question or do not know how to categorize their insurance, a lower proportion of write-in responses was preferable. Results showed that the write-in proportions, as well as the proportion of write-ins determined to be out-of-scope, did not statistically differ between versions. On the other hand,

the proportion of codable write-ins that reference terms associated with the Affordable Care Act or Children's Health Insurance Program was significantly lower in Test Version 1 than Control (4.7 percent vs 8.0 percent, respectively), indicating that the addition of these terms to the question text reduced write-ins for those coverage types.

Multiple coverage reporting was considered too high in production ACS. A lower proportion of multiple types of health insurance coverage could indicate a reduction in respondent confusion on how to categorize their coverage. Overall, the proportion of persons with multiple types of coverage was significantly lower in Test Version 1 than Control (14.4 percent vs 16.1 percent, respectively). This difference was driven by the self-response modes (mail and internet responses), where Test Version 1 had significantly lower proportions than Control.

A lower Gross Difference Rate (GDR) is preferable because it means fewer responses were different between the original interview and the CFU reinterview. Only the GDR for direct purchase coverage was significantly different between treatments, where the percent of inconsistent answers in Test Version 1 was lower than in Control (10.2 percent vs 13.7 percent, respectively). Mail and internet responses also had significantly lower GDR results for direct purchase coverage in Test Version 1 compared to Control.

Using another measure of reliability, none of the overall Indexes of Inconsistency (IOIs) for the specific coverage types and no coverage were statistically significantly different between treatments. There was one significant difference by mode of original interview. The IOI for internet respondents who reported direct-purchase coverage was significantly different, with Test Version 1 having the lower rate compared to Control (43.0 percent vs 55.2 percent, respectively).

Finally, no significant differences were found between Test Version 1 and Control when it comes to the proportions of persons reporting any type of health insurance coverage. By individual type of coverage, however, Medicaid coverage rates, along with direct purchase, were significantly different between treatments, with both rates being significantly lower in Test Version 1 compared to Control. It is unclear if a difference in estimates meant an improvement in reporting (i.e., reduction in bias), so these differences alone were not sufficient and must be considered along with the other priorities.

The analysis showed that Test Version 1 met the decision criteria when compared with the Control Treatment on all of the key measures. Therefore, comparisons of prevalence of coverage and type of coverage were not considered in the decision criteria.

Test Version 1 vs Test Version 2

The second part of the test focused on whether changing the question format would allow for more complete and accurate reporting. Test Version 2 used the same wording as Test Version 1, but changed the response type. Where Test Version 1 presented a series of

coverage types, each with its own yes/no response, Test Version 2 presented the full set of coverage types with an instruction to mark all that apply or mark "No health insurance coverage or health coverage plan." The primary decision criteria for evaluating Test Version 2 against Test Version 1 were item missing data rates and response reliability.

Complete item missingness occurs when the respondent does not mark any part of the question or provide a valid write-in (e.g., leaves the question completely blank). There were no significant differences in complete item missingness between Test Version 1 and Test Version 2, overall and by mode.

A lower GDR was preferable because it means fewer responses were different between the original interview and the CFU reinterview. There were no significant differences in GDRs for Test Version 1 compared to Test Version 2 for those who had specific health insurance types. Among uninsured persons, Test Version 2 had significantly lower overall GDRs than Test Version 1.

For tests of reliability using the IOI measure, there were no significant differences overall between question versions for those with specific health insurance types. When looking by mode, the IOI for Test Version 2 was significantly lower for Veteran's health care (VA) responses in the CAPI mode than Test Version 1, although both values fell into the moderate consistency category (between 20 and 50 percent IOI). For uninsured persons, Test Version 2 had a significantly lower overall IOI than Test Version 1; again, both values were in the moderate consistency category.

There were no significant differences in the proportion of persons reporting any coverage, and response distributions for most types of health insurance coverage were not statistically different between Test Version 1 and Test Version 2. The sole exception was VA coverage, which was significantly lower in Test Version 2 (1.8 percent) than in Test Version 1 (2.4 percent).

A difference in the proportion of multiple types of health insurance coverage could indicate a reduction in respondent confusion about how to categorize their coverage. The proportion of persons with reports of multiple coverage was significantly lower in Test Version 2 than in Test Version 1 (11.2 percent vs 14.4 percent, respectively). Reports of multiple coverage also were significantly lower in Test Version 2 than in Test Version 1 for every response mode.

Since respondents may write in a health insurance coverage when they misunderstand the question, and manually assigning write-ins is time-consuming and costly, a lower proportion of write-in responses was preferable. Test Version 2 had a significantly lower proportion of respondents who wrote in a health insurance type compared to Test Version 1 (4.0 percent vs 4.8 percent, respectively).

The analysis showed mixed evidence when comparing Test Version 2 to Test Version 1. In discussion with key Federal agency stakeholders, the consensus was that improved reliability for the uninsured population is a critical benefit of Test Version 2. Further, additional research into the consistency of Medicaid reporting in surveys with administrative records from the Centers for Medicare and Medicaid Services (CMS) is an important next step in understanding the nature of reporting error.

Conclusions

Based on the decision criteria, the results of the 2022 ACS Content Test provided evidence to support the implementation of Test Version 1 of the Health Insurance Coverage question over the Control version, with mixed evidence to support the implementation of Test Version 2 over Test Version 1. Overall, the interagency subcommittee focusing on the health insurance question supports the recommendation to implement Test Version 2 of the Health Insurance Coverage question.

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 the Health Insurance Coverage question.

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.

Prior to the beginning of testing, the Interagency Council on Statistical Policy (ICSP) Subcommittee for the ACS offered member agencies the opportunity to provide a representative for topic-level subcommittees. These subcommittees participated in development and testing activities and consulted throughout the decision-making process. The Health Insurance Subcommittee included Census Bureau staff and representatives from other federal agencies including: the National Center for Health Statistics (NCHS), the Office of Health

Policy in the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, the Center for Medicare and Medicaid Services (CMS), the Agency for Healthcare Research and Quality (AHRQ), the U.S. Department of Veterans Affairs (VA), and the Social Security Administration (SSA).

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] 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.

1.3. Field Testing Health Insurance Coverage in the 2022 ACS Content Test

1.3.1. Justification for Inclusion of Health Insurance Coverage in the Content Test

The Census Bureau first introduced a question collecting information on a person's health insurance coverage on the ACS in 2008. The purpose of the question is "to enable the Department of Health and Human Services (HHS) and other federal agencies to more accurately distribute resources and better understand state and local health insurance needs" (U.S. Census Bureau, 2007).

The purpose of testing the revised health insurance question was to enhance question reliability and validity. Since implementation in 2008, analyses have revealed some limitations in the current measure. Among these, research has found that Medicaid and other means-tested programs are underreported (O'Hara, 2010; Boudreaux et al., 2011; Lynch et al., 2011; Boudreaux et al., 2014; Boudreaux et al., 2015). Other research has found that direct-purchase coverage is overreported, in part due to misreporting of non-comprehensive health plans, such as plans that cover only dental, vision, or prescription drug expenses and do not include hospital or physician coverage (not in scope in the ACS), and reporting multiple coverage types for the same plan (Mach & O'Hara, 2011; Lynch et al., 2011; Boudreaux et al., 2014).

Moreover, revisions to the health insurance coverage question are needed to help capture changes to the health insurance landscape that occurred with and since the passage of the Patient Protection and Affordable Care Act (United States Congress, 2010).

In 2016, the Census Bureau tested changes to the health insurance question series as part of the 2016 ACS Content Test. Adjustments that were tested included reordering the health insurance types, adding additional instructions, updating the Medicaid question, adding information about the Health Insurance Marketplace, providing additional edit checks in the Computer-Assisted Telephone Interviewing (CATI) and Computer-Assisted Personal Interviewing (CAPI) instruments, and adding a premium and subsidy question. When the ACS added questions on health insurance to the ACS, health insurance marketplaces such as HeathCare.gov did not exist.

Although there was evidence that the tested changes made improvements in several areas (e.g., fewer reports of multiple coverage types, a lower direct-purchase rate, and fewer coverage type write-ins), the changes were not shown to uniformly improve the accuracy of health insurance estimates (Berchick et al., 2017). The primary objective was to increase the overall health insurance coverage rate, thinking that the underreporting of Medicaid and other means-tested programs (as detailed by prior research) would result in the underreporting of any health insurance coverage. However, the revised question did not have a higher insured rate than the control version, and the rates of Medicaid coverage did not differ between versions (Berchick et al., 2017). The decrease in the insured rate was associated with a decrease

in the proportion of respondents that reported having employer-based health insurance in the CAPI response mode (Berchick et al., 2017).

For these reasons, the report did not recommend a change in the Health Insurance Coverage question. Berchick et al. (2017) concluded that some of the question changes should be tested again in the future. The revisions to the health insurance coverage question in this test incorporated some of those previously tested changes.

The primary objectives of revising the health insurance coverage question were:

1. to improve measurement of public coverage and accuracy of direct-purchase coverage,
2. to reduce the reporting of single-service, non-comprehensive insurance plans, and
3. to reduce erroneous reports of multiple coverage.

This revised question would enhance question reliability and validity.

1.3.2. Cognitive Testing Development for Health Insurance Coverage

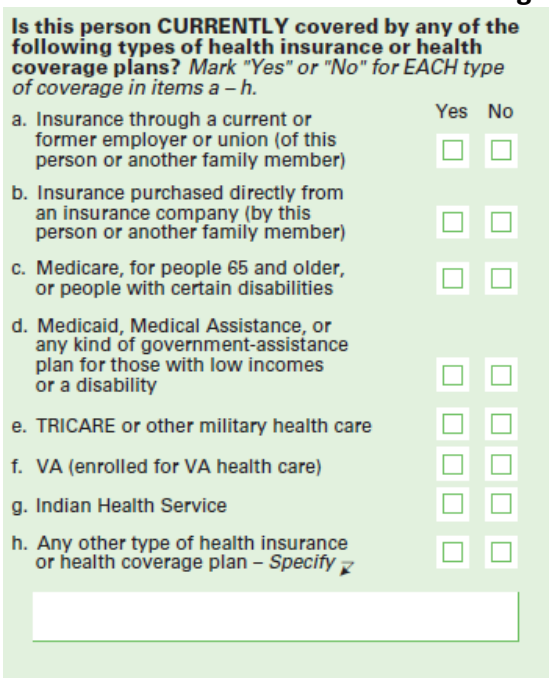
There was one version of the health insurance coverage question in Round 1 of the Content Test cognitive testing. It included some of the changes that continued in the test treatments, such as the instruction on single-service, non-comprehensive coverage plans, the reordered Medicare, Medicaid, and direct purchase options, and the modified wording for employer-based, Medicaid, and direct purchase options. In the Round 1 briefing report, RTI recommended including a "No coverage, Uninsured" option, suggesting that respondents would prefer a response category that allows them to specifically report no coverage (RTI International, 2022a). Round 1 of the cognitive testing also revealed confusion among interviewees regarding military health care (TRICARE) and Veteran's health care (VA) based on interview observations (RTI International, 2022a).

Round 2 of the Content Test cognitive testing, therefore, included two versions of the health insurance coverage question. Version 1 closely matched the version from Round 1, but reordered the TRICARE and VA options and reworded the direct purchase and VA options. Version 2 included the same changes as Version 1, but also included a "No coverage, Uninsured" option and changed the question format to a check-all-that-apply format. RTI reported that both versions performed well during Round 2, with a slight advantage for Version 2 (RTI International, 2022a). However, the results were inconclusive given the small number of interviews, and so we proceeded with both question versions for the field test.

1.3.3. Question Content

Control and test versions of each question are shown as they appear on the paper questionnaire. Automated versions of the questionnaire ask each category as an individual question for the Control and Test Version 1 or have respondents choose coverage options from a list (with a show card in CAPI) for Test Version 2.

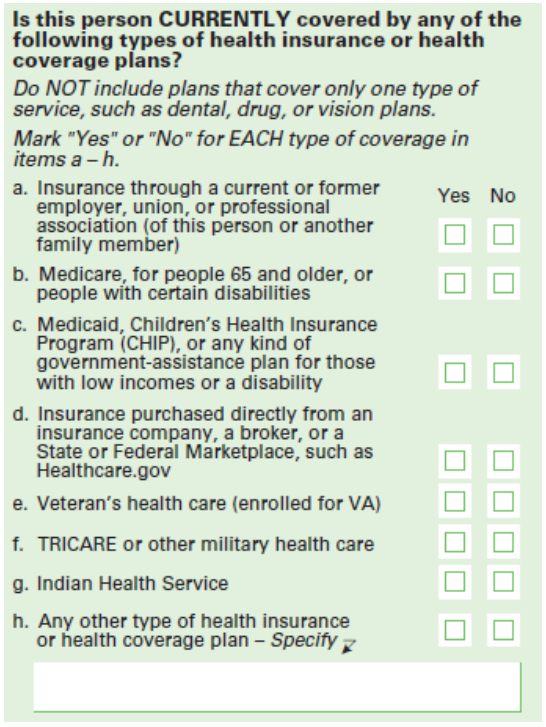
Figure 1. Control Version of the Health Insurance Coverage Question (Paper)



Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans? *Mark "Yes" or "No" for EACH type of coverage in items a – h.*

	Yes	No
a. Insurance through a current or former employer or union (of this person or another family member)	☐	☐
b. Insurance purchased directly from an insurance company (by this person or another family member)	☐	☐
c. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
e. TRICARE or other military health care	☐	☐
f. VA (enrolled for VA health care)	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan – *Specify*	☐	☐

Figure 2. Test Version 1 of the Health Insurance Coverage Question (Paper)



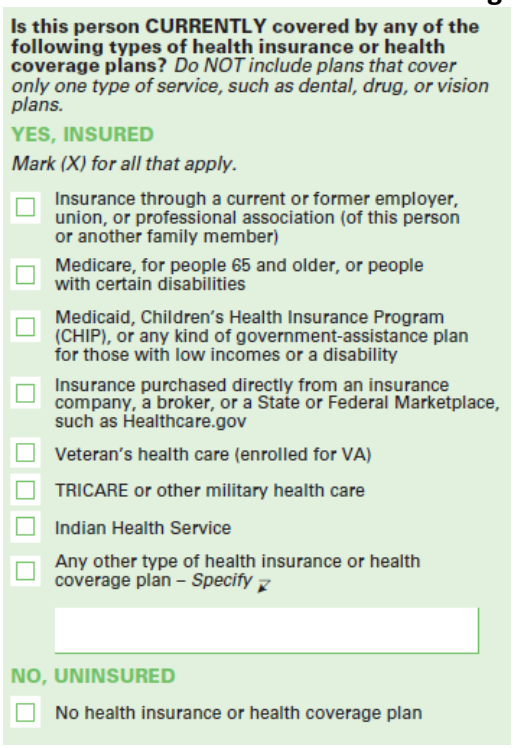
Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?

Do NOT include plans that cover only one type of service, such as dental, drug, or vision plans.

Mark "Yes" or "No" for EACH type of coverage in items a – h.

	Yes	No
a. Insurance through a current or former employer, union, or professional association (of this person or another family member)	☐	☐
b. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
c. Medicaid, Children's Health Insurance Program (CHIP), or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
d. Insurance purchased directly from an insurance company, a broker, or a State or Federal Marketplace, such as Healthcare.gov	☐	☐
e. Veteran's health care (enrolled for VA)	☐	☐
f. TRICARE or other military health care	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan – *Specify*	☐	☐

Figure 3. Test Version 2 of the Health Insurance Coverage Question (Paper)



The changes to the Test versions of the health insurance question include the following improvements:

1. Adding instruction to exclude single-service, non-comprehensive coverage plans

To address the objective of reducing the reporting of single-service, non-comprehensive coverage plans, Test Versions 1 and 2 of the health insurance question added an instruction to help focus respondents' attention on comprehensive coverage (e.g., coverage for hospital and physician services), and, therefore, to reduce overreporting of direct-purchase coverage. The instruction was positioned after the title question but before the health insurance type response choices and reads: *Do NOT include plans that cover only one type of service, such as dental, drug, or vision plans.*

2. Changing the question order of health insurance types

The order of the health insurance types in the Control version (which is the same as current ACS production) is as follows: (a) employer-based, (b) direct purchase, (c) Medicare, (d) Medicaid, (e) TRICARE, (f) VA health care, (g) Indian Health Service, (h) other (write-in).

The order in Test Versions 1 and 2 of the question is as follows: (a) employer-based, (b) Medicare, (c) Medicaid, (d) direct purchase, (e) Veteran's health care (VA), (f) TRICARE, (g) Indian Health Service, (h) other (write-in).

The direct purchase option was moved down the list (from second (b) to fourth (d) position), and Medicare and Medicaid were shifted up one position each. This reordering was designed to reduce over-reporting of direct-purchase insurance (Mach & O'Hara, 2011) and improve reporting of public coverage. The reordering was found to reduce direct-purchase coverage rates in the 2016 ACS Content Test (Berchick et al., 2017).

Round 1 of the Content Test cognitive testing revealed confusion among interviewees regarding military health care (TRICARE) and Veteran's health care (VA) based on interview observations (RTI International, 2022a). In order to improve reporting, TRICARE and Veteran's health care (VA) were reordered on the Test versions in later cognitive testing rounds and for the field test.

3. Rewording response options

A. Key terms were added to the direct purchase language in Test Versions 1 and 2 to improve measurement of health coverage in the current health insurance landscape. The description in the Control version (i.e., ACS production) is:

> *Insurance purchased directly from an insurance company (by this person or another family member)*

The description of direct-purchase coverage in the Test versions now reads:

> *Insurance purchased directly from an insurance company, a broker, or a State or Federal Marketplace, such as HealthCare.gov*

In order to improve measurement of direct-purchase coverage obtained through the Health Insurance Marketplace, "…or a State or Federal Marketplace, such as HealthCare.gov" was added to the direct-purchase coverage question. Since the Patient Protection and Affordable Care Act (ACA) was implemented in 2014, people have been able to buy health insurance through the Health Insurance Marketplace, usually through HealthCare.gov or a state-specific website. Although this coverage is direct purchase, individuals may not know how to report it. This change was intended to reduce reporting of Marketplace plans as Medicaid or as both Medicaid and direct purchase. However, it is worth noting that Medicaid coverage can be obtained through the Health Insurance Marketplace, so some ambiguity could still remain for respondents. Nonetheless, the inclusion of the terms "Marketplace" and "HealthCare.gov" may help orient respondents to identify this coverage as direct-purchase coverage.[3]

[3] Many state portals that provide access to Marketplace coverage also provide access to Medicaid coverage if individuals meet income eligibility and other criteria for these programs. Therefore, it is possible that some respondents may misreport Medicaid coverage accessed through a state portal as Marketplace coverage.

The term "a broker" was also added to the direct purchase question during cognitive testing. Some individuals purchasing coverage directly may purchase their coverage through a broker. Adding the term "a broker" may help orient respondents to identify this coverage as direct-purchase coverage.

