



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

November 13, 2023

2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM  
SERIES # ACS23-RER-12

MEMORANDUM FOR Donna Daily  
Chief, American Community Survey Office

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

Prepared by: Amy Steinweg *AS*  
Natalie Young *NY*  
Sharon Stern *SS*  
Social, Economic, and Housing Statistics Division

Lauren Contard *LC*  
Samantha Spiers *SS*  
Decennial Statistical Studies Division

Subject: 2022 American Community Survey Content Test Evaluation Report:  
Disability

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

If you have any questions about this report, please contact Lauren Contard at 301-763-3580.

Attachment

cc:

[acso.re.workgroup.list@census.gov](mailto:acso.re.workgroup.list@census.gov)

Lindsay Longsine DSSD  
Peter Massarone  
Broderick Oliver

Michael Risley  
Kelly Holder SEHSD  
Hyon Shin

*This page is intentionally blank.*

---

# **2022 American Community Survey Content Test Evaluation Report: Disability**

FINAL REPORT



**Amy Steinweg, Natalie Young, and Sharon Stern**  
Social, Economic, and Housing Statistics Division

**Lauren Contard and Samantha Spiers**  
Decennial Statistical Studies Division

*This page is intentionally blank.*

---

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY.....</b>	<b>7</b>
<b>1 BACKGROUND.....</b>	<b>9</b>
1.1 Proposals for New and Revised ACS Questions .....	10
1.2 Cognitive Testing.....	11
1.3 Field Testing Disability in the 2022 ACS Content Test .....	11
1.3.1 Justification for Inclusion of Disability in the Content Test .....	11
1.3.2 Census-Directed Cognitive Testing for Disability.....	17
1.3.3 Question Content.....	20
1.3.4 Research Questions .....	23
<b>2 METHODOLOGY.....</b>	<b>24</b>
2.1 Sample Design .....	24
2.2 Data Collection .....	25
2.3 Content Follow-Up Operation.....	27
2.3.1 Content Test Follow-Up Protocol .....	27
2.3.2 Content Test Follow-Up for Disability.....	28
2.4 Analysis Metrics.....	28
2.4.1 Unit-Level Analysis .....	29
2.4.2 Topic-Level Analysis .....	29
<b>3 DECISION CRITERIA.....</b>	<b>37</b>
<b>4 ASSUMPTIONS AND LIMITATIONS .....</b>	<b>38</b>
4.1 Assumptions .....	38
4.2 Limitations.....	39
<b>5 RESULTS .....</b>	<b>41</b>
5.1 Benchmark Results for Disability.....	42
5.2 Item Missing Data Rate Results for Disability .....	42
5.2.1 Item Missing Data Rates .....	42
5.2.2 Item Missing Data Rates for Proxy versus Self-Report Responses .....	47
5.2.3 Summary of Item Missing Data Rate Results.....	53
5.3 Response Distribution Results for Disability .....	54
5.3.1 Response Distributions for Overall Disability .....	54

5.3.2	Response Distributions by Disability Type .....	57
5.3.3	Response Distributions of Test Treatment Response Categories .....	62
5.3.4	Summary of Response Distribution Results.....	64
5.4	Response Reliability Results for Disability .....	66
5.4.1	Response Reliability for Overall Disability .....	67
5.4.2	Response Reliability by Disability Type .....	71
5.4.3	Response Reliability by Response Category in the Test Treatment .....	81
5.4.4	Summary of Response Reliability Results .....	82
5.5	Other Metrics: Respondent Burden Results for Disability .....	83
<b>6</b>	<b>CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>84</b>
<b>7</b>	<b>ACKNOWLEDGMENTS.....</b>	<b>88</b>
<b>8</b>	<b>REFERENCES .....</b>	<b>89</b>
	<b>APPENDIX A. Response Reliability Results for Same vs. Different CFU Respondent .....</b>	<b>93</b>

## LIST OF TABLES

Table 1. Age Universes by Disability Question .....	23
Table 2. Questions by Treatment .....	25
Table 3. Definitions of a Disability in the Control and Test Treatments.....	31
Table 4. Original Interview and CFU Reinterview Counts for Calculating GDR, IOI, and NDR.....	33
Table 5. Cross-Tab of Original Interview and CFU Results: Questions with Response Categories .....	35
Table 6. Decision Criteria for Disability.....	38
Table 7. Original Interview Unit Response Rates for Control and Test Treatments, Overall and Distributed by Mode .....	42
Table 8. Item Missing Data Rates: All Disability Questions .....	44
Table 9. Item Missing Data Rates: Hearing.....	45
Table 10. Item Missing Data Rates: Vision.....	45
Table 11. Item Missing Data Rates: Cognition .....	45
Table 12. Item Missing Data Rates: Ambulation.....	46
Table 13. Item Missing Data Rates: Self-Care .....	46
Table 14. Item Missing Data Rates: Independent Living .....	46
Table 15. Item Missing Data Rates: Communication (Test Version Only) .....	47
Table 16. Item Missing Data Rates for All Disability Questions: Self-Report.....	48
Table 17. Item Missing Data Rates for All Disability Questions: Proxy.....	48
Table 18. Item Missing Data Rates for Hearing: Self-Report .....	49
Table 19. Item Missing Data Rates for Hearing: Proxy .....	49
Table 20. Item Missing Data Rates for Vision: Self-Report.....	49
Table 21. Item Missing Data Rates for Vision: Proxy .....	50
Table 22. Item Missing Data Rates for Cognition: Self-Report .....	50
Table 23. Item Missing Data Rates for Cognition: Proxy .....	50
Table 24. Item Missing Data Rates for Ambulation: Self-Report.....	51
Table 25. Item Missing Data Rates for Ambulation: Proxy .....	51
Table 26. Item Missing Data Rates for Self-Care: Self-Report .....	51
Table 27. Item Missing Data Rates for Self-Care: Proxy .....	52
Table 28. Item Missing Data Rates for Independent Living: Self-Report.....	52
Table 29. Item Missing Data Rates for Independent Living: Proxy .....	52
Table 30. Item Missing Data Rates for Communication: Self-Report (Test Version Only) .....	53
Table 31. Item Missing Data Rates for Communication: Proxy (Test Version Only) .....	53
Table 32. Distribution of Overall Disability: Definition 1 .....	56
Table 33. Distribution of Overall Disability: Definition 2 .....	57
Table 34. Distribution of Disability: Hearing.....	58
Table 35. Distribution of Disability: Vision.....	59
Table 36. Distribution of Disability: Cognition .....	59
Table 37. Distribution of Disability: Ambulation.....	60

Table 38. Distribution of Disability: Self-Care .....	61
Table 39. Distribution of Disability: Independent Living .....	61
Table 40. Distribution of Disability: Communication (Test Version Only) .....	62
Table 41. Distribution of Test Treatment Response Categories: Hearing .....	63
Table 42. Distribution of Test Treatment Response Categories: Vision .....	63
Table 43. Distribution of Test Treatment Response Categories: Cognition .....	63
Table 44. Distribution of Test Treatment Response Categories: Ambulation .....	63
Table 45. Distribution of Test Treatment Response Categories: Self-Care .....	64
Table 46. Distribution of Test Treatment Response Categories: Independent Living .....	64
Table 47. Distribution of Test Treatment Response Categories: Communication .....	64
Table 48. CFU Reinterview Unit Response Rates for Control and Test Treatments, Overall and by Mode .....	66
Table 49. Distribution of Overall Disability: CFU Respondents vs. Nonrespondents .....	67
Table 50. GDR for Definition 1 of Disability: Overall Disability .....	69
Table 51. Adjusted IOI for Definition 1 of Disability: Overall Disability .....	69
Table 52. GDR for Definition 2 of Disability: Overall Disability .....	70
Table 53. Adjusted IOI for Definition 2 of Disability: Overall Disability .....	70
Table 54. GDR for Definition 1 of Disability: Hearing .....	72
Table 55. Adjusted IOI for Definition 1 of Disability: Hearing .....	72
Table 56. GDR for Definition 1 of Disability: Vision .....	72
Table 57. Adjusted IOI for Definition 1 of Disability: Vision .....	73
Table 58. GDR for Definition 1 of Disability: Cognition .....	73
Table 59. Adjusted IOI for Definition 1 of Disability: Cognition .....	73
Table 60. GDR for Definition 1 of Disability: Ambulation .....	74
Table 61. Adjusted IOI for Definition 1 of Disability: Ambulation .....	74
Table 62. GDR for Definition 1 of Disability: Self-Care .....	74
Table 63. Adjusted IOI for Definition 1 of Disability: Self-Care .....	75
Table 64. GDR for Definition 1 of Disability: Independent Living .....	75
Table 65. Adjusted IOI for Definition 1 of Disability: Independent Living .....	75
Table 66. GDR for Definition 1 of Disability: Communication (Test Version Only) .....	76
Table 67. Adjusted IOI for Definition 1 of Disability: Communication (Test Version Only) .....	76
Table 68. GDR for Definition 2 of Disability: Hearing .....	77
Table 69. Adjusted IOI for Definition 2 of Disability: Hearing .....	77
Table 70. GDR for Definition 2 of Disability: Vision .....	78
Table 71. Adjusted IOI for Definition 2 of Disability: Vision .....	78
Table 72. GDR for Definition 2 of Disability: Cognition .....	78
Table 73. Adjusted IOI for Definition 2 of Disability: Cognition .....	79
Table 74. GDR for Definition 2 of Disability: Ambulation .....	79
Table 75. Adjusted IOI for Definition 2 of Disability: Ambulation .....	79
Table 76. GDR for Definition 2 of Disability: Self-Care .....	80
Table 77. Adjusted IOI for Definition 2 of Disability: Self-Care .....	80