Additionally, in order to simplify the question text and to make space for added references to types of direct-purchase health insurance, the phrase "(by this person or another family member)" was removed. Removing this phrase highlights the different types and sources of direct-purchase coverage and may aid respondents in recognizing their coverage as direct purchase.

B. Reference to the Children's Health Insurance Program (CHIP) was added to the Medicaid question in Test Versions 1 and 2 to improve reporting of Medicaid and other means-tested coverage. This addition may aid respondents in recognizing CHIP as a means-tested insurance program, selecting Medicaid coverage, and therefore, potentially reducing the undercount of Medicaid coverage. The term "Medical Assistance" was removed and reference to CHIP was added so that the description of Medicaid coverage now reads:

   Medicaid, Children's Health Insurance Program (CHIP), or any kind of government-assistance plan for those with low incomes or a disability

C. The phrase "or professional association" was added to the employer-based insurance coverage question in Test Versions 1 and 2 to improve reporting of coverage under group plans (e.g., plans sponsored by trade associations or professional groups). The description of employer-based coverage now reads:

   Insurance through a current or former employer, union, or professional association (of this person or another family member)

D. "Veteran's health care" was substituted for "VA" in Test Versions 1 and 2 of the Health Insurance Coverage question. This change may improve reporting of Veteran's health care through the Veterans Health Administration and mitigate respondent confusion of Veteran's health care and other forms of military coverage, such as TRICARE. The description of Veteran's health care now reads:

   Veteran's health care (enrolled for VA)

4. Including a "No, uninsured" option as part of the Test Version 2 health insurance question

In the Round 1 briefing report from cognitive testing for the 2022 ACS Content Test, RTI recommended including a no coverage option, and Round 2 Version 2 of the health insurance question tested this option. The Round 1 briefing report suggested that respondents would prefer a response category that allows them to specifically report no coverage

(RTI International, 2022a). When this option was tested in Version 2 of the health insurance question in Round 2 of cognitive testing, however, the results were inconclusive given the small number of interviews. The Test Version 2 of the health insurance question incorporates a "No, uninsured" option as part of the health insurance question. This change may improve reporting of coverage by enabling respondents to actively select no coverage as an option and mitigate against erroneous reporting of no coverage by including the option as the last item on the insurance type list.

5. Changing the format of Test Version 2 to a check-all-that-apply format

Due to adding the no coverage option, Test Version 2 has a different layout than Test Version 1 and the Control version. The question layout is a check-all-that-apply or check "yes" format instead of individual yes/no questions for each coverage type (letters a-h). This change was expected to be most impactful in the self-response modes: mail and internet. In CAPI, we included a show card with the question, which allowed the respondent to select from the list of health insurance options. In the current production ACS, and in both the Content Test Control and Test Version 1, field representative's read each item separately and ask for a yes/no response (and therefore, no show card is needed).

The CAPI Test Version 2 instrument asks the respondent to select from a list of health insurance options:

> *Using this list, <are you/is NAME> currently covered by any of the following types of coverage…*
>
> - *Show flashcard with insurance types listed. If person says "no" or "none" to all types of coverage, mark "No health insurance or health coverage plans."*
> - *If needed, read response options.*

The CATI Test Version 2 instrument, which was used for phone CAPI interviews and Content Follow-up (CFU) interviews, could not use a show card over the phone, and instead asks the respondent to listen to a list of health insurance options first:

> *I am going to read you a list of different types of health insurance and health coverage. After I finish reading the list, please let me know whether <you are/NAME is> currently covered by any of the following types of coverage…*
>
> - *Read full list before noting the answer. If person says "no" or "none" to all types of coverage, mark "No health insurance or health coverage plans."*

1.3.4. Research Questions

The research questions are divided into several metrics. See Section 2.4.2 for a description of these metrics.

1.3.4.1. Research Questions: Test Version 1 vs. Control

The following list of research questions concern evaluation of Test Version 1 (Test treatment) versus Control of the Health Insurance Coverage question. The primary objective of these research questions was to address known limitations in the health insurance items, as described in Section 1.3.1.

Benchmarks

1. *How do the proportions of persons with any health insurance coverage in the Test treatment and the Control treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the National Health Interview Survey (NHIS)?*

2. *How do the proportions of persons with Medicaid coverage in the Test treatment and the Control treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC)? How do the proportions of persons with direct-purchase coverage compare?*

Item Missing Data Rates

3. *Is the complete item missing data rate different between the Test treatment and the Control treatment?*

4. *Is the proportion of partial responses different between the Test treatment and the Control treatment?*

5. *Are the complete item missing data rate and proportion of partial responses different between the Test treatment and the Control treatment when dividing responses by mode (paper, internet, CAPI)?*

6. *Are the item missing data rates for the Medicare, Medicaid, and direct purchase items different between the Test treatment and the Control treatment?*

7. *Are the item missing data rates for the specific combinations of response different between the Test treatment and the Control treatment? Specifically examining Medicare-Medicaid partial response and Medicaid-Direct Purchase partial response.*

8. *Are the item missing data rates for the TRICARE and VA boxes different between the Test treatment and the Control treatment?*

Response Distributions

9. *Are rates of having any health insurance coverage different between the Test treatment and the Control treatment?*

10. *Are rates of coverage by employer-based insurance, Medicare, Medicaid, direct-purchase insurance, VA, and TRICARE different between the Test treatment and the Control treatment?*

11. *Are the proportions of persons with multiple types of health insurance coverage different between the Test treatment and the Control treatment?*

12. *Are the proportions of persons who reported having both Medicaid and direct-purchase insurance different between the Test treatment and the Control treatment?*

13. *Are the proportions of persons who reported having both Medicare and direct-purchase insurance different between the Test treatment and the Control treatment? Are the proportions of persons who reported having both Medicare and Medicaid insurance different between the Test treatment and the Control treatment?*

14. *Are the proportions of persons who reported having both employer-based and direct-purchase insurance different between the Test treatment and the Control treatment?*

15. *Are the proportions of persons who write-in "other" type of health insurance coverage different between the Test treatment and the Control treatment?*

16. *Do the proportion of write-ins that reference "CHIP" or a state name for the CHIP program, "Marketplace", "HealthCare.gov", "ACA" or other terms associated with the ACA such as "Obamacare" differ between the Test treatment and the Control treatment?*

17. *Do the proportion of write-ins that are out-of-scope differ between the Test treatment and the Control treatment?*

18. *Are the rates and proportions described in questions 9 through 15 different between the Test treatment and the Control treatment when dividing responses by mode (paper, internet, CAPI)?*

19. *Are the above rates and proportions described in questions 9 through 15 different between the Test treatment and the Control treatment when dividing responses by age (under 19, 19-64, 65+) or by state group (i.e., whether or not the state expanded Medicaid eligibility)?*

Response Reliability

20. Are the measures of response reliability (GDR, IOI) different between the Test treatment and the Control treatment, for individual coverage types (items a to f) or no coverage?

21. Are the measures of response reliability (GDR, IOI) different between the Test treatment and the Control treatment within each collection mode (paper, internet, CAPI)?

Other Metrics

22. Is there a difference in help text use on the internet and CAPI instruments between the Test treatment and the Control treatment?

23. Is there a difference in behavior on the internet instrument between the Test treatment and the Control treatment with regards to:

- *a. Switching between answer choices while on the same screen?*
- *b. Going back to previous screens?*
- *c. Time spent on the question screen?*

24. How does the number of persons with Medicaid coverage in each treatment compare with Medicaid enrollment based on administrative records from the Center for Medicaid and Medicare Services?

1.3.4.2. Research Questions: Test Version 1 vs. Test Version 2

The following research questions were formulated to guide the specific analyses to be conducted between Test Version 1 (Test treatment) and Test Version 2 (Roster Test treatment) of the Health Insurance Coverage question.

Benchmarks

1. *How do the proportions of persons with any health insurance coverage in each test treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the National Health Interview Survey (NHIS)?*

2. *How do the proportions of persons with health insurance coverage by Medicaid in each test treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC)? How do the proportions by direct purchase compare?*

Item Missing Data Rates

3. *Is the complete item missing data rate in Test Version 2 different from the rate of complete missing or partial missing in Test Version 1?*

4. *Is the complete item missing data rate in Test Version 2 different from the rate of complete missing or partial missing in Test Version 1 when dividing responses by mode (paper, internet, CAPI)?*

Response Distributions

5. *Are rates of having any health insurance coverage different between Test Version 1 and Test Version 2?*

6. *Are rates of coverage by employer-based insurance, Medicare, Medicaid, direct-purchase insurance, VA, and TRICARE different between Test Version 1 and Test Version 2?*

7. *Are the proportions of persons with multiple types of health insurance coverage different between Test Version 1 and Test Version 2?*

8. *Are the proportions of persons who write-in an "other" type of health insurance coverage different between Test Version 1 and Test Version 2?*

9. *Is the proportion of persons with no health insurance coverage in Test Version 1 different from the proportion of uninsured persons in Test Version 2?*

10. *Are the above rates and proportions different between Test Version 1 and Test Version 2 when dividing responses by mode (paper, internet, CAPI)?*

11. *Are the above rates and proportions different between Test Version 1 and Test Version 2 when dividing responses by age (under 19, 19-64, 65+)?*

Response Reliability

12. *Are the measures of response reliability (GDR, IOI) different between Test Version 1 and Test Version 2, overall and when dividing responses by mode (paper, internet, CAPI)?*

Other Metrics

13. *Is there a difference in help text use on the internet and CAPI instruments between Test Version 1 and Test Version 2?*

14. Is there a difference in behavior on the internet instrument between Test Version 1 and Test Version 2 with regards to:

   a. Switching between answer choices while on the same screen?
   b. Going back to previous screens?
   c. Time spent on the question screen?

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 nearest neighbor method has been used in previous Content Tests; the method's purpose is for sampled housing units to be as similar as possible between treatments since members of housing unit addresses that are geographical "neighbors" are likely to have similar socioeconomic characteristics, and therefore are more likely to respond similarly to the ACS questions. 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 persons 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.[4]

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 is 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.

[4] 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.

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 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.[5]
- 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.[6] 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.[7]

[5] 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.

[6] 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.

[7] 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.

- 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.

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 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.[8]

2.3.2. Content Test Follow-Up for Health Insurance Coverage

For the CFU reinterview, the Health Insurance Coverage question was re-asked to households that completed the original interview. Each treatment path of the CFU instrument had the CATI version of the original interview question. The soft edit checks, though, were removed so that the respondent would not be prompted to change their answer.

For the Control and Test Version 1, the CATI and CAPI versions are almost identical. For Test Version 2, however, the CATI version omits the show card from the CAPI version (because it is impossible to show someone a card over the phone).[9] This omission makes the CFU version different from the original interview mail, internet, and CAPI in-person versions by not visually showing the respondent the full coverage options list. Therefore, it is difficult to determine the extent of the effect of these mode differences on the reliability of Test Version 2.

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 Health Insurance Coverage is attributable to differences in the wording and format. We tested these unit-level assumptions in both the original interview and the CFU interview. See Section 2.4.1 for details. The metrics that we used to evaluate Health Insurance Coverage 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. [10]

[8] 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.

[9] Refer to Section 1.3.3 for differences in CATI and CAPI Test Version 2 instruments.

[10] 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.

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.[11]

Comparisons between the Control and test versions of the Health Insurance Coverage question were conducted using a two-tailed t-test at the α=0.1 level of significance. The Content Test sample size was chosen to 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.[12] 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).

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 treatments in the original interviews and CFU reinterviews. These results are provided in a separate report (Spiers et al., 2023).[13]

2.4.2. Topic-Level Analysis

To evaluate the changes to the Health Insurance Coverage question, we calculated a variety of metrics, presented in Sections 2.4.2.1 through 2.4.2.5.

[11] 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).

[12] See Section 2.4.2.4 for the definition of Gross Difference Rate.

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

2.4.2.1. Benchmarks

For a variety of reasons, health insurance estimates vary across nationally representative surveys and administrative record systems (Berchick et al., 2017). To roughly gauge the accuracy of the responses to the Health Insurance Coverage question, we nominally compared select estimates derived from these data to similar estimates from two external reliable sources (i.e., a benchmark), the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the National Health Interview Survey (NHIS).

The CPS ASEC is the most widely used source of national health insurance estimates released by the Census Bureau, and the NHIS is the principal source of information on the health of the U.S. population released by the National Center for Health Statistics (NCHS). We made general comparisons to the most recent CPS ASEC and NHIS data available.

Comparisons among these surveys are complicated by differences in survey methods, reference periods, and universes. While the ACS and the NHIS measure health insurance coverage at the time of the interview, the CPS ASEC is collected between February and April of each year and captures coverage in the previous calendar year.[14] The ACS is a mixed mode survey that includes significant self-response (by mail and internet) where the other two rely on interviewers to complete the data collection. The context of the surveys varies as well. Where the NHIS is a health-focused survey, the CPS ASEC is primarily employment, income, and health insurance. Both of which contrast with the ACS, which includes a variety of housing, demographic, social, and economic questions in an entirely different manner. These contextual differences could prime respondents in different ways when they answer survey questions. All three surveys have different sampling, editing, and weighting procedures that may lead to different estimates. Finally, there are key differences in question wording that could create differences in estimates of health insurance coverage.

2.4.2.2. Item Missing Data Rates

To measure nonresponse to the Health Insurance Coverage question, we calculated its item missing data rate, the proportion of eligible persons for which a required response is 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.

The Control and Test Version 1 health insurance question included eight yes/no items plus a write-in box for people to explain their coverage (included below the eighth item “other” coverage). Given this structure, there are three classes of responses that were of interest when analyzing the item missing data rate.

[14] The CPS ASEC also includes a measure of coverage reported at the time of interview. However, the CPS ASEC is collected between February and April of each year and the current coverage measure thus reflects health coverage early in the survey year. In contrast, the ACS is collected throughout the year and provides a measure of annual coverage.

1. Complete missing: Person records with no response to any part of the question (completely blank).
2. Partial missing (or partial response): Person records with at least one indication of response and one indication of missing response. Indications of a response included checking a "yes" or "no" box or writing in a valid answer in the "other" field. Indication of a missing response was having at least one item with neither a "yes" or "no" box selected.
3. Complete response: Person records with either "yes" or "no" marked for each coverage type.

Since Test Version 2 is a check-all-that-apply format and not individual yes/no questions, the item missing data rate was defined slightly differently. Complete missing was the same as the other versions, completely blank. Partial missing was not defined for Test Version 2. If a person checked at least one of the nine boxes (eight types of coverage or "no coverage") or entered a valid coverage in the write-in box, the response was considered complete. Therefore, to have an equivalent comparison, we compared the complete item missing data rates in Test Version 2 to both complete and partial missingness in Test Version 1.

We compared item missing data rates via two-tailed t-tests.

2.4.2.3. Response Distributions

To assess how changes to the Health Insurance Coverage question affected the resulting estimates, we compared the response distributions of the Control and Test Version 1 and of Test Version 1 and Test Version 2. We calculated the response distributions as the proportion of valid responses in a category to all valid responses.

We compared response distributions for any health insurance coverage and for each type of health insurance coverage (employer-based, Medicare, Medicaid, direct purchase, Veteran coverage, and TRICARE) between treatments. We tested for significant differences in individual category proportions using two-tailed t-tests.

2.4.2.4. 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. For the 2022 ACS Content Test, response error was measured through response reliability or response bias, not both. This was done to reduce respondent burden and breakoffs during the CFU operation. For Health Insurance Coverage, we measured response error using response reliability.

A survey question has good response reliability if respondents tend to answer the question consistently. For the 2022 ACS Content Test, we measured response reliability for a given question by comparing the responses to this question in the original interview to the responses to this same question in the CFU reinterview.

Re-asking the same question of the same respondent allows us to measure simple response variance, using the following measures:

1. Gross difference rate (GDR)
2. Index of inconsistency (IOI)
3. L-fold index of inconsistency (IOI_L)

The first two measures, GDR and IOI, were calculated for individual response categories. The L-fold index of inconsistency was calculated for questions that had three or more mutually exclusive response categories, as a measure of overall reliability for the question.

In Table 2, "Yes" indicates that the unit is in the category of interest, according to the response from either the original interview or the CFU reinterview. "No" indicates that the unit is not reported to be in the category.

Table 2. Original Interview and CFU Reinterview Counts for Calculating GDR, IOI, and NDR

		Content Test original interview		reinterview totals
		Yes	No	
CFU reinterview	Yes	a	b	a + b
	No	c	d	c + d
original interview totals		a + c	b + d	n

Here, a, b, c, d, and n are counts, defined as follows:

a = units in category for both interview and reinterview
b = units not in category for original interview, but in category for reinterview
c = units in category for original interview, but not in category for reinterview
d = units in category for neither interview nor reinterview
n = total units in the universe = a + b + c + d

These counts were weighted to make them more representative of the population.

We calculated the GDR for this response category as:

$$GDR = \left(\frac{b+c}{n}\right) \times 100$$

To define the IOI, we must first discuss the variance of a category proportion estimate. If we are interested in the true proportion of a total population that is in a certain category, we can use the proportion of a survey sample in that category as an estimate. Under certain reasonable assumptions, it can be shown that the total variance of this proportion estimate is the sum of

two components, sampling variance (SV) and simple response variance (SRV). It can also be shown that an unbiased estimate of SRV is half of the GDR for the category.

The SV is the part of total variance resulting from the differences between all the possible samples of size n one might have selected. SRV is the part of total variance resulting from the aggregation of response error across all sample units. If the responses for all sample units were perfectly consistent, then SRV would be zero, and the total variance would be due entirely to SV. As the name suggests, the IOI is a measure of how much of total variance is due to inconsistency in responses, as measured by SRV. A preliminary definition of the IOI is:

$$IOI = \left(\frac{SRV}{SRV + SV}\right) \times 100$$

We can estimate SRV using the GDR, but also need to estimate the denominator (i.e., total variance) in this expression. Based on previous studies, the estimate we use for total variance is:

$$SRV + SV = \frac{p_1 q_2 + p_2 q_1}{2}$$

where:

$$p_1 = \frac{a + c}{n} = \text{original interview proportion in category}$$

$$q_1 = 1 - p_1 = \frac{b + d}{n} = \text{original interview proportion not in category}$$

$$p_2 = \frac{a + b}{n} = \text{CFU proportion in category}$$

$$q_2 = 1 - p_2 = \frac{c + d}{n} = \text{CFU proportion not in category}$$

In comparing relative reliability (or response error) between treatments, if the response categories are essentially the same, then we looked at the differences in the GDR and IOI for each response category. We tested the significance of these differences, using two-tailed t-tests.

If the response categories did not match up exactly between the compared treatments, we either collapsed response categories to form equivalent categories for comparison, or we conducted comparisons for the response categories where it made sense.

So far, we have only discussed response reliability with respect to single response categories. If a question has three or more response categories (or "comparison categories" in cases where it is necessary to collapse some response categories for comparison), we could also have measured the overall response reliability of a question using the L-fold index of inconsistency, IOI_L. It is possible to look at the difference in IOI_L between treatments and test for significance as with the single category measures. We did not examine IOI_L for this topic.