Table 78. GDR for Definition 2 of Disability: Independent Living .....	80
Table 79. Adjusted IOI for Definition 2 of Disability: Independent Living .....	81
Table 80. GDR for Definition 2 of Disability: Communication (Test Version Only) .....	81
Table 81. IOI for Definition 2 of Disability: Communication (Test Version Only) .....	81
Table 82. IOI <sub>L</sub> by Disability Type (Test Treatment Only) .....	82
Table 83. GDR for Definition 1 of Disability for Same CFU Respondent: Overall Disability .....	93
Table 84. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Overall Disability .....	93
Table 85. GDR for Definition 1 of Disability for Different CFU Respondent: Overall Disability....	93
Table 86. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Overall Disability.....	94
Table 87. GDR for Definition 1 of Disability for Same CFU Respondent: Hearing .....	94
Table 88. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Hearing .....	94
Table 89. GDR for Definition 1 of Disability for Same CFU Respondent: Vision .....	95
Table 90. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Vision .....	95
Table 91. GDR for Definition 1 of Disability for Same CFU Respondent: Cognition .....	95
Table 92. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Cognition .....	96
Table 93. GDR for Definition 1 of Disability for Same CFU Respondent: Ambulation.....	96
Table 94. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Ambulation .....	96
Table 95. GDR for Definition 1 of Disability for Same CFU Respondent: Self-Care .....	97
Table 96. IOI for Definition 1 of Disability for Same CFU Respondent: Self-Care.....	97
Table 97. GDR for Definition 1 of Disability for Same CFU Respondent: Independent Living.....	97
Table 98. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Independent Living .....	98
Table 99. GDR for Definition 1 of Disability for Same CFU Respondent: Communication (Test Version Only).....	98
Table 100. IOI for Definition 1 of Disability for Same CFU Respondent: Communication (Test Version Only).....	98
Table 101. GDR for Definition 1 of Disability for Different CFU Respondent: Hearing .....	99
Table 102. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Hearing ...	99
Table 103. GDR for Definition 1 of Disability for Different CFU Respondent: Vision .....	99
Table 104. IOI for Definition 1 of Disability for Different CFU Respondent: Vision.....	100
Table 105. GDR for Definition 1 of Disability for Different CFU Respondent: Cognition .....	100
Table 106. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Cognition .....	100
Table 107. GDR for Definition 1 of Disability for Different CFU Respondent: Ambulation .....	101
Table 108. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Ambulation .....	101
Table 109. GDR for Definition 1 of Disability for Different CFU Respondent: Self-Care .....	101
Table 110. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Self-Care	102

Table 111. GDR for Definition 1 of Disability for Different CFU Respondent: Independent Living .....	102
Table 112. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Independent Living .....	102
Table 113. GDR for Definition 1 of Disability for Different CFU Respondent: Communication (Test Version Only).....	103
Table 114. IOI for Definition 1 of Disability for Different CFU Respondent: Communication (Test Version Only).....	103

#### **LIST OF FIGURES**

Figure 1. Control Version of the Disability Questions (Paper).....	20
Figure 2. Test Version of the Disability Questions (Paper) .....	21

## EXECUTIVE SUMMARY

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

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 (of which one test treatment contained the Disability Test question set).