Suppose a question has L response categories. Let X_{ij} be the weighted count of sample units (households or persons) for which we have CFU responses in category *i* and original interview responses in category *j*. Here, both *i* and *j* range from 1 to L. Table 3 shows a cross-tabulation of the original interview and CFU results for a generic analysis topic. Note that if L = 2, then Table 3 is equivalent to Table 2.

Table 3. Cross-Tab of Original Interview and CFU Results: Questions with Response Categories

		Original Interview categories						
		1	2	…	j	…	L	CFU totals
CFU categories	1	X_{11}	X_{12}	…	X_{1j}	…	X_{1L}	X_{1+}
	2	X_{21}	X_{22}	…	X_{2j}	…	X_{2L}	X_{2+}
	…	…	…	…	…	…	…	…
	i	X_{i1}	X_{i2}	…	X_{ij}	…	…	X_{i+}
	…	…	…	…	…	…	…	…
	L	X_{L1}	X_{L2}	…	X_{Lj}	…	X_{LL}	X_{L+}
Original interview totals		X_{+1}	X_{+2}	…	X_{+j}	…	X_{+L}	$T = \sum_{i=1}^{L} \sum_{j=1}^{L} X_{ij}$

Now define the following proportions:

$$p_{ij} = \frac{X_{ij}}{T}$$

$$p_{+j} = \frac{X_{+j}}{T}$$

$$p_{i+} = \frac{X_{i+}}{T}$$

The IOI_L is calculated as

$$\mathrm{IOI_L} = \frac{1 - \sum_{i=1}^{L} p_{ii}}{1 - \sum_{i=1}^{L} (p_{i+} p_{+i})} \times 100$$

It can be shown that the IOI_L is a weighted sum of the L category IOI values (Biemer, 2011), but this formula is easier for calculation.

The IOI metrics can be biased if the parallel measures assumption is violated, i.e., if the errors in the original interview and CFU reinterview are positively or negatively correlated (Biemer, 2011). We checked this assumption by testing if the net difference rate (NDR) is significantly different from zero. The NDR is the difference between the original interview proportion of positive responses ("Yes" or in the category of interest) and the CFU proportion of positive responses. The NDR is calculated as follows:

$$NDR = (p_1 - p_2) \times 100 = \left(\frac{c-b}{n}\right) \times 100$$

If the NDR is significantly positive or negative, the assumption of "parallel measures" necessary for the SRV and IOI to be valid is not satisfied (Biemer, 2011). In these situations, we use the following adjustment of the IOI, developed by Flanagan (2001):

$$\text{IOI}_{\text{adjusted}} = \frac{\frac{n^2(b+c) - n(c-b)^2}{n-1}}{(a+c)(c+d) + (a+b)(b+d)} \times 100$$

2.4.2.5. Other Metrics for Health Insurance Coverage

For the Health Insurance Coverage analyses, we examined two additional metrics: 1) internet and CAPI paradata and 2) Medicaid enrollment data. These metrics were for informational purposes and were not a part of the decision criteria.

Paradata refers to auxiliary data collected in a survey that describe the data collection process (Beaumont, 2005; Couper, 1998; Couper & Lyberg, 2005; Kreuter et al, 2010; Kreuter, Couper, & Lyberg, 2010). There are multiple types of paradata; each one can be used for different purposes. The 2022 ACS Content Test Respondent Burden Report (Virgile et al., 2023) examined internet and CAPI paradata to determine if any question changes or additions would cause undue burden on future ACS respondents. The analysis included examining median survey completion time and median time spent on particular topics, rates of clicking a "help" access screen within topics, breakoff rates, and form completeness rates.

The research questions in Section 1.3.4 using internet and CAPI paradata include some metrics not covered by the Virgile et al. (2023) report. Additional analysis of answer switching and instrument navigation (backing up) will be documented separately and are not covered in this report.[15]

We also compared estimates from both Test Version 1 and the Control version with Medicaid enrollment data from administrative data released through the Centers for Medicare and Medicaid Services (CMS). Although ACS estimates of the number of people covered by Medicaid and CHIP tend to be lower than enrollment counts from CMS, we still nominally

[15] See research questions 22a and 22b in Section 1.3.4.1 and research questions 14a and 14b in Section 1.3.4.2.

compared the number of people with Medicaid coverage in the Control and Test treatments with enrollment data from CMS.

3. DECISION CRITERIA

Before field testing the Health Insurance Coverage question, a team of subject matter experts identified and prioritized which of the research questions presented in Section 1.3.4 would determine which question version would be recommended for inclusion in the ACS.

As stated in previous sections, the primary objectives of revising the health insurance questions were to improve measurement of public coverage and accuracy of direct-purchase coverage, to reduce overcount of single-service, non-comprehensive insurance plans, and to reduce erroneous reports of multiple coverage. These misreports may arise in large part through respondent confusion in distinguishing coverage type or not recognizing non-comprehensive coverage as such. Therefore, the decision criteria as specified below adjudicates which version best reduces respondent confusion and improves respondent reports of coverage type.

3.1. Decision Criteria for Control vs Test Version 1

The purpose of testing the revised Health Insurance Coverage question was to enhance question reliability and validity. While external research may suggest a particular ACS estimate is too low or too high, the underlying cause of the difference may be related to survey methods differences or an array of non-survey variation. Therefore, preferences for higher or lower rates in Test Version 1 compared with the Control version were not primary decision criteria (see Section 1.3). Rather, these were included in the decision criteria only when the quality measures identified as priority 1, 2, 3, and 4 did not lead to a conclusive decision.

Research Questions 1, 2, 13, 14, 22, 23 and 24 are only for information and were not a factor in the decision.

The most important results of this analysis when drawing a conclusion about Test Version 1 to the Control version are, in order of priority:

Table 4. Decision Criteria for Health Insurance Coverage: Control vs Test Version 1

Priority	Research Question*	Decision Criteria Control vs Test Version 1
1	15, 16, 17	We expect the proportion of write-ins to be lower when the respondent understands the question and/or how to categorize their insurance. In addition, coding write-in responses is time-consuming and costly. The test version incorporates two changes which may be expected to reduce the volume of write-ins: (1) the inclusion of the instruction to exclude single-service, non-comprehensive insurance plans; and (2) the references to "CHIP", "Marketplace" and "HealthCare.gov" in the health insurance question text. Therefore, • A lower proportion of persons who write-in a type of health insurance coverage is preferable. • A lower proportion of write-ins that reference the terms that the new wording attempts to clarify is preferable. • A lower proportion of write-ins that are determined to be out-of-scope is preferable.
2	11	A difference in proportions of multiple coverage could indicate a reduction in possible respondent confusion on how to categorize their coverage. Therefore, a lower rate of multiple coverage is preferable.
3	3-8	A decrease in complete missingness is preferable. Specifically, fewer responses with all parts left blank is preferable. In general, lower item missing data rates are preferred.
4	20, 21	Higher test-retest reliability is preferable. A lower GDR and IOI are preferable.
	9	External evaluations suggest the ACS current coverage rate is still too low. If Test Version 1 and the Control version do not differ on other quality measures, the version which has a lower uninsured rate is preferable.
	10	External evaluations suggest the ACS public coverage rate is too low. If Test Version 1 and the Control version do not differ on other quality measures, then the version which has a higher rate of Medicaid coverage is preferable. It may be unclear if a change in estimates means an improvement in reporting (i.e., reduction in bias), and so these changes alone are not sufficient and must be considered along with the other priorities.

Priority	Research Question*	Decision Criteria Control vs Test Version 1
5	12	Having Medicaid and direct-purchase coverage concurrently is highly unlikely. Thus, the version which has fewer cases where both are selected is preferable. In some states, Medicaid and direct purchase Marketplace plans may be accessed through the same online portal. Therefore, some respondents may be answering correctly from their perspective – it is one plan that is Medicaid and purchased through a state website. This is a situation which can be managed in the editing process. Hence, while we generally prefer fewer reports of both Medicaid and direct purchase, when analyzed in the fuller context, we may accept higher reports of both types.
	18, 19	Preference for higher or lower rates by mode, age, and Medicaid-expansion state status will depend on the coverage type. See the decision criteria for Research Questions 9-12 for more information.

*Research questions within a priority are organized with smallest research question number listed first.
Research questions not included in the decision criteria are for research purposes only.

3.2. Decision Criteria for Test Version 1 vs Test Version 2

We created Test Version 2 to address a specific recommendation from cognitive testing. The Round 1 briefing report suggested that respondents would prefer a response category that allows them to specifically report no coverage (RTI International, 2022a). In Round 2 of cognitive testing, Version 2 of the health insurance question showed promise in self-response (RTI International, 2022a). If Test Version 1 showed improvement over Control (meets the decision criteria in Table 4), then we considered whether Test Version 2 resolved issues identified with the current question format without introducing additional error.

Research Questions 1, 2, 13, and 14 are only for information and were not a factor in the decision.

The most important results of this analysis when drawing a conclusion about Test Version 1 to Test Version 2 were, in order of priority:

Table 5. Decision Criteria for Health Insurance Coverage: Test Version 1 vs Test Version 2

Priority	Research Question*	Decision Criteria Test Version 1 vs Test Version 2
1	3, 4	Filling missing data with logical edits or allocations requires making assumptions and adding uncertainty that are not measured in the final statistics. In general, lower item missing data rates are preferred. However, the check-all-that-apply question format is almost guaranteed to have fewer missing data because any mark is considered a complete response. In order to determine that Test Version 2 is an improvement over Test Version 1, we compare missingness within mode. For all modes but especially self-response modes, a decrease in complete missingness is preferable.
2	12	Ask/Re-ask reliability measures consistency of response. Generally, if people are able to answer the same question with the same answer several weeks apart, this is evidence of a well-understood question. Both Test Version 1 and Test Version 2 use the same wording. If the check-all-that-apply approach in Test Version 2 is successful, the GDR and IOI should be lower than Test Version 1. However, due to the potential CATI mode impact on Test Version 2, we may not be able to draw any conclusions regarding the comparative response reliability of the versions.
3	5, 9	External evaluations suggest the ACS current coverage rate is still too low. Test Version 2 provides respondents the option to select “NO, UNINSURED: No health insurance coverage or plan”. The version which has a lower uninsured rate is preferable.
	6	External evaluations suggest the ACS public coverage rate is too low. The version which has a higher rate of Medicaid coverage is preferable.
	7	A difference in proportions of multiple coverage could indicate a reduction in possible respondent confusion on how to categorize their coverage. The version which has a lower rate of multiple coverage is preferable.
	8	We expect the proportion of write-ins to be lower when the respondent understands the question and/or how to categorize their insurance. In addition, coding write-in responses is time-consuming and costly. Therefore, the version which has a lower proportion of persons who write-in a type of health insurance coverage is preferable.
	10, 11	Preference for higher or lower rates by mode and age will depend on the coverage type.

*Research questions within a priority are organized with smallest research question number listed first.
Research questions not included in the decision criteria are for research purposes only.

4. ASSUMPTIONS AND LIMITATIONS

4.1. Assumptions

- 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 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 were similar between treatments.

4.2. Limitations

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

- 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.

- 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 effect.

- 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 do 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 Health Insurance Coverage. The comparisons presented assume homogeneity of the response distributions for the three treatments, prior to the field test. We tested this assumption via unit-level (i.e., address level) analyses. The results are presented in (Spiers et al., 2023).

For the original interview, 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.

For the CFU reinterview, the response rates overall and by mode of original interview were not significantly different between the Control and Test treatments. However, there were CFU response rate differences between the Test and Roster treatments overall and within some original interview modes. These differences may affect some of the response error analyses for Test Version 1 and Test Version 2 because they involve those treatments. Additionally, there were CFU response distribution differences for tenure and language of response among some treatment and mode comparisons. We think the impact of these characteristic differences are minimal to the Health Insurance Coverage response error analyses.

5.1. Results for Control vs Test Version 1

The following sections give the benchmarks, item missing data rates, response distributions, response reliability, and other metrics results for comparing Control and Test Version 1. The research questions and tables reference the Test treatment, which is the treatment that used Test Version 1 of the Health Insurance Coverage question (Table 1).

5.1.1. Benchmark Results

How do the proportions of persons with any health insurance coverage in the Test treatment and the Control treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the National Health Interview Survey (NHIS)? How do the proportions of persons with Medicaid coverage compare? How do the proportions of persons with direct-purchase coverage compare?

To roughly gauge the accuracy of the responses to the Health Insurance Coverage question, we nominally compared select estimates from each question version to similar estimates from external benchmarks. We compared estimates from the 2022 CPS ASEC and the 2022 NHIS, both widely used sources for national health insurance coverage. The benchmark comparisons were only for information and were not a factor in the decision for choosing a question version.

Estimates of health insurance coverage from the CPS ASEC and the NHIS may be different than those from the ACS due to differences in survey design and focus.[16] The surveys also differ regarding coverage reference period. ACS Content Test estimates measured current coverage at the time of interview (September – November 2022). The 2022 CPS ASEC and 2022 NHIS also measured current coverage at time of interview, but the interview periods were February – April 2022 and October – December 2022, respectively.

Table 6 shows estimates of health insurance coverage from the 2022 ACS Content Test, the 2022 CPS ASEC current coverage, and the 2022 NHIS. As we will discuss in Section 5.1.3, the insured rate was 92.2 percent for the Control version and 91.9 percent for Test Version 1.[17] These estimates were nominally close to the benchmark estimates. According to the 2022 CPS ASEC, 91.5 percent had coverage at the time the survey was conducted. The ACS estimates of the overall insured rate are also similar to the 2022 NHIS Early Release estimates, which indicate that 91.5 percent had health insurance at the time of interview (July – December 2022).

The private coverage rate was 66.0 percent for the Control version and 65.8 percent for Test Version 1; the NHIS rate was 60.6 percent, lower than in the Content Test. For public coverage, the opposite occurred; the NHIS rate was 39.7 percent while the Control and Test Version 1 rates were 36.5 percent and 35.9 percent, respectively. When compared to the CPS ASEC, the public and private coverage estimates were very similar (35.7 and 65.0 percent, respectively).

The employer-based coverage rate in the Control treatment (53.5 percent) was nominally closer to the CPS ASEC estimate (53.2 percent) than that for Test Version 1 (55.2 percent). On the other hand, direct-purchase coverage estimates appeared higher in the Control version (12.5 percent) compared to Test Version 1 (9.2 percent) and the CPS ASEC (10.0 percent); the direct-purchase coverage estimates were significantly different between Control and Test Version 1.

For Medicare coverage, the rate appeared lower in the CPS ASEC (18.4 percent) compared to the Content Test (20.7 percent and 21.1 percent for the Control and Test Version 1, respectively). On the other hand, the Medicaid rate appeared higher in the CPS ASEC (18.8

[16] See Section 2.4.2.1 for more information on differences between the ACS, CPS ASEC, and NHIS.

[17] Unless otherwise noted, Control and Test Version 1 coverage rates were not statistically significantly different. Refer to Section 5.1.3 for more information on treatment comparisons.

percent) compared to the Control (17.1 percent) and Test Version 1 (15.5 percent); the Medicaid rates were significantly different between Control and Test Version 1.

Table 6. Comparison of 2022 ACS Content Test Control and Test Version 1 to External Benchmarks for Health Insurance Coverage

Category	Content Test Control	Content Test Version 1	2022 CPS ASEC Current Coverage	2022 NHIS Early Release
Reference Period	Sept – Nov 2022	Sept – Nov 2022	Feb – April 2022	July – Dec 2022
Any insurance	92.2 (0.3)	91.9 (0.4)	91.5 (0.1)	91.5 (0.4)
Any private	66.0 (0.6)	65.8 (0.5)	65.0 (0.2)	60.6 (0.7)
Any public	36.5 (0.5)	35.9 (0.6)	35.7 (0.2)	39.7 (0.6)
Employer-based	53.5 (0.6)	55.2 (0.6)	53.2 (0.2)	--
Direct purchase	12.5 (0.3)	9.2 (0.3)	10.0 (0.1)	--
Medicare	20.7 (0.4)	21.1 (0.4)	18.4 (0.1)	--
Medicaid	17.1 (0.5)	15.5 (0.4)	18.8 (0.2)	--

Sources: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
U.S. Census Bureau, Current Population Survey, 2022 Annual Social and Economic Supplement.
National Center for Health Statistics, National Health Interview Survey, 2022.
Note: Standard errors are in parentheses.

5.1.2. Item Missing Data Rate Results

Is the complete item missing data rate different between the Test treatment and the Control treatment?

High item nonresponse could indicate that a question lacks clarity, is sensitive, or is too difficult to answer. In general, lower item missing data rates are preferred.

To measure item nonresponse to the Health Insurance Coverage question, we examined several definitions of missingness. First, we examined complete item missingness, which is the proportion of persons who do not respond "yes" or "no" to any part of the question or provide a valid write-in (e.g., leaves the question completely blank). The universe includes all persons with at least one question answered in the detailed person section of the questionnaire who did not break off before the Health Insurance Coverage question.[18] Refer to Section 2.4.2.2 for more information on this metric.

Table 7 shows the complete item missing data rates for Control and Test Version 1, overall and by mode. There were no statistically significant differences in any of the complete item missing data rates between question versions.

[18] The detailed person questions used to determine breakoffs before Health Insurance Coverage were school enrollment, grade level attending, educational attainment, language spoken at home, ancestry, and migration.

Table 7. Complete Item Missing Data Rates for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall	7.4 (0.3)	6.8 (0.3)	0.7 (0.5)	0.33
Self-Response	7.6 (0.3)	7.2 (0.3)	0.4 (0.4)	0.33
Internet	8.3 (0.3)	7.6 (0.3)	0.7 (0.5)	0.33
Mail	3.8 (0.4)	5.1 (0.5)	-1.3 (0.6)	0.20
CAPI	6.9 (0.9)	5.5 (0.7)	1.4 (1.2)	0.33

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Is the proportion of partial responses different between the Test treatment and the Control treatment?

Partial item missingness is the proportion of persons who responded “yes” or “no” to at least one item but did not respond “yes” or “no” to all items in the question. This definition also includes the proportion of persons with a codable write-in but left all other health insurance items blank. For example, if a person selected “yes” to employer-based coverage but left the remaining boxes blank, then that response would count as a partial response. The universe includes all persons who responded “yes” or “no” to at least one item or provided content in the write-in field (i.e., complete item missingness was excluded). Refer to Section 2.4.2.2 for more information on this metric.

Table 8 presents the proportion of partial responses for Control and Test Version 1, overall and by mode. Overall, the proportion of partial responses was significantly lower in Test Version 1 than in Control (38.5 percent vs 40.7 percent, respectively). This difference was driven by the internet mode, where the proportion of partial responses in internet was significantly lower in Test Version 1 than in Control. The difference in internet contributed to the proportions of partial responses in self-response (internet and mail modes combined) being significantly lower in Test Version 1 than in Control. The proportions for mail and CAPI were not significantly different between question versions.[19]

[19] The proportion of partial responses for CAPI are low due to the instrument design of the personal interviews.