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 housing units that completed the original Content Test interview, a Content Follow-Up telephone reinterview was conducted to measure response reliability.

The National Center for Health Statistics (NCHS) proposed that the Census Bureau modify the Disability questions in the ACS. NCHS recommended that the Census Bureau use the Washington Group Short Set on Functioning (WG-SS) as a replacement to the existing Disability question set in the ACS (ACS-6). The WG-SS differs from the ACS-6 in that it includes a communication difficulty question and uses graded response categories (rather than a yes/no response). It also contains a number of wording differences. Prior NCHS evaluations of the WG-SS indicate that it is a valid measure of functional disability with advantages over the ACS-6. The WG-SS is more consistent with current conceptualizations of functional disability as existing on a continuum, and thus better aligns with the World Health Organization's (WHO) International Classification of Functioning, Disability, and Health (ICF). The WG-SS is also internationally comparable and is the disability measure recommended by the United Nations. As such, adopting a version of the WG-SS would potentially allow ACS estimates of disability prevalence in the United States to be compared to estimates from other countries. The WG-SS has already been incorporated into several other federal surveys, including the principal health survey for the U.S – the National Health Interview Survey (NHIS) – as well as the National Health

and Nutrition Examination Survey (NHANES), the National Study of Family Growth (NSFG), and the Census Bureau’s experimental Household Pulse Survey.

For the 2022 ACS Content Test, a Control version that included the current ACS measure of Disability (ACS-6) was compared to a Test version that instead included the proposed Disability measure (a version of the WG-SS). While the analysis team developed a number of research questions to better understand how the proposed Disability questions work in the ACS environment, the *a priori* decision criteria for Disability had two main components. Decision criterion number 1, which was assigned the highest priority level, concerned response reliability. In general, higher response reliability is preferred, but it was established *a priori* that lower reliability in the Test treatment may be acceptable when considering the previously documented validity of the WG-SS. Decision criterion number 2, which was assigned second priority, involved item missing data rates. While lower item missing data rates are generally preferred, the analysis team recognized that the Test treatment might have higher item missing data rates than Control simply because of the more complex response choices and the addition of the question about communication difficulty (as additional items generally increase the chance of missing data). Thus, in advance of analyzing the field test data, the decision criteria established that a small increase in item missing data rates would be deemed acceptable. It should be noted that there is more than one way to define Disability in the WG-SS, due to the use of graded response options. Two definitions were considered by the analysis team: the standard definition recommended by NCHS (Definition 1), and an alternative, broader definition of Disability (Definition 2).<sup>1</sup>

Response reliability (Decision Criterion 1) was measured by comparing responses from the original interview and a Content Follow-Up (CFU) interview and was computed using two metrics: the Gross Difference Rate (GDR) and the Index of Inconsistency (IOI). Results were mixed; they did not consistently favor one version of the Disability question set over another. When GDR – a simple measure of response variance – was used as the metric, the Test treatment generally had better response reliability than the Control when using the recommended definition of Disability (Definition 1) and worse response reliability than Control when using the alternative, broader definition of Disability (Definition 2). Results were less consistent when response reliability was instead measured using the more complex metric of IOI. Using IOI, the Test treatment sometimes had worse response reliability than the Control treatment, while in other cases no significant difference was detected between the treatments. Results for item missing data rates (Decision Criterion 2) were less complex than the response

---

<sup>1</sup> The standard definition of Disability (Definition 1) dichotomizes the graded answer scale to indicate disability when respondents report “a lot of difficulty” or “cannot do at all.” The broader definition (Definition 2) also includes those who report “some difficulty.”

reliability results. Overall, item missing data rates were sometimes higher in the Test treatment, compared to Control, but only by about 1 percentage point.

Because determining whether the Test treatment met the two decision criteria involved interpretation of the data, the Census Bureau consulted with the Disability Subcommittee, which is an interagency group consisting of Census Bureau staff and representatives from other federal agencies, including NCHS. Through a series of emails and meetings, the Disability Subcommittee reviewed and discussed the results of the 2022 ACS Content Test for Disability. Overall, it was concluded that the Test treatment met both decision criteria. The recommendation of the Disability Subcommittee is to move forward with the Test version of the Disability question set, which is based on the Washington Group Short Set and also includes Independent Living difficulty, for a total of 7 questions. It is recommended that Definition 1 of Disability be used as the primary Census Bureau measure, as field test results did not indicate that the alternative approach (Definition 2) performed better than the standard method recommended by the Washington Group (Definition 1). Whether estimates based on Definition 2 of Disability will also be released in data products will depend on results of subsequent analyses.

Notably, consistent with prior NCHS evaluations, results from the 2022 ACS Content Test indicate that the revised Disability measure will not be comparable to the current ACS measure. When the definition of Disability recommended by NCHS was used (Definition 1), the estimated percent of the U.S. population with any disability was about 40 percent lower in Test than in Control (8.1 percent (SE = 0.2) in Test, compared to 13.9 percent (SE = 0.3) in Control). This difference is consistent with other analyses that compared the ACS-6 with the WG-SS. Cognitive interviews conducted prior to the field test suggest that the revised Disability measure captures a smaller population and a population with a higher level of functional difficulty, relative to current ACS measure. It should be further noted that while we expect overall Disability estimates to be lower when the new questions are put into production, the magnitude of difference for each individual Disability type may vary. Overall, the population identified as having a disability under the revised Disability measure will not be the same as the one identified by the current measure. This difference is important for data users, who will need to address the implications of the Disability measure change in their own work.

## **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 Disability.

### **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.

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 for changes related to their topic and consulted throughout the decision-making process. The Disability Subcommittee included Census Bureau staff and representatives from other federal agencies including: The National Center for Health Statistics (NCHS), the National Center on Birth Defects and Developmental Disabilities (NCBDDD), the Department of Human Development and Disability (DHDD), the Center for Medicare and Medicaid Services (CMS), the Agency for Healthcare Research and Quality (AHRQ), the U.S. Department of Veterans Affairs (VA), the Bureau of Justice Statistics (BJS), the U.S. Equal Opportunity Employment Commission (EEOC), the Office of Disability Employment Policy (ODEP), 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.<sup>2</sup> Cognitive interviews were conducted virtually, in English and Spanish.<sup>3</sup> 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 Fee, Home Heating Fuel, and Means of Transportation to Work. For the most part, the changes to these questions are expected to either impact a small population or result in a small change in the data that would not be detectable in the Content Test. The subject matter experts recommended that cognitive testing was sufficient for these questions and that field testing was not necessary; the Interagency Council on Statistical Policy Subcommittee on the ACS agreed with this recommendation. Content changes for these topics will be implemented in production ACS in 2024.

## 1.3 Field Testing Disability in the 2022 ACS Content Test

### 1.3.1 Justification for Inclusion of Disability in the Content Test

The National Center for Health Statistics (NCHS) proposed that the Census Bureau use the Washington Group Short Set on Functioning (WG-SS) as a replacement to the existing Disability question set in the ACS (ACS-6). The NCHS is the principal health statistics agency for the United States. Over the past two decades, NCHS has invested substantial resources in the development and evaluation of a standardized measure of disability for use in national surveys and censuses worldwide. It is worth noting that both the ACS-6 and the WG-SS operationalize the concept of “disability” by asking about a specific set of *functional difficulties*. As such, these measures of disability are best understood as measures of *functional* disability – as opposed to, for example,

---

<sup>2</sup> 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.

<sup>3</sup> 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.

work disability or self-identification as “disabled.” When referring to these measures of “disability” throughout this report, it should be understood that we are referring to functional disability. The following sections give background on the development of the WG-SS, the development of the ACS-6, and previous research on how the WG-SS performs in the ACS environment.

#### **1.3.1.1 Development of the Washington Group Short Set on Functioning**

The WG-SS was developed by the Washington Group on Disability Statistics (WG), a City Group created by the United Nations (UN) Statistical Commission to improve the quality and international comparability of disability statistics worldwide.<sup>4</sup> The group was created in response to issues raised at the 2001 International Seminar on the Measurement of Disability about differences in how disability is conceptualized and operationalized within surveys internationally, and the resulting variation in the validity, reliability, and cross-national comparability of national disability estimates. A primary goal of the Washington Group is to establish agreed upon standards for the measurement of disability in national surveys and censuses, in part to improve cross-survey and cross-national comparability of disability estimates (Washington Group on Disability Statistics, 2020). The Washington Group, whose Secretariat is located at NCHS, includes representatives from statistical agencies from across the world. The first product developed by the Washington Group was a short set of questions for use in censuses and surveys – the WG-SS. The WG-SS was endorsed by the Washington Group in 2006, following cognitive testing in the U.S. and fifteen other countries, as well as field testing in the U.S., Argentina, Brazil, Gambia, Paraguay, and Vietnam (Madans, 2017).