Table 8. Proportion of Partial Responses for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall	38.5 (0.5)	40.7 (0.6)	-2.1 (0.8)	0.04*
Self-Response	50.4 (0.6)	53.2 (0.7)	-2.8 (1.0)	0.01*
Internet	49.6 (0.7)	52.9 (0.7)	-3.3 (1.1)	0.01*
Mail	54.3 (1.3)	54.6 (1.2)	-0.4 (1.4)	0.80
CAPI	1.7 (0.3)	1.3 (0.2)	0.4 (0.4)	0.53

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the item missing data rates for the Medicare, Medicaid, and direct purchase items different between the Test treatment and the Control treatment?

Table 9 shows the item missing data rates for the Medicare, Medicaid, and direct purchase items for Control and Test Version 1. The three coverage options were presented in a different order in the Test treatment than the Control treatment (refer to Section 1.3.3).

Both Medicare and Medicaid had lower item missing data rates in Test Version 1 than Control (24.7 percent vs 27.4 percent for Medicare; 30.1 percent vs 32.1 percent for Medicaid, respectively). The rates for direct-purchase coverage were not statistically significantly different between the question versions.

Table 9. Medicare, Medicaid, and Direct Purchase Item Missing Data Rates for Control and Test Version 1

Coverage Type	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Medicare	24.7 (0.6)	27.4 (0.5)	-2.7 (0.8)	<0.01*
Medicaid	30.1 (0.5)	32.1 (0.6)	-2.0 (0.8)	0.03*
Direct purchase	30.4 (0.5)	31.3 (0.6)	-0.9 (0.8)	0.25

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the item missing data rates for the specific combinations of response different between the Test treatment and the Control treatment? Specifically examining Medicare-Medicaid partial response and Medicaid-Direct Purchase partial response.

Table 10 shows the item missing data rates for Medicare combined with Medicaid (i.e., either Medicare or Medicaid boxes had no response) and Medicare combined with direct purchase (i.e., either Medicare or direct purchase boxes had no response) for Control and Test Version 1. Both missingness combinations were significantly different between treatments, with the item

missing data rates being lower in Test Version 1 than Control (33.5 percent vs 35.9 percent for Medicare and Medicaid; 34.0 percent vs 36.5 percent for Medicare and direct purchase, respectively).

Table 10. Item Missing Data Rates for Specific Coverage Combinations for Control and Test Version 1

Coverage Type	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Medicare & Medicaid	33.5 (0.6)	35.9 (0.6)	-2.4 (0.9)	0.01*
Medicare & direct purchase	34.0 (0.6)	36.5 (0.6)	-2.4 (0.9)	0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the item missing data rates for the TRICARE and VA boxes different between the Test treatment and the Control treatment?

Table 11 presents the item missing data rates for TRICARE and VA items for Control and Test Version 1. The TRICARE and VA coverage options were reversed in Test Version 1 compared to Control (see Section 1.3.3). Both item missing data rates were significantly different between treatments, with the rates being lower in Test Version 1 than in Control (32.8 percent vs 35.1 percent for TRICARE; 33.1 percent vs 35.5 percent for VA, respectively).

Table 11. TRICARE and VA Item Missing Data Rates for Control and Test Version 1

Coverage Type	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
TRICARE	32.8 (0.5)	35.1 (0.6)	-2.3 (0.8)	0.01*
VA	33.1 (0.6)	35.5 (0.6)	-2.3 (0.8)	0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

5.1.3. Response Distribution Results

For analysis of response distributions, the results include persons who selected “yes” for a type of health insurance or entered a valid health insurance coverage in the write-in field. Section 2.4.2.3 provides additional information on the calculation of these metrics.

Are rates of having any health insurance coverage different between the Test treatment and the Control treatment?

The question version with a higher insured rate (or a lower uninsured rate) is generally preferable. However, a rate that meets this expectation may still contain error. Therefore, changes in the overall insured rate were not a primary criterion for evaluating the Control and Test treatments. Section 1.3.1 includes additional information on these considerations.

Health insurance coverage estimates are meant to include comprehensive medical and hospital plans. People reporting employer-based coverage, Medicare, Medicaid, direct-purchase coverage, VA, or TRICARE were considered to be insured. As in the production ACS, this analysis excludes persons with only Indian Health Service. Additionally, if someone only filled out the write-in box and that entry is not codable, out-of-scope, or falls under no coverage, the person does not have health insurance coverage.

Table 12 displays the response distributions for any insurance coverage for Control and Test Version 1, overall and by mode. There were no statistically significant differences between versions in the proportions of persons reporting any type of health insurance coverage.

Table 12. Response Distributions for Any Insurance Coverage for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	P-Value
Overall	91.9 (0.4)	92.2 (0.3)	-0.3 (0.4)	0.43
Self-Response	93.8 (0.3)	93.9 (0.3)	-0.1 (0.4)	0.77
Internet	93.8 (0.3)	93.9 (0.3)	-0.1 (0.4)	0.73
Mail	94.3 (0.6)	94.3 (0.7)	0.1 (0.9)	0.95
CAPI	85.7 (0.9)	86.7 (0.8)	-0.9 (1.2)	0.42

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Are rates of coverage by employer-based insurance, Medicare, Medicaid, direct-purchase insurance, VA, and TRICARE different between the Test treatment and the Control treatment?

The health insurance coverage rates in Table 13 includes six different types of insurance coverage: three types of private coverage (employer-based, direct purchase, and TRICARE) and three types of public coverage (Medicare, Medicaid, and VA). Table 13 shows the response distributions individually for these specific health insurance coverage types for Control and Test Version 1. Only direct purchase and Medicaid coverage rates were significantly different between versions, with both rates being significantly lower in Test Version 1 compared to Control (9.2 percent and 12.5 percent for direct-purchase coverage; 15.5 percent and 17.1 percent for Medicaid, respectively). The remainder of Section 5.1.3 explores these results in detail, along with the specific coverage type rates by mode.

Table 13. Response Distributions for Specific Health Insurance Coverage Types for Control and Test Version 1

Coverage Type	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Private	65.8 (0.5)	66.0 (0.6)	-0.2 (0.7)	0.79
Employer-based	55.2 (0.6)	53.5 (0.6)	1.8 (0.8)	0.12
Direct purchase	9.2 (0.3)	12.5 (0.3)	-3.3 (0.4)	<0.01*
TRICARE	2.7 (0.2)	2.8 (0.2)	-0.1 (0.3)	0.79
Public	35.9 (0.6)	36.5 (0.5)	-0.6 (0.7)	0.79
Medicare	21.1 (0.4)	20.7 (0.4)	0.4 (0.5)	0.79
Medicaid	15.5 (0.4)	17.1 (0.5)	-1.6 (0.6)	0.08*
VA	2.4 (0.1)	2.1 (0.1)	0.3 (0.2)	0.66
Any insurance[†]	91.9 (0.4)	92.2 (0.3)	-0.3 (0.4)	0.43

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
[†] Any insurance row repeated from Table 12. Therefore, p-values were not adjusted with those from the other coverages.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Employer-Based Coverage

Employer-based coverage was the first option in both versions of the health insurance question. The one change to the item was the addition of the phrase "or professional association" to Test Version 1 to improve reporting of coverage under group plans (e.g., plans sponsored by trade associations or professional groups). Table 14 shows the response distributions for employer-based coverage for Control and Test Version 1, overall and by mode. There were no statistically significant differences in employer-based coverage rates between question versions.

Table 14. Response Distributions for Employer-Based Health Insurance by Mode for Control and Test Treatments

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall[†]	55.2 (0.6)	53.5 (0.6)	1.8 (0.8)	0.12
Internet	62.1 (0.7)	60.8 (0.6)	1.3 (0.9)	0.30
Mail	45.0 (1.4)	43.4 (1.3)	1.7 (1.8)	0.35
CAPI	43.0 (1.4)	40.2 (1.5)	2.9 (2.0)	0.30

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
[†] Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Direct-Purchase Coverage

The Test treatment included several changes aimed at reducing the overreporting of direct-purchase coverage. First, the instruction added at the beginning of the question ("*Do NOT include plans that cover only one type of service, such as dental, drug, or vision plans."*) was designed to reduce erroneous reports of non-comprehensive plans. Second, reordering the options (moving the direct purchase option from second to fourth position) was also part of the strategy to minimize error in reporting this type. Finally, adding certain phrases, specifically HealthCare.gov and Marketplace, was designed to help people choose different coverage appropriately (refer to Section 1.3.3 for more details). Since prior research suggested that ACS direct-purchase coverage rates were higher than other surveys' rates (Mach & O'Hara, 2011), a lower direct-purchase coverage rate was considered preferable.

Table 15 shows the response distributions for direct-purchase coverage for Control and Test Version 1, overall and by mode. The coverage rates were significantly different between treatments in all modes, with all of them being significantly lower in Test Version 1 compared to Control.

Table 15. Response Distributions for Direct-Purchase Health Insurance by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall[†]	9.2 (0.3)	12.5 (0.3)	-3.3 (0.4)	<0.01*
Internet	9.6 (0.4)	13.1 (0.4)	-3.6 (0.6)	<0.01*
Mail	12.7 (0.7)	17.1 (0.8)	-4.4 (1.0)	<0.01*
CAPI	6.4 (0.6)	8.2 (0.8)	-1.8 (1.1)	0.10*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

[†] Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Medicare

The Medicare item moved from the third position to the second position in the order of types. The text was the same in both versions. Table 16 shows the response distributions for Medicare for Control and Test Version 1, overall and by mode. There were no statistically significant differences in Medicare coverage rates between question versions.

Table 16. Response Distributions for Medicare by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall†	21.1 (0.4)	20.7 (0.4)	0.4 (0.5)	0.79
Internet	18.2 (0.4)	17.5 (0.4)	0.6 (0.6)	0.73
Mail	42.6 (1.5)	41.8 (1.2)	0.8 (1.7)	0.96
CAPI	17.3 (0.9)	17.3 (0.9)	0.1 (1.3)	0.96

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

† Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Medicaid

Test Version 1 included several changes designed to improve the reporting of Medicaid, such as moving the Medicaid option from fourth to third and explicitly mentioning CHIP (refer to Section 1.3.3 for more details). A higher rate of Medicaid coverage was preferable, but only if the treatments did not differ on other quality measures.

Table 17 shows the response distributions for Medicaid for Control and Test Version 1, overall and by mode. The coverage rates were significantly different between treatments overall and in internet, with both rates being significantly lower in Test Version 1 compared to Control. The overall Medicaid rate was 15.5 percent for Test Version 1 and 17.1 percent for Control. For internet, the rate was 11.3 percent and 13.1 percent for Test Version 1 and Control, respectively. There were no statistically significant differences in the coverage rates for the mail and CAPI modes.

Table 17. Response Distributions for Medicaid by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall†	15.5 (0.4)	17.1 (0.5)	-1.6 (0.6)	0.08*
Internet	11.3 (0.4)	13.1 (0.5)	-1.8 (0.7)	0.03*
Mail	14.9 (1.0)	15.3 (0.8)	-0.4 (1.2)	0.73
CAPI	26.8 (1.1)	28.6 (1.3)	-1.8 (1.8)	0.65

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

† Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

TRICARE

We reversed the order of the TRICARE and VA options in Test Version 1 after cognitive testing revealed confusion among some interviewees (RTI International, 2022a).

Table 18 shows the response distributions for TRICARE for Control and Test Version 1, overall and by mode. There were no statistically significant differences in TRICARE coverage rates between versions.

Table 18. Response Distributions for TRICARE by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall†	2.7 (0.2)	2.8 (0.2)	-0.1 (0.3)	0.79
Internet	3.1 (0.2)	3.1 (0.2)	0.1 (0.3)	0.97
Mail	2.5 (0.3)	3.4 (0.4)	-0.9 (0.5)	0.21
CAPI	1.9 (0.4)	1.8 (0.4)	<0.1 (0.6)	0.97

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Veteran's Health Care (VA)

"Veteran's health care" was added to "enrolled for VA" in the Test treatment of the Health Insurance Coverage question. We introduced this change to improve reporting of Veteran's health care through the Veterans Health Administration and mitigate respondent confusion between Veteran's health care and other forms of military coverage, such as TRICARE. Table 19 shows the response distributions for VA coverage for Control and Test Version 1, overall and by mode. There were no statistically significant differences in VA coverage rates between versions.

Table 19. Response Distributions for VA Health Insurance by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall†	2.4 (0.1)	2.1 (0.1)	0.3 (0.2)	0.66
Internet	2.2 (0.1)	1.9 (0.1)	0.2 (0.2)	0.55
Mail	3.4 (0.3)	3.7 (0.4)	-0.3 (0.5)	0.58
CAPI	2.5 (0.4)	1.7 (0.2)	0.8 (0.4)	0.21

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 13. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the proportions of persons with multiple types of health insurance coverage different between the Test treatment and the Control treatment?

One of the primary objectives for revising the Health Insurance Coverage question was to reduce erroneous reports of multiple coverage, since prior research has found that direct-

purchase coverage is overreported, partly due to misreporting single-service, non-comprehensive insurance plans and reporting multiple coverage types for the same plan (Mach & O'Hara, 2011; Lynch et al., 2011; Boudreaux et al., 2014). The treatment with the lower rate of multiple coverage could indicate a reduction in possible respondent confusion on how to categorize their coverage. We defined a person as having multiple coverage if they selected "yes" to more than one type of health coverage except for Indian Health Service (i.e., if they selected any combination of employment-based coverage, Medicare, Medicaid, direct-purchase coverage, TRICARE, VA coverage, or specified an in-scope coverage type or plan).

Table 20 shows the response distributions for persons who reported multiple types of coverage for Control and Test Version 1, overall and by mode. Overall, the proportion of persons with multiple types of coverage was significantly less in Test Version 1 than in Control (14.4 percent and 16.1 percent, respectively). This difference was driven by the self-response modes, where in mail, internet, and mail and internet combined, Test Version 1 had significantly lower proportions than Control. There were no statistically significant differences between versions in the proportions for the CAPI mode.

Table 20. Response Distributions for Persons Who Reported Multiple Types of Health Insurance Coverage for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall	14.4 (0.4)	16.1 (0.3)	-1.7 (0.5)	<0.01*
Self-Response	14.8 (0.4)	17.5 (0.4)	-2.7 (0.5)	<0.01*
Internet	12.5 (0.4)	14.9 (0.4)	-2.4 (0.6)	<0.01*
Mail	26.0 (1.3)	29.6 (1.1)	-3.6 (1.5)	0.03*
CAPI	13.1 (0.8)	11.7 (0.8)	1.4 (1.2)	0.24

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the proportions of persons who reported having both Medicaid and direct-purchase insurance different between the Test treatment and the Control treatment?

Having Medicaid and direct-purchase coverage concurrently is highly unlikely. However, in some states, Medicaid and direct purchase Marketplace plans can be accessed through the same online portal. Respondents choosing both options may be answering correctly from their perspective – it is one plan that is Medicaid and purchased through a state website. In general, the treatment which has fewer cases where both are selected was preferable.

Table 21 shows the response distributions for persons who reported both Medicaid and direct-purchase coverage for Control and Test Version 1, overall and by mode. There were no statistically significant differences in the rates overall. The only significant difference by mode

was in internet, where Test Version 1 had lower rates of both options being selected than Control (0.2 percent vs 0.4 percent, respectively).

Table 21. Response Distributions for Persons Who Reported Both Medicaid and Direct-Purchase Coverage by Mode for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall	0.4 (<0.1)	0.5 (0.1)	-0.1 (0.1)	0.36
Internet	0.2 (<0.1)	0.4 (0.1)	-0.2 (0.1)	0.01*
Mail	0.6 (0.2)	0.8 (0.2)	-0.2 (0.2)	0.36
CAPI	0.7 (0.1)	0.4 (0.1)	0.3 (0.2)	0.36

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the proportions of persons who reported having both Medicare and direct-purchase insurance different between the Test treatment and the Control treatment? Are the proportions of persons who reported having both Medicare and Medicaid insurance different between the Test treatment and the Control treatment? Are the proportions of persons who reported having both employer-based and direct-purchase insurance different between the Test treatment and the Control treatment?

We also examined other coverage combinations that could be erroneous reports by the respondent. Persons over 65 are nearly universally eligible for Medicare coverage, yet many persons may supplement their Medicare coverage with a direct-purchase plan or may purchase a Medicare Advantage plan. We also were concerned about the interaction between the Medicare and Medicaid boxes. Cognitive testing revealed some individuals with dual coverage reported only Medicare or only Medicaid. Test Version 1 included several question changes that may have reduced confusion between the two insurance options, such as moving the direct purchase response option further down the list and including an instruction to exclude single-service, non-comprehensive insurance plans.

Table 22 shows the response distributions for persons who reported both Medicare and direct-purchase coverage, Medicare and Medicaid, and employer-based and direct-purchase coverage for Control and Test Version 1. All of the coverage combination rates were significantly different between versions, with the rates being significantly lower in Test Version 1 (3.7 percent for Medicare and direct-purchase; 1.8 percent for Medicare and Medicaid; and 0.7 percent for employer-based and direct-purchase coverage) compared with Control (4.7 percent, 2.5 percent, and 2.0 percent, respectively).

Table 22. Response Distributions for Persons Who Reported Specific Combinations of Health Insurance Coverage for Control and Test Version 1

Coverage Type	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Medicare and direct purchase	3.7 (0.2)	4.7 (0.2)	-1.0 (0.2)	<0.01*
Medicare and Medicaid	1.8 (0.1)	2.5 (0.1)	-0.7 (0.2)	<0.01*
Employer-based and direct purchase	0.7 (0.1)	2.0 (0.2)	-1.4 (0.2)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the proportions of persons who write-in "other" type of health insurance coverage different between the Test treatment and the Control treatment?

Respondents may write in a health insurance coverage when they misunderstand the question or do not know how to categorize their insurance. Coding write-in responses is time-consuming and costly. Therefore, a lower proportion of persons who write-in a type of health insurance coverage was preferable.

We expected the volume of write-ins to be lower in Test Version 1 due to (1) including the instruction to exclude single-service, non-comprehensive insurance plans and (2) adding references to "CHIP", "Marketplace" and "HealthCare.gov" in the question text. Table 23 shows the response distributions for coverage type write-ins for Control and Test Version 1, overall and by mode. None of the write-in proportions were statistically significantly different between versions.

Table 23. Response Distributions for Health Insurance Coverage Type Write-Ins for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	P-Value
Overall	4.8 (0.2)	4.5 (0.3)	0.3 (0.3)	0.35
Self-Response	4.9 (0.2)	4.6 (0.3)	0.4 (0.3)	0.24
Internet	3.9 (0.2)	3.7 (0.3)	0.2 (0.4)	0.65
Mail	9.9 (0.8)	8.5 (0.7)	1.5 (1.0)	0.16
CAPI	4.4 (0.6)	4.4 (0.7)	<0.1 (0.8)	0.99

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Do the proportions of write-ins that reference "CHIP" or a state name for the CHIP program, "Marketplace", "HealthCare.gov", "ACA" or other terms associated with the ACA such as "Obamacare" differ between the Test treatment and the Control treatment?

We also expected the proportion of write-ins referencing terms added to the health insurance question text to be lower in Test Version 1. Table 24 shows the proportion of codable write-ins that reference terms associated with the ACA or CHIP program for Control and Test Version 1. The Test Version 1 proportion was significantly lower than the Control proportion (4.7 percent vs 8.0 percent, respectively), indicating that the addition of these terms to the question text did reduce write-ins for those coverage types.

Table 24. Proportion of Write-Ins that Reference Terms associated with the ACA or CHIP Program for Control and Test Version 1

	Version 1 Percent	Control Percent	Difference	P-Value
ACA or CHIP write-ins	4.7 (1.2)	8.0 (1.3)	-3.3 (1.8)	0.07*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Do the proportions of write-ins that are out-of-scope differ between the Test treatment and the Control treatment?

A write-in is out-of-scope if it specifies single-service, non-comprehensive insurance plans, such as prescription plans, vision, or dental plans, accidental coverage, or coverage for specific conditions. Some people write in disability insurance or long-term care insurance. These are also out of scope. The instruction to exclude single-service, non-comprehensive insurance plans was designed to reduce the number of write-ins that are out-of-scope for the ACS.

Table 25 shows the response distributions for coverage type write-ins that were out-of-scope for Control and Test Version 1, overall and by mode. None of the out-of-scope write-in proportions were statistically significantly different between versions.

Table 25. Response Distributions for Health Insurance Coverage Type Write-Ins that were Out-of-Scope for Control and Test Version 1

Mode	Version 1 Percent	Control Percent	Difference	P-Value
Overall	27.0 (2.5)	27.8 (2.5)	-0.8 (3.6)	0.83
Self-Response	29.2 (2.2)	28.9 (2.6)	0.4 (3.7)	0.92
Internet	26.8 (2.7)	29.9 (3.0)	-3.1 (4.4)	0.48
Mail	33.9 (4.0)	26.7 (4.7)	7.1 (6.4)	0.26
CAPI	19.2 (6.8)	24.2 (6.2)	-5.0 (9.2)	0.59

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Response Distributions by Age or State Medicaid Expansion Status

The eligibility of some health insurance coverage types varies due to a person's age or whether a person lives in a state that expanded Medicaid coverage. Therefore, it was also important to examine any differences in coverage rates between the treatments among these subgroups.

Table 26 shows the response distributions for any type of coverage, employer-based coverage, direct-purchase coverage, Medicare, and Medicaid by age group for Control and Test Version 1.[20] The age groups align with enrollment eligibility of Medicare and Medicaid; children under the age of 19 are eligible for Medicaid/CHIP coverage and adults aged 65 and over are eligible for Medicare.[21]

The response distributions for any coverage were not statistically significantly different between versions among any age group. However, some of the response distributions for specific coverage types were significantly different by age group.

In Table 13, the overall rate of employer-based coverage was not statistically significantly different between Control and Test Version 1. For adults ages 19 to 64, however, the rate of employer-based coverage was significantly higher in Test Version 1 than in Control (65.5 percent vs 63.3 percent, respectively).

The overall direct-purchase coverage rate for Test Version 1 was significantly lower than that for Control (see Table 13). This result held when examining the rates by age group, where for each age group, the rate of direct-purchase coverage was significantly lower in Test Version 1 than in Control.

[20] We omitted some age groups for employer-based and Medicare because of how rare it is to have persons in those age groups with those coverage types.
[21] Individuals under 65 may be eligible for Medicare coverage if they have specific illnesses or disabilities.

The overall Medicare rate was not statistically significantly different between Control and Test Version 1 (Table 13), but for adults aged 65 and over, the rate of Medicare coverage was significantly higher in Test Version 1 than in Control (90.1 percent vs 87.4 percent, respectively).

The overall Medicaid rate for Test Version 1 was significantly lower than that for Control (Table 13). This result is consistent for adults ages 19 to 64 and adults aged 65 and over, where the rates of Medicaid coverage were significantly lower in Test Version 1 than in Control (12.0 percent vs 14.0 percent for those ages 19 to 64; 7.2 percent vs 9.2 percent for those age 65 and older, respectively). For those under the age of 19, the rate of Medicaid coverage was not statistically significantly different between versions.

Table 26. Response Distributions for Health Insurance Coverage by Age for Control and Test Version 1

Age	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall				
Under 19	94.2 (0.8)	94.4 (0.7)	-0.2 (1.0)	0.81
19-64	88.6 (0.4)	89.3 (0.4)	-0.7 (0.5)	0.40
65+	98.9 (0.2)	98.4 (0.2)	0.5 (0.3)	0.19
Employer-based				
19-64	65.5 (0.6)	63.3 (0.5)	2.2 (0.7)	<0.01*
65+	26.6 (0.9)	25.8 (0.9)	0.8 (1.2)	0.52
Direct purchase				
Under 19	3.1 (0.4)	5.9 (0.5)	-2.8 (0.6)	<0.01*
19-64	8.2 (0.3)	10.8 (0.4)	-2.6 (0.5)	<0.01*
65+	17.8 (0.7)	23.6 (0.7)	-5.8 (1.0)	<0.01*
Medicare				
65+	90.1 (0.5)	87.4 (0.6)	2.7 (0.7)	<0.01*
Medicaid				
Under 19	34.6 (1.2)	34.4 (1.2)	0.2 (1.7)	0.90
19-64	12.0 (0.4)	14.0 (0.4)	-2.0 (0.5)	<0.01*
65+	7.2 (0.5)	9.2 (0.5)	-2.1 (0.7)	0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 27 shows the response distributions for any coverage, direct-purchase coverage, Medicaid, and other coverage by state Medicaid expansion status for Control and Test Version 1. Provisions of the ACA allowed states to expand Medicaid eligibility. When the Content Test was in data collection (September – November 2022), 38 states and the District of Columbia had expanded their Medicaid eligibility (Table 27).

The response distributions for any coverage and other coverage were not statistically significantly different between versions among expansion states nor non-expansion states. The direct-purchase coverage rates were significantly different in both expansion and non-expansion states, with the rates being lower in Test Version 1 than in Control (8.4 percent vs 12.1 percent for expansion states; 11.0 percent vs 13.4 percent for non-expansion states, respectively). Like age group, this result corresponds to the one for the overall direct-purchase coverage rate (Table 13). Only the Medicaid rate in expansions states was significantly different between versions, with Medicaid coverage being lower in Test Version 1 than in Control (16.7 percent vs 18.5 percent, respectively).

Table 27. Response Distributions for Health Insurance Coverage by State Medicaid Expansion Status for Control and Test Version 1

State Expansion Status	Version 1 Percent	Control Percent	Difference	Adjusted P-Value
Overall				
Expansion^	93.2 (0.4)	93.7 (0.4)	-0.5 (0.5)	0.66
Non-expansion	88.7 (0.7)	88.6 (0.5)	0.1 (0.9)	0.96
Direct purchase				
Expansion^	8.4 (0.4)	12.1 (0.4)	-3.7 (0.6)	<0.01*
Non-expansion	11.0 (0.5)	13.4 (0.7)	-2.4 (0.8)	<0.01*
Medicaid				
Expansion^	16.7 (0.6)	18.5 (0.5)	-1.8 (0.8)	0.04*
Non-expansion	12.8 (0.7)	13.9 (0.7)	-1.1 (0.9)	0.23
Other coverage†				
Expansion^	5.0 (0.3)	4.5 (0.3)	0.5 (0.4)	0.31
Non-expansion	4.3 (0.4)	4.7 (0.4)	-0.3 (0.5)	0.54

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ Medicaid expansion states are AK, AR, AZ, CA, CO, CT, DC, DE, HI, IA, ID, IL, IN, KY, LA, MA, MD, ME, MI, MN, MO, MT, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, RI, UT, VA, VT, WA, and WV.

† Other coverage refers to cases where respondents wrote a response in the write-in field.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

5.1.4. Response Reliability Results

We measured response reliability, specifically simple response variance, through re-asking the Health Insurance Coverage question versions on the CFU reinterview. We focused on two measures: GDR and IOI.[22] See Section 2.4.2.4 for more information on these metrics.

[22] All IOI rates presented used the IOI adjusted formula since the parallel measures assumption was violated for some comparisons.

Are the measures of response reliability (GDR, IOI) different between the Test treatment and the Control treatment, for individual coverage types (items a to f) or no coverage?

Table 28 shows the GDRs for individual coverage types and no coverage for Control and Test Version 1. A lower GDR is preferable because it means fewer responses were different choices between the original interview and the CFU reinterview. Only the GDR for direct-purchase coverage was significantly different between the versions, with the percent of inconsistent answers in Test Version 1 being lower than in Control (10.2 percent vs 13.7 percent, respectively). There were no statistically significant differences between treatments for the other coverage types and no coverage.

Table 28. Difference in GDR for Specific Health Insurance Coverage Types between Control and Test Version 1

Coverage Type	Version 1 GDR Percent	Control GDR Percent	Difference	Adjusted P-Value
Employer-based	7.9 (0.5)	8.6 (0.5)	-0.7 (0.6)	0.86
Direct purchase	10.2 (0.5)	13.7 (0.6)	-3.5 (0.9)	<0.01*
Medicare	3.5 (0.3)	4.0 (0.3)	-0.5 (0.4)	0.86
Medicaid	5.5 (0.4)	6.1 (0.5)	-0.6 (0.7)	0.86
TRICARE	0.8 (0.1)	0.9 (0.2)	-0.1 (0.2)	0.86
VA	1.4 (0.1)	1.1 (0.2)	0.3 (0.2)	0.86
No coverage^	3.7 (0.3)	3.6 (0.4)	0.1 (0.5)	0.86

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the GDRs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 29 shows the IOIs for Control and Test Version 1 for the same coverage type comparisons as before. There is a general rule that IOI values less than 20 percent indicate low inconsistency, 20 to 50 percent indicate moderate inconsistency, and over 50 percent indicate high inconsistency (Federal Committee on Statistical Methodology, 2001). For both versions, employer-based, Medicare, and TRICARE were in the low inconsistency category; Medicaid, VA, and no coverage were in the moderate inconsistency category; and direct purchase was in the high inconsistency category.[23] None of the IOIs were statistically significantly different between versions for the specific coverage types and no coverage.

[23] The low, moderate, and high inconsistency categorization was based solely on the IOI percent without any statistical testing.

Table 29. Difference in IOI for Specific Health Insurance Coverage Types between Control and Test Version 1

Coverage Type	Version 1 IOI Percent	Control IOI Percent	Difference	P-Value
Employer-based	16.0 (0.9)	17.1 (1.0)	-1.1 (1.2)	0.36
Direct purchase	50.8 (2.1)	56.0 (2.3)	-5.2 (3.4)	0.12
Medicare	8.9 (0.6)	10.2 (0.8)	-1.3 (1.0)	0.18
Medicaid	22.9 (1.7)	22.1 (1.8)	0.7 (2.6)	0.78
TRICARE	11.1 (2.0)	13.1 (1.9)	-2.1 (2.6)	0.44
VA	22.8 (2.5)	23.7 (3.2)	-0.9 (4.1)	0.83
No coverage^	27.9 (2.4)	31.2 (2.8)	-3.3 (3.8)	0.38

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the IOIs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Are the measures of response reliability (GDR, IOI) different between the Test treatment and the Control treatment within each collection mode (paper, internet, CAPI)?

We also calculated measures of response reliability within each original interview data collection mode (mail, internet, and CAPI). Because the CFU reinterview was conducted by phone, examining GDR and IOI within each mode allows us to see any potential reliability effects between a specific mode and phone.

Table 30. Difference in GDR for Specific Health Insurance Coverage Types between Control and Test Version 1, Mail Mode

Coverage Type	Version 1 GDR Percent	Control GDR Percent	Difference	Adjusted P-Value
Employer-based	7.6 (0.9)	9.1 (1.2)	-1.5 (1.5)	0.50
Direct purchase	13.8 (1.2)	18.9 (1.6)	-5.0 (2.0)	0.08*
Medicare	5.3 (0.8)	6.1 (0.8)	-0.8 (1.2)	0.50
Medicaid	4.7 (0.6)	6.5 (1.1)	-1.8 (1.3)	0.50
TRICARE	1.0 (0.3)	0.7 (0.2)	0.3 (0.4)	0.50
VA	2.5 (0.5)	1.3 (0.3)	1.2 (0.6)	0.20
No coverage^	2.0 (0.4)	4.0 (1.2)	-2.0 (1.2)	0.50

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the GDRs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 31. Difference in IOI for Specific Health Insurance Coverage Types between Control and Test Version 1, Mail Mode

Coverage Type	Version 1 IOI Percent	Control IOI Percent	Difference	Adjusted P-Value
Employer-based	15.6 (1.8)	18.6 (2.3)	-3.1 (3.1)	0.84
Direct purchase	53.1 (3.8)	54.2 (4.1)	-1.1 (5.8)	0.84
Medicare	10.5 (1.5)	11.9 (1.5)	-1.4 (2.4)	0.84
Medicaid	16.9 (2.7)	23.7 (3.7)	-6.8 (4.5)	0.66
TRICARE	14.8 (4.6)	11.4 (3.6)	3.4 (6.0)	0.84
VA	31.8 (5.6)	21.4 (4.6)	10.4 (6.9)	0.66
No coverage^	20.2 (3.8)	37.2 (7.7)	-17.0 (8.5)	0.32

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the IOIs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 30 and Table 31 show the GDRs and IOIs between Control and Test Version 1 for those with a mail response in the original interview, respectively. The results are the same as when examined overall; the GDR for direct-purchase coverage was significantly lower in Test Version 1 than in Control (13.8 percent vs 18.9 percent, respectively), and none of the other GDRs or IOIs were statistically significantly different between versions.

Table 32. Difference in GDR for Specific Health Insurance Coverage Types between Control and Test Version 1, Internet Mode

Coverage Type	Version 1 GDR Percent	Control GDR Percent	Difference	Adjusted P-Value
Employer-based	7.1 (0.5)	8.3 (0.6)	-1.2 (0.7)	0.67
Direct purchase	8.4 (0.5)	12.7 (0.7)	-4.3 (0.9)	<0.01*
Medicare	2.8 (0.3)	3.1 (0.3)	-0.3 (0.4)	0.67
Medicaid	4.3 (0.4)	4.6 (0.6)	-0.3 (0.7)	0.67
TRICARE	0.6 (0.1)	0.9 (0.2)	-0.2 (0.2)	0.67
VA	1.1 (0.1)	1.0 (0.2)	0.1 (0.2)	0.67
No coverage^	3.3 (0.4)	3.5 (0.5)	-0.3 (0.6)	0.67

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the GDRs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 33. Difference in IOI for Specific Health Insurance Coverage Types between Control and Test Version 1, Internet Mode

Coverage Type	Version 1 IOI Percent	Control IOI Percent	Difference	Adjusted P-Value
Employer-based	14.9 (1.1)	17.2 (1.2)	-2.3 (1.5)	0.76
Direct purchase	43.0 (2.4)	55.2 (2.7)	-12.2 (3.7)	0.01*
Medicare	8.2 (0.8)	9.5 (0.8)	-1.4 (1.1)	0.76
Medicaid	22.0 (2.3)	19.5 (2.1)	2.4 (3.2)	0.89
TRICARE	7.9 (1.7)	11.6 (2.3)	-3.7 (2.9)	0.76
VA	20.7 (2.7)	20.9 (3.4)	-0.2 (4.3)	0.96
No coverage^	28.4 (3.1)	34.5 (4.1)	-6.1 (5.3)	0.76

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the IOIs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 32 and Table 33 show the GDRs and IOIs between Control and Test Version 1 for those with an internet response in the original interview, respectively. Similar to the overall rates, only the GDR for direct-purchase coverage was significantly different between versions, with the rate being lower in Test Version 1 than Control (8.4 percent vs 12.7 percent, respectively). Additionally, among original interview internet responses, the IOI for direct-purchase coverage was significantly lower in Test Version 1 than Control (43.0 percent vs 55.2 percent, respectively). For internet, the IOI rate for direct-purchase coverage in Test Version 1 indicates moderate inconsistency while the IOI rate for Control indicates high inconsistency. None of the other GDRs or IOIs were statistically significantly different between treatments.

Table 34. Difference in GDR for Specific Health Insurance Coverage Types between Control and Test Version 1, CAPI Mode

Coverage Type	Version 1 GDR Percent	Control GDR Percent	Difference	Adjusted P-Value
Employer-based	10.6 (1.3)	9.1 (1.0)	1.4 (1.4)	0.99
Direct purchase	12.9 (1.4)	12.8 (1.6)	<0.1 (2.5)	0.99
Medicare	4.2 (0.8)	4.9 (0.8)	-0.7 (1.1)	0.99
Medicaid	9.5 (1.4)	9.7 (1.2)	-0.2 (1.9)	0.99
TRICARE	1.1 (0.4)	1.1 (0.5)	<0.1 (0.7)	0.99
VA	1.5 (0.4)	1.3 (0.5)	0.1 (0.6)	0.99
No coverage^	6.1 (1.0)	3.5 (0.5)	2.6 (1.1)	0.17

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the GDRs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 35. Difference in IOI for Specific Health Insurance Coverage Types between Control and Test Version 1, CAPI Mode

Coverage Type	Version 1 IOI Percent	Control IOI Percent	Difference	P-Value
Employer-based	21.1 (2.7)	18.3 (2.0)	2.8 (2.8)	0.32
Direct purchase	75.2 (5.3)	63.1 (6.9)	12.1 (9.1)	0.18
Medicare	10.9 (2.0)	13.1 (2.0)	-2.1 (2.7)	0.44
Medicaid	29.0 (3.8)	26.9 (3.2)	2.0 (5.1)	0.69
TRICARE	22.2 (10.4)	20.4 (7.9)	1.8 (13.2)	0.89
VA	20.3 (7.0)	36.3 (11.7)	-16.1 (13.3)	0.23
No coverage^	30.2 (4.8)	22.6 (3.1)	7.7 (6.2)	0.22

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

^ No coverage is mutually exclusive of any coverage (i.e., any persons not defined as having any coverage). Therefore, the IOIs for any coverage are the same as those for no coverage, and not included separately in the table.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Table 34 and Table 35 show the GDRs and IOIs between Control and Test Version 1 for those with a CAPI response in the original interview, respectively. None of the GDRs or IOIs for the specific coverage types and no coverage were statistically significantly different between versions.

5.1.5. Other Metrics Results

Is there a difference in help text use on the internet and CAPI instruments between the Test treatment and the Control treatment? Is there a difference in behavior on the internet instrument between the Test treatment and the Control treatment?

Analysis for these research questions were covered by the 2022 Content Test Respondent Burden analysis (Virgile et al., 2023). In addition to examining help text use, we also discuss median completion times in internet and CAPI, breakoff rates in internet, and form completeness rates in all modes. Refer to Virgile et al. (2023) for detailed information on all results discussed. Additional analysis of answer switching and instrument navigation (backing up) will be documented separately and are not covered in this report.