Since 2006, the WG-SS has been included in censuses and surveys in over 90 countries worldwide based on Washington Group tallies (Miller et al., 2020). It is the question set recommended for use in censuses by the United Nations Statistical Division (UNSD) and the United Nations Economic Commission for Europe for measuring disability internationally, for monitoring the UN Convention on the Rights of Persons with Disabilities, and for disaggregating Sustainable Development Goal Indicators and other international commitments (United Nations, 2017; United Nations Economic Commission for Europe, 2015; United Nations General Assembly, 2007; United Nations General Assembly, 2015; United Nations Economic and Social Commission for Asia and the Pacific, 2012). The WG-SS is also used by the U.S. Agency for International Development (USAID), the World Bank, and the Demographic and Health Surveys (DHS) Program. The WG-SS is currently collected as part of the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), and the National Survey of Family Growth (NSFG). In addition, the WG-SS has been included in the Census Bureau’s experimental Household Pulse Survey.

---

<sup>4</sup> United Nations “City Groups” are informal groups of experts, primarily from national statistical agencies, to discuss and address specific thematic challenges in the development and implementation of statistical methodologies. Some of these groups are known as “City Groups,” in reference to the place where they held their first meeting.



The WG-SS is a set of six questions designed to identify the population with functional disability (in a census or survey format but the questions are also recommended for use in administrative systems), namely those at greater risk than the general population for participation restrictions due to the presence of difficulties in six core functional domains, if appropriate accommodations are not made. The questions ask whether people have difficulty performing basic universal activities in six domains of functioning: Vision, Hearing, Ambulation, Cognition, Self-Care, and Communication. The response options allow for a continuum of functional difficulty to be reported: 1) no difficulty, 2) some difficulty, 3) a lot of difficulty, 4) cannot do at all. Different definitions of disability status can be created using the WG-SS, but if using a dichotomous measure, the Washington Group recommends defining a respondent as having a disability if they report “a lot of difficulty” or “cannot do at all” for at least one activity in the question set. Those who report “no difficulty” or “some difficulty” for all questions are not considered to have a disability. The WG-SS is designed to provide comparable data cross-nationally for populations living in a variety of cultures with varying economic resources.

### **1.3.1.2 Development of the Current ACS Disability Question Set (ACS-6)**

The current ACS Disability question set (ACS-6) was developed around the same time as the WG-SS in a related effort to improve disability statistics collected in federal surveys. This effort, initiated by OMB, emerged in response to data user concerns about disability questions included in the 2000 Census and in the ACS. OMB’s Interagency Committee for the ACS established an ACS Subcommittee on Disability Measurement in 2003. NCHS was asked to spearhead an evaluation of the ACS Disability question set, with the help of other federal agencies (Brault, 2009). The ACS Subcommittee reviewed agency mandates and determined that information on disability was necessary for at least two major reasons: 1) to monitor whether persons with disabilities are being prevented from full participation in society as outlined in the 1990 Americans with Disabilities Act and 2) to estimate the number of persons eligible for service programs offered by state and federal governments (Brault et al., 2007).

Similar to the WG, the ACS Subcommittee on Disability Measurement used the International Classification of Functioning, Disability and Health (ICF) as a conceptual guide for identifying disability domains and an approach to question construction based on a definition of disability located at the person-level, conceptualizing limitations or difficulties as possible risk factors associated with restrictions to full participation in society (World Health Organization, 2001). The resulting question set has come to be known as the ACS-6. Like the WG-SS, the ACS-6 is a measure of functional disability: it consists of six questions designed to measure limitations individuals experience in six domains of functioning. Many of these domains overlap with those measured by the WG-SS. Specifically, the ACS-6 asks whether people have difficulty with vision, hearing, ambulation, cognition, self-care, and/or independent living. Unlike the graded response categories on the WG-SS, the ACS-6 asks individuals to select either “yes” or “no” when responding to each question. Although some members of the ACS Subcommittee on Disability Measurement in 2006 believed that graded response categories would more

accurately reflect the continuum of functional abilities, the strict size limit of the ACS paper form restricted the committee's ability to consider such an option. As such, a dichotomous yes/no response was used for each of the six questions in the ACS-6 and, in some cases, the word "serious" was added to the question stem wording to narrow the level of difficulty that would qualify as a 'yes' response. An individual is considered to have a disability if they report difficulty with at least one activity – that is, if they respond "yes" to at least one item in the question set.