For the entire survey, the median completion time was nominally faster in the Control treatment compared to the Test treatment, with a difference of about 15 seconds in internet and about 45 seconds in CAPI. For only the Health Insurance Coverage question, the median completion times were nominally similar between question versions; they were 44 seconds in Control and Test Version 1 in internet, and about 40 seconds in Control and Test Version 1 in CAPI (Control 37 seconds and Test Version 1 43 seconds).

The other measures of respondent burden using internet paradata were help screen access rates and breakoff rates. For the Health Insurance Coverage question, the help screen access

rates were not significantly different between Control and Test Version 1. For the entire survey, the overall breakoff rates were significantly lower in the Control treatment compared with the Test treatment. However, when looking at breakoffs at the Health Insurance Coverage question, the rates were not significantly different between question versions.

Form completeness rates measured the number of questions on the form that were answered among those that should have been answered, based on Content Test response data. None of the mail form completeness rates were significantly different between treatments. For internet, the form completeness rates were significantly higher in the Control treatment than the Test treatment, both overall and for the Detailed Person section of the questionnaire. For CAPI, overall form completeness rates were significantly higher in the Control treatment than the Test treatment, but there were no significant differences by section. It is unclear how, and if, these mode differences in form completeness between treatments impacted the Health Insurance Coverage analyses.

How does the number of persons with Medicaid coverage in each treatment compare with Medicaid enrollment based on administrative records from the Center for Medicaid and Medicare Services (CMS)?

Table 36 shows the number of persons with Medicaid coverage in Control and Test Version 1. Because the Content Test weights did not include the production weighting adjustments for population controls, we calculated the number of persons with Medicaid coverage using the Medicaid coverage rates (Table 13) and the U.S. population estimate as of July 1, 2021, from the Population and Housing Unit Estimates Special Tabulations Program (U.S. Census Bureau, 2022c). Table 36 also includes Medicaid and CHIP enrollment counts based on administrative records from the CMS. The three time periods (September 2021, October 2021, and November 2021) correspond to when the 2022 Content Test was field tested.

ACS estimates of the number of persons covered by Medicaid and CHIP tend to be lower than enrollment counts from CMS, which corresponds to the adjusted Content Test counts. The number of persons with Medicaid coverage in the Test treatment was nominally less than the number of persons in the Control treatment due to the Medicaid coverage rates being significantly different between treatments (Table 13).

Table 36. Adjusted Number of Persons with Medicaid Coverage in Control and Test Treatments compared with CMS Medicaid Enrollment Counts

Medicaid Coverage	Content Test Version 1	Content Test Control	CMS Sept 2021	CMS Oct 2021	CMS Nov 2021
Number of persons^	51,550,000	56,870,000	84,814,198	85,320,683	85,785,045

Sources: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Centers for Medicare and Medicaid Services (CMS), Medicaid and CHIP Eligibility and Enrollment Performance Indicator Data as of February 23, 2023.
^ Content Test number of persons rounded to four significant digits and adjusted to the U.S. population estimate as of July 1, 2021, from U.S. Census Bureau, Population and Housing Unit Estimates Special Tabulations Program.

5.2. Results for Test Version 1 vs Test Version 2

The following sections give the benchmark, item missing data rate, response distribution, response reliability, and other metrics results for comparing Test Version 1 (included on the Test treatment) and Test Version 2 (included on the Roster Test treatment) (Table 1). The research questions and decision criteria for Test Version 2 are specific to the proposed reformat of the question from a series of yes/no items to a check-all-that-apply approach. The response options are the same on both versions, but Test Version 2 also includes a "no health insurance or health coverage plan" option.

5.2.1. Benchmark Results

How do the proportions of persons with any health insurance coverage in each test treatment compare to the proportions found in the most recent Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the National Health Interview Survey (NHIS)? How do the proportions of persons with Medicaid coverage compare? How do the proportions of persons with direct-purchase coverage compare?

As with the previous benchmark comparisons (Section 5.1.1), we nominally compared select estimates from each test question version to similar estimates from external benchmarks to roughly gauge the accuracy of the responses to the Health Insurance Coverage question. We compared estimates from the 2022 CPS ASEC and the 2022 NHIS, both widely used sources for national health insurance coverage. The benchmark comparisons were only for information and were not a factor in the decision for choosing a question version.

Estimates of health insurance coverage from the CPS ASEC and the NHIS may be different than those from the ACS due to differences in survey design and focus.[24] The surveys also differ regarding coverage reference period. ACS Content Test estimates measured current coverage at the time of interview (September – November 2022). The 2022 CPS ASEC and 2022 NHIS also

[24] See Section 2.4.2.1 for more information on differences between the ACS, CPS ASEC, and NHIS.

measured current coverage at time of interview, but the interview periods were February – April 2022 and October – December 2022, respectively.

Table 37 shows estimates of health insurance coverage from the 2022 ACS Content Test, the 2022 CPS ASEC current coverage, and the 2022 NHIS. As we will discuss in Section 5.2.3, the insured rate was 91.9 percent for Test Version 1 and 92.3 for Test Version 2.[25] These estimates were very close to the benchmark estimates. According to the 2022 CPS ASEC, 91.5 percent had coverage at the time the survey was conducted. The ACS estimates of the overall insured rate are also similar to the 2022 NHIS Early Release estimates, which indicate that 91.5 percent had health insurance at the time of interview (July – December 2022).

The private coverage rate was 65.8 percent for Test Version 1 and 64.2 percent for Test Version 2; and the NHIS rate was 60.6 percent, lower than in the Content Test. For public coverage, the opposite happens; the NHIS rate was 39.7 percent while the Test Version 1 and Test Version 2 rates were 35.9 percent and 35.2 percent, respectively. When compared to the CPS ASEC, the public and private coverage estimates are very similar.

The employer-based coverage rate in Test Version 2 (54.0 percent) appeared closer to the CPS ASEC estimate (53.2 percent) than did that for Test Version 1 (55.2 percent). On the other hand, the direct-purchase coverage rate in Test Version 1 (9.2 percent) appeared closer to the CPS ASEC estimate (10.0 percent) than that from Test Version 2 (8.3 percent).

For Medicare coverage, the rate seemed lower in the CPS ASEC (18.4 percent) compared to the Content Test (21.1 percent and 20.9 percent for the Test Version 1 and Test Version 2, respectively). On the other hand, the Medicaid rate seemed higher in the CPS ASEC (18.8 percent) compared to Test Version 1 (15.5 percent) and Test Version 2 (15.1 percent).

[25] None of the Test Version 1 and Test Version 2 coverage rates in Table 37 were statistically significantly different. Refer to Section 5.2.3 for more information on version comparisons.

Table 37. Comparison of 2022 ACS Content Test Version 1 and Version 2 to External Benchmarks for Health Insurance Coverage

Category	Content Test Version 1	Content Test Version 2	2022 CPS ASEC Current Coverage	2022 NHIS Early Release
Reference Period	Sept – Nov 2022	Sept – Nov 2022	Feb – April 2022	July – Dec 2022
Any insurance	91.9 (0.4)	92.3 (0.3)	91.5 (0.1)	91.5 (0.4)
Any private	65.8 (0.5)	64.2 (0.5)	65.0 (0.2)	60.6 (0.7)
Any public	35.9 (0.6)	35.2 (0.6)	35.7 (0.2)	39.7 (0.6)
Employer-based	55.2 (0.6)	54.0 (0.6)	53.2 (0.2)	--
Direct purchase	9.2 (0.3)	8.3 (0.3)	10.0 (0.1)	--
Medicare	21.1 (0.4)	20.9 (0.4)	18.4 (0.1)	--
Medicaid	15.5 (0.4)	15.1 (0.4)	18.8 (0.2)	--

Sources: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
U.S. Census Bureau, Current Population Survey, 2022 Annual Social and Economic Supplement.
National Center for Health Statistics, National Health Interview Survey, 2022.
Note: Standard errors are in parentheses.

5.2.2. Item Missing Data Rate Results

Is the complete item missing data rate in Test Version 2 different from the rate of complete missing or partial missing in Test Version 1?

As mentioned previously, generally lower item missing data rates are preferable since high item nonresponse could indicate that a question lacks clarity, is sensitive, or is too difficult to answer.

Test Version 1 was a series of yes/no items, and missingness can be differentiated into complete missingness (no marks on any part) or partial missingness (some indication of a response, but not a complete response).[26] Test Version 2, however, was a check-all-that-apply format, and any mark counts as a complete response. Therefore, missingness in this version was a completely blank response. Only respondents who did not select a type, did not mark "No health insurance or health coverage plan," and did not write in a response were considered completely missing. To have an equivalent comparison, we compared both complete missingness and partial missingness in Test Version 1 to complete missingness in Test Version 2. Like Test Version 1, the universe included all persons with at least one question answered in the detailed person section of the questionnaire who did not break off before the Health Insurance Coverage question.

Table 38 shows the complete item missing data rates for Test Version 1 and Test Version 2, overall and by mode. Note the Test Version 1 complete item missing data rates are the same as in Table 7 (Section 5.1.2). In Test Version 2, the proportion of persons who did not check any

[26] Section 5.1.2 describes the missingness metric for Test Version 1 in more detail.

response option of the Health Insurance Coverage question (i.e., left the question completely blank) was 7.3 percent, which was not statistically significantly different from the proportion for Test Version 1 (7.4 percent). Further, there were no statistically significant differences in the complete item missing data rates for any mode between question versions.

Table 38. Complete Item Missing Data Rates for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	P-Value
Overall	7.4 (0.3)	7.3 (0.2)	0.1 (0.4)	0.71
Self-Response	7.6 (0.3)	7.5 (0.3)	<0.1 (0.4)	0.91
Internet	8.3 (0.3)	8.3 (0.3)	0.1 (0.4)	0.90
Mail	3.8 (0.4)	3.9 (0.4)	-0.1 (0.5)	0.87
CAPI	6.9 (0.9)	6.5 (0.7)	0.4 (1.2)	0.72

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Table 39 shows the proportion of partial responses for Test Version 1 and complete missingness for Test Version 2, overall and by mode. The proportion of partial responses in Table 39 differed from those in Table 7 (Section 5.1.2) due to differences in the universes. Table 39 includes respondents with complete item missingness in the rate calculation.

Overall, the proportion of partial responses for Test Version 1 was significantly higher than complete missingness for Test Version 2. This difference was driven by the self-response modes, where the proportions of partial responses for Test Version 1 in internet, mail, and self-response modes (internet and mail combined) were over five times higher than complete missingness for Test Version 2.

The results in Table 38 and Table 39 suggest that complete missingness in Test Version 1 and Test Version 2 are measuring a similar phenomenon, skipping the question entirely. The partial-missing response pattern is unique to asking a series of yes/no questions in a self-response environment. The Test Version 1 proportion for CAPI was not statistically significantly different from Test Version 2 complete missingness, but CAPI partial responses were low in general due to the instrument design of personal interviews.

Table 39. Proportion of Partial Responses for Test Version 1 and Complete Missingness for Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall	43.1 (0.6)	7.3 (0.2)	35.8 (0.6)	<0.01*
Self-Response	54.1 (0.6)	7.5 (0.3)	46.6 (0.7)	<0.01*
Internet	53.8 (0.7)	8.3 (0.3)	45.5 (0.8)	<0.01*
Mail	56.0 (1.3)	3.9 (0.4)	52.1 (1.4)	<0.01*
CAPI	8.5 (0.9)	6.5 (0.7)	2.0 (1.2)	0.10

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

5.2.3. Response Distribution Results

Are rates of having any health insurance coverage different between Test Version 1 and Test Version 2?

External evaluations suggest the ACS current coverage rate is too low (see Section 5.1.3 for more information). Thus, the question version with a higher insured rate (or a lower uninsured rate) was preferable. Test Version 2 included an uninsured option in the question that could have potentially affected any coverage rates. However, changes in the overall insured rate were not a primary criterion for evaluating Test Version 1 and Test Version 2.

Table 40 shows the response distributions for any insurance coverage for Test Version 1 and Test Version 2, overall and by mode. As described in Section 5.1.3, coverage by Indian Health Services alone is not considered comprehensive coverage. There were no statistically significant differences between versions in the proportions of persons reporting having any type of health insurance coverage.

Table 40. Response Distributions for Any Insurance Coverage for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	P-Value
Overall	91.9 (0.4)	92.3 (0.3)	-0.4 (0.4)	0.37
Self-Response	93.8 (0.3)	94.1 (0.2)	-0.3 (0.4)	0.40
Internet	93.8 (0.3)	94.4 (0.3)	-0.7 (0.4)	0.11
Mail	94.3 (0.6)	93.0 (0.8)	1.3 (1.0)	0.19
CAPI	85.7 (0.9)	86.6 (0.8)	-0.9 (1.3)	0.51

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Are rates of coverage by employer-based insurance, Medicare, Medicaid, direct-purchase insurance, VA, and TRICARE different between Test Version 1 and Test Version 2?

The any insurance coverage rate includes six different types of insurance coverage: three types of private coverage (employer-based, direct purchase, and TRICARE) and three types of public coverage (Medicare, Medicaid, and VA). Table 41 displays the response distributions for these specific health insurance coverage types for Test Version 1 and Test Version 2. Only VA coverage rates were significantly different between question versions, with Test Version 2 being significantly lower than Test Version 1 (1.8 percent vs 2.4 percent, respectively). We discuss the results for employer-based, direct purchase, Medicare, and Medicaid coverage in detail in the following sections, along with the specific coverage type rates by mode.

Table 41. Response Distributions for Specific Health Insurance Coverage Types for Test Version 1 and Test Version 2

Coverage Type	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Private	65.8 (0.5)	64.2 (0.5)	1.6 (0.7)	0.11
Employer-based	55.2 (0.6)	54.0 (0.6)	1.3 (0.7)	0.44
Direct purchase	9.2 (0.3)	8.3 (0.3)	0.9 (0.4)	0.21
TRICARE	2.7 (0.2)	2.7 (0.2)	<0.1 (0.2)	0.89
Public	35.9 (0.6)	35.2 (0.6)	0.8 (0.9)	0.89
Medicare	21.1 (0.4)	20.9 (0.4)	0.2 (0.6)	0.89
Medicaid	15.5 (0.4)	15.1 (0.4)	0.4 (0.6)	0.89
VA	2.4 (0.1)	1.8 (0.1)	0.6 (0.2)	0.01*
Any insurance[†]	91.9 (0.4)	92.3 (0.3)	-0.4 (0.4)	0.37

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067

[†] Any insurance row repeated from Table 40. Therefore, p-values were not adjusted with those from the other coverages.

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Employer-Based Coverage

Table 42 shows the response distributions for employer-based coverage for Test Version 1 and Test Version 2, overall and by mode. There were no statistically significant differences in employer-based coverage rates between versions.

Table 42. Response Distributions for Employer-Based Health Insurance by Mode for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall†	55.2 (0.6)	54.0 (0.6)	1.3 (0.7)	0.44
Internet	62.1 (0.7)	61.9 (0.6)	0.2 (0.9)	0.86
Mail	45.0 (1.4)	44.0 (1.4)	1.1 (2.0)	0.86
CAPI	43.0 (1.4)	39.7 (1.4)	3.3 (2.0)	0.28

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 41. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Direct-Purchase Coverage

Table 43 shows the response distributions for direct-purchase coverage for Test Version 1 and Test Version 2, overall and by mode. Among mail responses, Test Version 2 had significantly lower direct-purchase coverage rates than Test Version 1. Previous research suggested that the ACS rates of direct purchase were too high (Mach & O'Hara, 2011), suggesting that Test Version 2 may be an improvement over Test Version 1. However, the overall direct-purchase rates were not statistically significantly different between question versions. Further, Test Version 1 (9.2 percent) was significantly lower than Control (12.5 percent) (Table 13). The difference between Version 1 and Version 2 for mail respondents does not in and of itself suggest improvement. The rates among internet and CAPI responses were not statistically significantly different between question versions.

Table 43. Response Distributions for Direct-Purchase Health Insurance by Mode for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall†	9.2 (0.3)	8.3 (0.3)	0.9 (0.4)	0.21
Internet	9.6 (0.4)	8.6 (0.4)	1.0 (0.5)	0.15
Mail	12.7 (0.7)	10.2 (0.7)	2.5 (1.0)	0.03*
CAPI	6.4 (0.6)	6.6 (0.6)	-0.1 (0.9)	0.87

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 41. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Medicare

Table 44 shows the response distributions for Medicare for Test Version 1 and Test Version 2, overall and by mode. There were no statistically significant differences in Medicare coverage rates between question versions.

Table 44. Response Distributions for Medicare by Mode for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall†	21.1 (0.4)	20.9 (0.4)	0.2 (0.6)	0.89
Internet	18.2 (0.4)	17.8 (0.4)	0.4 (0.6)	0.94
Mail	42.6 (1.5)	43.0 (1.3)	-0.4 (1.9)	0.94
CAPI	17.3 (0.9)	17.3 (0.7)	0.1 (1.1)	0.94

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 41. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Medicaid

Previous research has shown that Medicaid and other means-tested programs are underreported in the ACS (O'Hara, 2010; Boudreaux et al., 2011; Lynch et al., 2011; Boudreaux et al., 2014; Boudreaux et al., 2015). Thus, the version which has a higher rate of Medicaid coverage was preferable.

Table 45 shows the response distributions for Medicaid for Test Version 1 and Test Version 2, overall and by mode. Among mail responses, Test Version 2 had significantly lower Medicaid coverage rates than Test Version 1 (11.4 percent vs 14.9 percent, respectively). The rates among internet and CAPI responses were not statistically significantly different between question versions.

Table 45. Response Distributions for Medicaid by Mode for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall†	15.5 (0.4)	15.1 (0.4)	0.4 (0.6)	0.89
Internet	11.3 (0.4)	10.5 (0.4)	0.8 (0.6)	0.38
Mail	14.9 (1.0)	11.4 (0.7)	3.5 (1.2)	0.01*
CAPI	26.8 (1.1)	28.2 (1.1)	-1.4 (1.6)	0.39

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
† Overall row repeated from Table 41. Therefore, p-values were not adjusted with those from the other modes.
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

After consultation with the interagency committee for health insurance questions, one additional analysis was proposed. In order to explore the differences between Test Version 1 and Test Version 2, the analysis that produced the results shown in Table 45 was reperformed comparing persons who reported Medicaid only (with no other second coverage). When persons with multiple coverage were removed from the analysis, the Medicaid rates for Test Version 2 were no longer significantly different from Test Version 1 overall or by mode.

Are the proportions of persons with multiple types of health insurance coverage different between Test Version 1 and Test Version 2?

The version which has a lower rate of multiple coverage was preferable since lower rates of multiple coverage could indicate a reduction in respondent confusion on how to categorize their coverage.

Table 46 shows the response distributions for persons who reported multiple types of coverage for Test Version 1 and Test Version 2, overall and by mode. Overall, the proportion of persons with multiple types of coverage was significantly lower in Test Version 2 (11.2 percent) than Test Version 1 (14.4 percent). In all modes, the proportions of responses of multiple coverage were significantly lower in Test Version 2 than Test Version 1. These differences suggest that the check-all-that-apply format of Test Version 2 contributed to lower rates of multiple coverage compared with the individual yes/no questions format.

Table 46. Response Distributions for Persons Who Reported Multiple Types of Health Insurance Coverage for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall	14.4 (0.4)	11.2 (0.3)	3.2 (0.5)	<0.01*
Self-Response	14.8 (0.4)	11.9 (0.3)	2.8 (0.5)	<0.01*
Internet	12.5 (0.4)	9.8 (0.3)	2.7 (0.5)	<0.01*
Mail	26.0 (1.3)	22.2 (0.9)	3.8 (1.5)	0.01*
CAPI	13.1 (0.8)	8.8 (0.7)	4.3 (1.1)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Are the proportions of persons who write-in an "other" type of health insurance coverage different between Test Version 1 and Test Version 2?

We assume that respondents write in a health insurance coverage when they misunderstand the question or do not know how to categorize their insurance. Additionally, coding write-in responses is time-consuming and costly. The version which has a lower proportion of persons who write-in a type of health insurance coverage is generally preferable.

Table 47 shows the response distributions for coverage type write-ins for Test Version 1 and Test Version 2, overall and by mode. Overall, the write-in proportion was significantly lower in Test Version 2 than Test Version 1 (4.0 percent vs 4.8 percent, respectively). This difference was significant in self-response (internet and mail combined), with the write-in proportion in self-response being significantly lower in Test Version 2 than Test Version 1 (4.3 percent vs 4.9 percent, respectively). The differences were not statistically significant for the modes individually (internet, mail, or CAPI).

Table 47. Response Distributions for Health Insurance Coverage Type Write-Ins for Test Version 1 and Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall	4.8 (0.2)	4.0 (0.2)	0.8 (0.2)	<0.01*
Self-Response	4.9 (0.2)	4.3 (0.2)	0.7 (0.3)	0.04*
Internet	3.9 (0.2)	3.4 (0.2)	0.5 (0.3)	0.11
Mail	9.9 (0.8)	8.6 (0.7)	1.4 (1.0)	0.18
CAPI	4.4 (0.6)	3.1 (0.5)	1.3 (0.7)	0.11

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Is the proportion of persons with no health insurance coverage in Test Version 1 different from the proportion of uninsured persons in Test Version 2?

Test Version 2 contains an explicit "no health insurance or health coverage plan" option. By contrast, Test Version 1 does not have an explicit no coverage option. Rather, Test Version 1 has a series of yes/no questions and as described in Section 5.2.2, a complicated partial response answer pattern in the self-response modes. Because of this uncertainty in respondents' behavior, we examined two different definitions of uninsured in Test Version 1:

- Definition 1 – any person who selected "no" to all items a-f **and** the write-in box was "no" or the write-in text coded to be out-of-scope.

- Definition 2 – any person who had a partial or complete response and had no indication of coverage (did not select "yes" to any items a-f **and** the write-in field was either blank or the write-in text was coded to be out-of-scope).

We defined uninsured in Test Version 2 as any person who met at least one of the following: (a) they selected the no coverage option, (b) they selected IHS coverage and no other coverage, including not providing a write-in, or (c) the only indication of coverage was a write-in that was determined to be out-of-scope. The universe for all definitions of uninsured for Test Version 1 and Test Version 2 excluded persons who left the entire question blank.

Table 48 displays the response distributions for uninsured Definition 1 in Test Version 1 and uninsured in Test Version 2, overall and by mode. Note that this definition of "uninsured" is not the inverse of "any coverage" in previous tables. The rates were not statistically different overall between question versions. In self-response (internet and mail combined), the percentage of persons uninsured was higher in Test Version 2 (6.0 percent) than in Test Version 1 (5.0 percent), driven by the differences within the mail response mode, where Test Version 2 had 7.3 percent uninsured compared to 4.6 percent for Test Version 1.

Table 48. Response Distributions for Uninsured Persons (Definition 1) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall	7.0 (0.3)	7.8 (0.2)	-0.8 (0.4)	0.16
Self-Response	5.0 (0.3)	6.0 (0.2)	-1.0 (0.4)	0.03*
Internet	5.1 (0.3)	5.7 (0.3)	-0.6 (0.4)	0.27
Mail	4.6 (0.6)	7.3 (0.8)	-2.7 (1.0)	0.03*
CAPI	13.2 (0.9)	13.1 (0.8)	0.2 (1.3)	0.90

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 49 presents the response distributions for uninsured Definition 2 in Test Version 1 and uninsured in Test Version 2, overall and by mode. None of the proportions were significantly different between Test Version 1 and Test Version 2. These results suggest that some respondents who select no to only some options might select a no coverage option when offered.

Table 49. Response Distributions for Uninsured Persons (Definition 2) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 Percent	Version 2 Percent	Difference	P-Value
Overall	8.2 (0.4)	7.8 (0.2)	0.4 (0.4)	0.31
Self-Response	6.3 (0.3)	6.0 (0.2)	0.3 (0.4)	0.45
Internet	6.3 (0.3)	5.7 (0.3)	0.5 (0.4)	0.19
Mail	6.3 (0.7)	7.3 (0.8)	-1.0 (1.1)	0.35
CAPI	14.2 (0.9)	13.1 (0.8)	1.1 (1.3)	0.38

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Are the above rates and proportions different between Test Version 1 and Test Version 2 when dividing responses by age (under 19, 19-64, 65+)?

The eligibility of some health insurance coverage types varies due to a person's age or whether a person lives in a state that expanded Medicaid coverage. Therefore, it was also important to examine any differences in coverage rates between the treatments among these subgroups.

Table 50 shows the response distributions for any coverage, employer-based coverage, direct-purchase coverage, Medicare, and Medicaid by age group for Test Version 1 and Test Version 2.[27] As with the Control and Test Version 1 comparison in Section 5.1.3, the age groups align with enrollment eligibility of Medicare and Medicaid.

Table 50. Response Distributions for Health Insurance Coverage by Age for Test Version 1 and Test Version 2

Age	Version 1 Percent	Version 2 Percent	Difference	Adjusted P-Value
Overall				
Under 19	94.2 (0.8)	95.6 (0.4)	-1.4 (0.9)	0.35
19-64	88.6 (0.4)	88.7 (0.4)	-0.1 (0.6)	0.86
65+	98.9 (0.2)	98.8 (0.2)	<0.1 (0.3)	0.86
Employer-based				
19-64	65.5 (0.6)	64.6 (0.6)	0.8 (0.8)	0.28
65+	26.6 (0.9)	24.2 (0.7)	2.3 (1.0)	0.04*
Direct purchase				
Under 19	3.1 (0.4)	3.6 (0.4)	-0.4 (0.6)	0.51
19-64	8.2 (0.3)	7.9 (0.4)	0.3 (0.5)	0.51
65+	17.8 (0.7)	14.0 (0.5)	3.9 (1.0)	<0.01*
Medicare				
65+	90.1 (0.5)	88.4 (0.6)	1.6 (0.8)	0.03*
Medicaid				
Under 19	34.6 (1.2)	34.4 (1.3)	0.2 (1.7)	0.93
19-64	12.0 (0.4)	11.6 (0.4)	0.4 (0.6)	0.93
65+	7.2 (0.5)	6.2 (0.4)	0.9 (0.6)	0.46

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

The response distributions for any coverage were not statistically significantly different between treatments among any age group. While the overall rates of employer-based coverage, direct-purchase coverage, Medicare, and Medicaid were not significantly different between Test Version 1 and Test Version 2 in Table 41, some of the response distributions for specific coverage types were significantly different by age group.

[27] We omitted some age groups for employer-based and Medicare because how rare it is to have persons in those age groups with those coverage types.

For adults aged 65 and over, the rates of employer-based coverage, direct-purchase coverage, and Medicare were significantly lower in Test Version 2 than Test Version 1. These differences could indicate that this age group reacted differently to the question versions. While lower rates of direct-purchase and employer-based coverage for this population may suggest that Test Version 2 is an improvement over Test Version 1, lower rates of Medicare coverage do not support the same. Benchmark CPS ASEC data suggest that about 94.0 percent of the population 65 and over is covered by Medicare.[28] While higher reported Medicare rates among this age group are not necessarily preferred, the fact that Test Version 2 nominally differs from benchmarks more than Test Version 1 suggests that Medicare reporting among adults aged 65 and over may be more accurate in Test Version 1. The rates of Medicaid coverage, though, were not significantly different among any age group.

5.2.4. Response Reliability Results

As in Section 5.1.4, we measured response reliability through re-asking the Health Insurance Coverage question versions on the CFU reinterview, focusing on two measures: GDR and IOI.[29] See Section 2.4.2.4 for more information on these metrics.

Since Test Version 1 was a series of yes/no questions, there were few format differences in the question version between the various modes. Test Version 2, however, was a check-all-that-apply format that was delivered differently over the phone compared to the other modes; mail, internet, and CAPI allow for a visual element that is not possible when speaking over the phone. Therefore, the response reliability measures for Test Version 2 may be impacted by the mode differences between the original interview and the CFU reinterview.

Are the measures of response reliability (GDR, IOI) different between Test Version 1 and Test Version 2, overall and when dividing responses by mode (paper, internet, CAPI)?

Table 51 shows the GDRs for specific health insurance coverage types for Test Version 1 and Test Version 2. A lower GDR is preferable because it means fewer responses were different between the original interview and the CFU reinterview. There were no significant differences between question versions for any of the coverage types.

[28] Source: U.S. Census Bureau, Current Population Survey, 2022 Annual Social and Economic Supplements (CPS ASEC)

[29] All IOI rates presented used the IOI adjusted formula since the parallel measures assumption was violated for some comparisons.

Table 51. Difference in GDR for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2

Coverage Type	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Employer-based	7.9 (0.5)	6.9 (0.5)	1.0 (0.6)	0.51
Direct purchase	10.2 (0.5)	9.3 (0.5)	0.9 (0.7)	0.78
Medicare	3.5 (0.3)	3.9 (0.3)	-0.4 (0.4)	0.78
Medicaid	5.5 (0.4)	5.0 (0.4)	0.5 (0.6)	0.78
TRICARE	0.8 (0.1)	0.7 (0.2)	0.1 (0.2)	0.78
VA	1.4 (0.1)	1.3 (0.2)	0.1 (0.2)	0.78

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 52 shows the IOIs for Test Version 1 and Test Version 2 for the same coverage type comparisons as before. Again, the general rule is that IOI values less than 20 percent indicate low inconsistency, 20 to 50 percent indicate moderate inconsistency, and over 50 percent indicate high inconsistency (Federal Committee on Statistical Methodology, 2001). For both versions, employer-based, Medicare, and TRICARE were in the low inconsistency category, and Medicaid and VA were in the moderate inconsistency category; direct purchase was on the edge of categories, being in the high inconsistency category for Test Version 1 and the moderate category for Test Version 2.[30] However, none of the IOIs were statistically significantly different between question versions for the specific coverage types.

Table 52. Difference in IOI for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2

Coverage Type	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Employer-based	16.0 (0.9)	13.9 (0.9)	2.1 (1.2)	0.36
Direct purchase	50.8 (2.1)	49.3 (2.1)	1.5 (3.0)	0.62
Medicare	8.9 (0.6)	10.2 (0.7)	-1.3 (0.9)	0.61
Medicaid	22.9 (1.7)	20.5 (1.6)	2.4 (2.4)	0.62
TRICARE	11.1 (2.0)	12.5 (2.5)	-1.5 (3.0)	0.62
VA	22.8 (2.5)	30.9 (3.2)	-8.1 (3.9)	0.23

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

[30] The low, moderate, and high inconsistency categorization was based solely on the IOI percent without any statistical testing.

We also calculated measures of response reliability within each original interview data collection mode (mail, internet, and CAPI) to see if there were any potential reliability effects between a specific mode and phone.

Table 53 and Table 54 show the GDRs and IOIs for specific coverage types between Test Version 1 and Test Version 2 for those with a mail response in the original interview, respectively. None of the GDRs or IOIs were statistically significantly different between question versions.

Table 53. Difference in GDR for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, Mail Mode

Coverage Type	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Employer-based	7.6 (0.9)	7.2 (0.7)	0.4 (1.1)	0.80
Direct purchase	13.8 (1.2)	13.1 (1.3)	0.7 (1.8)	0.80
Medicare	5.3 (0.8)	6.4 (0.9)	-1.2 (1.3)	0.80
Medicaid	4.7 (0.6)	4.9 (0.6)	-0.2 (0.7)	0.80
TRICARE	1.0 (0.3)	0.8 (0.2)	0.3 (0.4)	0.80
VA	2.5 (0.5)	1.3 (0.3)	1.2 (0.6)	0.32

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 54. Difference in IOI for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, Mail Mode

Coverage Type	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Employer-based	15.6 (1.8)	14.5 (1.4)	1.0 (2.2)	0.94
Direct purchase	53.1 (3.8)	56.8 (3.9)	-3.7 (5.4)	0.94
Medicare	10.5 (1.5)	12.9 (1.7)	-2.4 (2.6)	0.94
Medicaid	16.9 (2.7)	25.0 (3.1)	-8.1 (3.8)	0.22
TRICARE	14.8 (4.6)	15.3 (4.6)	-0.5 (6.6)	0.94
VA	31.8 (5.6)	27.2 (5.9)	4.6 (8.4)	0.94

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 55 and Table 56 show the GDRs and IOIs for specific coverage types between Test Version 1 and Test Version 2 for those with an internet response in the original interview, respectively. None of the GDRs or IOIs were significantly different between question versions.

Table 55. Difference in GDR for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, Internet Mode

Coverage Type	Version 1 GDR Percent	Version 2 GDR Percent	Difference	P-Value
Employer-based	7.1 (0.5)	6.5 (0.6)	0.6 (0.7)	0.39
Direct purchase	8.4 (0.5)	8.2 (0.6)	0.2 (0.7)	0.80
Medicare	2.8 (0.3)	2.6 (0.3)	0.2 (0.4)	0.63
Medicaid	4.3 (0.4)	3.3 (0.5)	0.9 (0.6)	0.11
TRICARE	0.6 (0.1)	0.5 (0.1)	0.2 (0.2)	0.43
VA	1.1 (0.1)	1.2 (0.2)	-0.1 (0.3)	0.75

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Table 56. Difference in IOI for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, Internet Mode

Coverage Type	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Employer-based	14.9 (1.1)	13.4 (1.1)	1.5 (1.5)	0.91
Direct purchase	43.0 (2.4)	43.3 (2.6)	-0.3 (3.0)	0.91
Medicare	8.2 (0.8)	7.8 (0.9)	0.3 (1.2)	0.91
Medicaid	22.0 (2.3)	16.1 (2.1)	5.9 (3.2)	0.37
TRICARE	7.9 (1.7)	7.2 (1.9)	0.7 (2.6)	0.91
VA	20.7 (2.7)	26.4 (4.0)	-5.7 (5.1)	0.91

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 57 and Table 58 show the GDRs and IOIs for specific coverage types between Test Version 1 and Test Version 2 for those with a CAPI response in the original interview, respectively. None of the GDRs were significantly different between question versions. For IOI, only the index for VA coverage was significantly different between versions, with the IOI being higher for Version 2 than Version 1 (49.5 percent vs 20.3 percent, respectively). However, both index values were still in the moderate consistency category, and this result did not appear among any other modes.

Table 57. Difference in GDR for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, CAPI Mode

Coverage Type	Version 1 GDR Percent	Version 2 GDR Percent	Difference	P-Value
Employer-based	10.6 (1.3)	7.9 (1.2)	2.7 (1.6)	0.10
Direct purchase	12.9 (1.4)	9.5 (1.5)	3.4 (2.1)	0.10
Medicare	4.2 (0.8)	5.4 (0.7)	-1.2 (1.0)	0.25
Medicaid	9.5 (1.4)	9.2 (1.2)	0.3 (1.9)	0.88
TRICARE	1.1 (0.4)	1.4 (0.5)	-0.3 (0.7)	0.67
VA	1.5 (0.4)	1.8 (0.5)	-0.3 (0.6)	0.59

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result.

Table 58. Difference in IOI for Specific Health Insurance Coverage Types between Test Version 1 and Test Version 2, CAPI Mode

Coverage Type	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Employer-based	21.1 (2.7)	15.9 (2.4)	5.2 (3.4)	0.36
Direct purchase	75.2 (5.3)	59.8 (6.5)	15.4 (8.6)	0.36
Medicare	10.9 (2.0)	15.6 (2.1)	-4.7 (2.9)	0.36
Medicaid	29.0 (3.8)	26.5 (3.3)	2.5 (5.3)	0.64
TRICARE	22.2 (10.4)	32.0 (13.8)	-9.8 (15.5)	0.64
VA	20.3 (7.0)	49.5 (8.6)	-29.2 (9.5)	0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

We also calculated measures of response reliability for uninsured persons. The definitions are the same as those in Table 48 and Table 49 in Section 5.2.3. As with the specific coverage types, we calculated measures overall and within each original interview data collection mode (mail, internet, and CAPI).

Table 59. Difference in GDR for Uninsured Persons (Definition 1) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Overall	3.4 (0.3)	2.5 (0.3)	0.9 (0.4)	0.09*
Internet	3.2 (0.3)	2.2 (0.4)	1.0 (0.5)	0.09*
Mail	2.4 (0.5)	1.6 (0.4)	0.8 (0.7)	0.45
CAPI	4.5 (0.8)	3.9 (0.5)	0.6 (1.0)	0.51

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 60. Difference in IOI for Uninsured Persons (Definition 1) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Overall	28.8 (2.5)	20.5 (2.2)	8.2 (3.1)	0.03*
Internet	31.9 (3.6)	22.4 (3.1)	9.5 (4.9)	0.11
Mail	27.6 (5.9)	13.0 (4.2)	14.6 (7.4)	0.11
CAPI	24.7 (4.5)	21.6 (3.7)	3.1 (5.8)	0.59

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 59 and Table 60 show the GDRs and IOIs for uninsured Definition 1 in Test Version 1 and uninsured for Test Version 2, respectively. For GDR, the overall percent of inconsistent answers in Test Version 2 was significantly lower than that in Test Version 1 (2.5 percent vs 3.4 percent, respectively). This result appeared in the internet mode, where the GDR for Test Version 2 was significantly lower than that in Test Version 1 (2.2 percent vs 3.2 percent, respectively), but not in the other original interview modes. For IOI, the overall index was significantly lower in Test Version 2 than Test Version 1 (20.5 percent vs 28.8 percent, respectively). However, across the original interview modes, none of the IOIs were significantly different between versions.

Table 61. Difference in GDR for Uninsured Persons (Definition 2) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Overall	3.7 (0.3)	2.5 (0.3)	1.2 (0.4)	0.03*
Internet	3.2 (0.4)	2.2 (0.4)	1.0 (0.5)	0.13
Mail	2.1 (0.4)	1.6 (0.4)	0.6 (0.6)	0.29
CAPI	6.0 (1.0)	3.9 (0.5)	2.1 (1.1)	0.13

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 62. Difference in IOI for Uninsured Persons (Definition 2) for Test Version 1 and Uninsured Persons for Test Version 2

Mode	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Overall	27.8 (2.4)	20.5 (2.2)	7.2 (3.2)	0.09*
Internet	28.3 (3.1)	22.4 (3.1)	5.9 (4.7)	0.21
Mail	20.6 (4.1)	13.0 (4.2)	7.5 (5.9)	0.21
CAPI	29.9 (4.8)	21.6 (3.7)	8.3 (6.1)	0.21

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 61 and Table 62 show the GDRs and IOIs for uninsured Definition 2 in Test Version 1 and uninsured for Test Version 2, respectively. Generally, the results were similar to those for uninsured Definition 1. For GDR, the overall percent of inconsistent answers in Test Version 2 was significantly lower than that in Test Version 1 (2.5 percent vs 3.7 percent, respectively). For IOI, the overall index was significantly lower in Test Version 2 than Test Version 1 (20.5 percent vs 27.8 percent, respectively). However, unlike with Definition 1, across the original interview modes, none of the GDRs or IOIs were significantly different between question versions.

Lastly, we calculated measures of response reliability for multiple coverage. We defined multiple coverage for the purpose of reliability metrics using two definitions:

- Definition 1 – Persons who selected multiple coverage types in the original interview and CFU reinterview (but not necessarily the *same* coverage types).
- Definition 2 – Persons who selected *the same* multiple coverage types in the original interview and CFU reinterview.

Again, we calculated measures overall and within each original interview data collection mode (mail, internet, and CAPI).

Table 63 and Table 64 show the GDRs and IOIs for multiple coverage Definition 1 in Test Version 1 and Test Version 2, respectively. For GDR, the overall percent of inconsistent answers was not significantly different between versions. Across original interview modes, only the GDR for internet was significantly different between versions, with the GDR being lower in Test Version 2 than Test Version 1 (9.5 percent vs 11.4 percent, respectively). For IOI, the overall index was significantly higher in Test Version 2 than Test Version 1 (46.4 percent vs 40.1 percent, respectively). This result was driven by the CAPI mode, where the IOI for Test Version 2 was significantly higher than that in Test Version 1 (59.4 percent vs 42.6 percent, respectively).

Table 63. Difference in GDR for Multiple Coverage (Definition 1) for Test Version 1 and Test Version 2

Mode	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Overall	12.9 (0.5)	11.9 (0.5)	1.0 (0.7)	0.48
Internet	11.4 (0.5)	9.5 (0.6)	1.9 (0.7)	0.04*
Mail	18.2 (1.4)	19.6 (1.4)	-1.4 (2.0)	0.77
CAPI	13.4 (1.3)	12.9 (1.4)	0.6 (1.9)	0.77

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 64. Difference in IOI for Multiple Coverage (Definition 1) for Test Version 1 and Test Version 2

Mode	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Overall	40.1 (1.4)	46.4 (1.7)	-6.3 (2.2)	0.02*
Internet	39.3 (1.7)	41.6 (2.0)	-2.3 (2.5)	0.35
Mail	42.0 (2.8)	49.8 (3.2)	-7.9 (4.3)	0.13
CAPI	42.6 (4.0)	59.4 (5.2)	-16.7 (6.2)	0.02*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 65 and Table 66 show the GDRs and IOIs for multiple coverage Definition 2 in Test Version 1 and Test Version 2, respectively. For GDR, the overall percent of inconsistent answers was significantly different between versions, with the GDR being lower in Test Version 2 than Test Version 1 (14.3 percent vs 16.8 percent, respectively). This result was driven by the internet mode, where the GDR for internet was significantly lower in Test Version 2 than Test Version 1 (11.3 percent vs 14.3 percent, respectively). For IOI, the overall index was not significantly different between versions. Across original interview modes, only the IOI for CAPI was significantly different between versions, with the IOI being higher in Test Version 2 than Test Version 1 (79.5 percent vs 63.5 percent, respectively). Considering both IOI and GDR, and considering both definitions, there was inconclusive evidence whether Test Version 1 or Test Version 2 had higher reliability with regard to multiple coverage.

Table 65. Difference in GDR for Multiple Coverage (Definition 2) for Test Version 1 and Test Version 2

Mode	Version 1 GDR Percent	Version 2 GDR Percent	Difference	Adjusted P-Value
Overall	16.8 (0.6)	14.3 (0.6)	2.5 (0.8)	<0.01*
Internet	14.3 (0.6)	11.3 (0.7)	3.0 (0.9)	<0.01*
Mail	24.7 (1.7)	24.2 (1.6)	0.5 (2.4)	0.83
CAPI	18.5 (1.6)	15.6 (1.5)	2.8 (2.1)	0.34

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

Table 66. Difference in IOI for Multiple Coverage (Definition 2) for Test Version 1 and Test Version 2

Mode	Version 1 IOI Percent	Version 2 IOI Percent	Difference	Adjusted P-Value
Overall	54.8 (1.5)	59.1 (1.7)	-4.2 (2.3)	0.21
Internet	51.5 (1.8)	51.5 (2.1)	<0.1 (2.7)	1.00
Mail	57.6 (3.0)	63.6 (3.6)	-6.0 (4.9)	0.45
CAPI	63.5 (4.6)	79.5 (4.9)	-16.0 (6.4)	0.05*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0067
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. Significance was tested using a two-tailed t-test at the α=0.1 level. An asterisk (*) indicates a statistically significant result. P-values have been adjusted for multiple comparisons using the Hochberg method.

5.2.5. Other Metrics Results

Is there a difference in help text use on the internet and CAPI instruments between Test Version 1 and Test Version 2? Is there a difference in behavior on the internet instrument between Test Version 1 and Test Version?

Analysis for these research questions were covered by the 2022 Content Test Respondent Burden analysis (Virgile et al., 2023). In addition to examining help text use, we also discuss median completion times in internet and CAPI, breakoff rates in internet, and form completeness rates in all modes. Refer to Virgile et al. (2023) for detailed information on all results discussed. Additional analysis of answer switching and instrument navigation (backing up) will be documented separately and are not covered in this report.

For the entire survey, the median completion time was nominally faster in Test Version 2 compared to the Test Version 1, with a difference of about 30 seconds in internet and about 40 seconds in CAPI. For only the Health Insurance Coverage question, the median completion times were nominally faster in Test Version 2 than Test Version 1. The median completion times for internet were 31 seconds in Test Version 2 and 44 seconds in Test Version 1, and for CAPI were 34 seconds in Test Version 2 and 43 seconds in Test Version 1. These nominal differences

indicate that the check-all-that-apply format of Test Version 2 took less time for respondents to answer compared to the individual yes/no question format of Test Version 1.

The other measures of respondent burden using internet paradata were help screen access rates and breakoff rates. For the Health Insurance Coverage question, the help screen access rates were significantly higher in Test Version 1 than Test Version 2. This result could indicate that the format differences between Test Version 1 and Test Version 2 impacted respondents' understanding of the question. For breakoff rates, both overall and when looking at the Health Insurance Coverage question, the rates were not significantly different between treatments.

Form completeness rates measured the number of questions on the form that were answered among those that should have been answered, based on Content Test response data. None of the mail form completeness rates were significantly different between treatments. For internet, the form completeness rates were significantly higher in Test Version 2 than Test Version 1, both overall and for the Detailed Person section of the questionnaire. For CAPI, there were no significant differences in rates between treatments, overall and within questionnaire section. It is unclear how, and if, the internet mode differences in form completeness between treatments impacted the Health Insurance Coverage analyses.

6. CONCLUSIONS AND RECOMMENDATIONS

The American Community Survey is the principal source of detailed health insurance coverage information for state and sub-state geographies due to its large sample size. The purpose of testing the revised health insurance question was to enhance question reliability and validity.

Control vs Test Version 1

Based on the decision criteria, the results of the 2022 ACS Content Test provided evidence to support the implementation of Test Version 1 of the Health Insurance Coverage question over the Control version. Our primary decision criterion concerned health insurance write-ins. Respondents may write in a health insurance coverage when they misunderstand the question or do not know how to categorize their insurance. Therefore, a lower proportion of write-in responses was preferable. Results showed that the write-in proportions, as well as the proportion of write-ins determined to be out-of-scope, did not statistically differ between versions. On the other hand, the proportion of codable write-ins that referenced terms associated with the ACA or CHIP program was significantly lower in Test Version 1 than Control, indicating that the addition of these terms to the question text reduced write-ins for those coverage types.

A difference in the proportion of multiple types of health insurance coverage could indicate a reduction in respondent confusion on how to categorize their coverage. Thus, the second decision criterion stated that the version of the question with the lower rate of multiple

coverage was preferable. Overall, the proportion of persons with multiple types of coverage was significantly lower in Test Version 1 than the Control version. This difference was driven by the self-response modes (mail and internet responses), where Test Version 1 had significantly lower proportions than Control.

The third criterion concerned item missing data rates. High item nonresponse might indicate that a question lacked clarity, was sensitive, or was too difficult to answer. In general, lower item missing data rates were preferable. Although there were no significant differences between treatments in any of the overall item missing data rates, the proportion of partial responses was significantly lower in Test Version 1 than Control. The item missing data rates for Medicare, Medicaid, VA, TRICARE, Medicare combined with Medicaid, and Medicare combined with direct purchase were also significantly lower in Test Version 1 compared to the Control version.

For the fourth criterion, Test Version 1 also performed better than the Control version for the following reliability metrics: overall significantly lower GDR for direct-purchase coverage; significantly lower GDR for direct-purchase coverage in mail and internet; and significantly lower IOI for direct-purchase coverage in internet. On the other hand, no significant differences were found between Test Version 1 and Control when it comes to the response reliability indicators including overall IOI, IOI rates in mail and CAPI, and GDR rates in CAPI.

Overall, the results presented supported Test Version 1. It showed that Test Version 1 met the decision criteria when compared with Control on all of the key measures. Therefore, comparisons of prevalence of coverage and type of coverage were not considered in the decision criteria. The ultimate decision regarding what change to recommend for the ACS depended upon the analysis of Test Version 2 compared with Test Version 1.

Test Version 1 vs Test Version 2

We created Test Version 2 to address a recommendation from cognitive testing: including a response category that allows respondents to specifically report no coverage. Because the Content Test results provided evidence to support the implementation of Test Version 1 of the Health Insurance Coverage question over the Control version, we considered the decision criteria for whether Test Version 2 resolved issues identified with the current question format without introducing additional error.

Based on the decision criteria, the results of the 2022 ACS Content Test show mixed evidence to support the implementation of Test Version 2 over Test Version 1. The primary decision criterion concerned item missing data rates in Test Version 2 compared to Test Version 1, with lower item missing data rates being preferable. There was no significant difference in the complete item missing data rates between question versions, and there were no statistical differences by mode of response.

Because of the question format, partial item missingness was not defined for Test Version 2. The partial-missing response pattern of Test Version 1 and Control is unique to asking a series of yes/no questions in a self-response environment. In order to fill partially missing data (as Census would do with Test Version 1), assumptions need to be made and imputation methods used. In a production environment, Test Version 2 has a distinct advantage of no partial missing data.

A secondary criterion for acceptance of Test Version 2 was response reliability, that is, the ability of respondents to correctly identify the same health insurance in a follow-up interview. There were no significant differences in overall GDR or IOI measures between Test Version 1 and Test Version 2 for those who have health insurance. When examined by mode of original interview, only the IOI for VA coverage was significantly different between versions, with the IOI being higher for Test Version 2 than Test Version 1. For multiple coverage, whether Test Version 1 or Test Version 2 had higher reliability varied across definitions and reliability measures.

For uninsured persons, overall GDR and IOI were significantly lower in Test Version 2 than Test Version 1. Because a major change between question versions was the inclusion of the no coverage option, increased reliability among those without insurance may indicate people are better able to report an absence of coverage in Test Version 2.

Given the earlier evidence that people underreport Medicaid, Medical Assistance, and other means tested state-provided health plans, part of the third evaluation criterion stated that the version of the question which produced a higher Medicaid coverage rate was preferable. Overall Medicaid coverage rates were not significantly different between versions. However, when examining the rates by mode of response, Test Version 2 had significantly lower Medicaid coverage rates than Test Version 1 for mail responses, while the rates among internet and CAPI responses were not significantly different between question versions. This finding suggests that more research is needed to understand how public coverage for people with low income (Medicaid, CHIP, etc.) is reported.

Both versions had health insurance coverage rates that were not significantly different from each other. When looking at specific coverage type rates, only VA coverage was significantly different between versions, with the proportion being significantly lower in Test Version 2 than Test Version 1. Looking at coverage types by age, there were no significant differences in reported coverage for children under age 19, and working age adults ages 19 to 64. However, adults aged 65 and over reported significantly lower rates of employer-based coverage, direct-purchase coverage, and Medicare in Test Version 2 than Test Version 1.

Test Version 2 performed better than Test Version 1 for certain other aspects of health insurance reporting evaluated among the third criterion. Test Version 2 had significantly lower rates of multiple coverage and a significantly lower proportion of write-in responses. These

differences may indicate respondents were better able to designate their health insurance into the presented categories with the Test Version 2 response format, and a reduced number of write-ins shortens the time-intensive and costly process of classifying write-in responses.

In summary, there was mixed evidence whether Test Version 2 performed better than Test Version 1. There were no significant differences in complete item missing data rates, our main criterion. However, there were some improvements in our secondary criterion, as Test Version 2 had significantly higher reliability than Test Version 1 for the uninsured. For our third priority criterion there were mixed results, with improvements in reports of multiple coverage overall and across modes, as well as write-in responses overall and in self-response modes. The lower rates of Medicaid reporting among mail responses in Test Version 2 was a concern. The interagency subcommittee supporting the health insurance questions found that the supplementary comparisons of Medicaid-only were helpful in understanding that much of the difference between Test Version 1 and Test Version 2 was among people reporting multiple coverages.

Overall, the recommendation is to implement Test Version 2 of the Health Insurance Coverage question.

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

Beaumont, J-F. (2005). On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. *Survey Methodology*, 31(2), 227-231. https://www.surveypractice.org/article/3036-paradata-in-survey-research

Berchick, E., O'Hara, B., Heimel, S., & Chase Sawyer, R. (2017). *2016 American Community Survey evaluation report: Health insurance coverage, premiums and subsidies*. U.S. Census Bureau. Retrieved February 23, 2022, from https://www.census.gov/library/working-papers/2017/acs/2017_Berchick_01.html

Biemer, P. (2011). *Latent class analysis of survey error.* John Wiley & Sons, Inc.

Boudreaux, M., Call, K., Turner, J., & Fried, B. (2014). *Estimates of direct purchase from the ACS and Medicaid misreporting: Is there a link?* State Health Access Data Assistance Center. Retrieved February 23, 2022, from https://www.shadac.org/sites/default/files/publications/SHADACBrief38_DirectPurchase_Web.pdf

Boudreaux, M. H., Call, K. T., Turner, J., Fried, B., & O'Hara, B. (2015). Measurement error in public health insurance reporting in the American Community Survey: Evidence from record linkage. *Health Services Research*, 50(6), 1973-1995. https://doi.org/10.1111/1475-6773.12308

Boudreaux, M., Ziegenfuss, J. Y., Graven, P., Davern, M., & Blewett, L. A. (2011). Counting uninsurance and means-tested coverage in the American Community Survey: A comparison to the Current Population Survey. *Health Services Research*, 46(1), 210-231. https://doi.org/10.1111/j.1475-6773.2010.01193.x

Couper, M.P. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the 1998 Joint Statistical Meetings, Survey Research Methods Section* (pp. 41-49). American Statistical Association.

Couper, M.P., & Lyberg, L. (2005). The use of paradata in survey research. *Proceedings of the 55th Session of the International Statistic Institute.* https://www.isi-web.org/isi.cbs.nl/iamamember/CD6-Sydney2005/ISI2005_Presentations/Thu%20HA2%20940%20LYBERG%20Lars%20-%20TM05.pdf

Federal Committee on Statistical Methodology. (2001). *Measuring and reporting sources of error in surveys (Statistical Policy Working Paper 31).* Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved May 24, 2023, from https://www.fcsm.gov/assets/files/docs/spwp31.pdf

Flanagan, P.E. (2001). Measurement errors in survey response. (Unpublished doctoral dissertation). University of Maryland, Baltimore County.

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 2016 American Community Survey Content Test sample.* U.S. Census Bureau. DSSD 2016 American Community Survey Memorandum Series #ACS16-S-22.

Kreuter, F., Casas-Cordero, C., & German Data Forum (RatSWD). (2010). Paradata. In *Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences* (1st ed., pp. 509-530). Verlag Barbara Budrich.

Kreuter, F., Couper, M., & Lyberg, L. (2010). The use of paradata to monitor and manage survey data collection. *Proceedings of the 2010 Joint Statistical Meetings, Section on Survey Research Methods* (pp. 282-296). American Statistical Association. http://www.asasrms.org/Proceedings/y2010/Files/306107_55863.pdf

Lynch, V., Kenney, G. M., Haley, J., & Resnick, D. M. (2011). *Improving the validity of the Medicaid/CHIP estimates on the American Community Survey: The role of logical coverage edits*. U.S. Census Bureau. Retrieved February 24, 2022, from https://www.census.gov/library/working-papers/2011/demo/improving-the-validity-of-the-medicaid-chip-estimates-on-the-acs.html

Mach, A., & O'Hara, B. (2011). *Do people really have multiple health insurance plans? Estimates of nongroup health insurance in the American Community Survey*. U.S. Census Bureau. Retrieved February 24, 2022, from https://www.census.gov/library/working-papers/2011/demo/SEHSD-WP2011-28.html

O'Hara, B. (2010). *Is there an undercount of Medicaid participants in the 2006 ACS Content Test?* U.S. Census Bureau. Retrieved February 24, 2022, from https://www.census.gov/library/working-papers/2010/acs/2010_OHara_01.html

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 July 19, 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 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 27, 2023, from https://www.census.gov/library/working-papers/2023/acs/2023_Spiers_01.html

U.S. Census Bureau. (2007). *Subjects planned for the 2010 Census and American Community Survey: Federal legislative and program uses*. Retrieved February 28, 2022, from https://www2.census.gov/programs-surveys/acs/operations_admin/Final_2010_Census_and_American_Community_Survey_Subjects_Notebook.pdf

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

U.S. Census Bureau. (2022c). *Annual estimates of the resident population for the United States, regions, states, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2022 (NST-EST2022-POP)*. Retrieved May 31, 2023, from https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-total.html

U.S. Congress. (2010). *PUBLIC LAW 111-148, 111TH CONGRESS, An Act Entitled the Patient Protection and Affordable Care Act*. Retrieved September 29, 2022, from https://www.govinfo.gov/content/pkg/PLAW-111publ148/html/PLAW-111publ148.htm

Virgile, M., Weng, C. F., Mills, G., & Spiers, S. (2023). *2022 American Community Survey Content Test evaluation report: Respondent burden*. Retrieved November 27, 2023, from https://www.census.gov/content/dam/Census/library/working-papers/2023/acs/2023_Virgile_01.pdf