In January-March 2006, a content test was conducted to evaluate the question set that would become the ACS-6. Both the Census Bureau and NCHS were involved in content testing efforts. The test concluded that the new question set (ACS-6) resulted in improvements over the previous question set, in terms of reliability and response and its ability to identify the population of persons with disabilities. In 2008, the ACS-6 replaced the previous question set in the ACS. Further, the new ACS Disability question set was also incorporated into a number of other surveys, including the Survey of Income and Program Participation (SIPP), the Current Population Survey (CPS), the National Survey of Crime Victimization (NCVS), the National Health Interview Survey (NHIS), and the National Survey of Family Growth (NSFG). In 2011, in response to the Affordable Care Act being signed into law in 2010, the U.S. Department of Health and Human Services (HHS) made a programmatic change, recommending that the ACS-6 question set be used in all surveys that are conducted or sponsored by HHS (U.S Department of Health and Human Services, 2011).

Over time, a number of federal surveys have either switched (or are planning to switch) from the ACS-6 to the WG-SS. Due to a shared conceptual basis and approach to measuring disability, the WG-SS and ACS-6 question sets are similar in many ways (see Section 1.3.3). The main differences between the sets are in the sixth domain of functioning measured and the response categories. Both sets use the following five domains: vision, hearing, ambulation, cognition, and self-care. As the sixth domain, the WG-SS includes difficulty with communication and the ACS-6 uses difficulty with independent living. In terms of response categories, the WG-SS uses graded response choices, while the ACS uses a dichotomous yes/no response.

#### **1.3.1.3 NHIS Analysis: Comparing the WG-SS and ACS Disability Question Set**

Based on findings from cognitive interviews and field tests conducted by NCHS and other stakeholders within the U.S., as well as in other countries, the WG-SS is a valid, reliable measure of functional disability with cross-national comparability (Altman, 2016). In addition, there have been opportunities to evaluate the WG-SS alongside the ACS-6 prior to the current Content Test. In 2011-2012, for example, the NHIS included both the WG-SS and the ACS-6. Both sets of questions were asked of the same respondents, allowing for a comparison of individuals' responses across the two question sets (Weeks et al., 2021).

Weeks et al. (2021) published results of their comparison of the WG-SS and the ACS disability questions using the 2011-2012 NHIS data. The first step in the analysis was to assess

consistency in disability status across the two question sets. Notably, the authors decided to focus on just the five functional domains that the WG-SS and the ACS-6 have in common: Hearing, Seeing, Ambulation, Cognition, and Self-care. The questions on Communication (from the WG-SS), and on Independent Living (from the ACS-6), were not included within their analysis. Individuals were considered to have a disability according to the ACS measure if they had a value of ‘yes’ for at least one of the five domains, while they were considered to have a disability according to the WG-SS measure if they reported “a lot of difficulty” or “cannot do it all” for at least one of the five domains.

Given the high percentage of respondents reporting “no difficulty” on both question sets, nonagreement between the ACS and Washington Group measures was generally low. However, the ACS version of the questions identified a larger percentage of the population as having a disability. A detailed analysis of the response categories was then conducted on the NHIS data to assess the nature of the differences in identification. Responses of “no difficulty” in the WG-SS were highly concordant with “no” responses to the ACS questions, and responses of “a lot of difficulty” or “cannot do at all” to the WG-SS questions were highly concordant with “yes” responses to the ACS questions. In contrast, respondents who reported “some difficulty” to at least one of the five WG-SS questions included in analysis did not consistently fall into either the “yes” or “no” response category for the five ACS questions used in analysis: 63 percent responded “no” to all five relevant ACS questions, while 37 percent responded “yes” to at least one of the five ACS questions. Notably, respondents who fell into the “some difficulty” category for the WG-SS and responded “yes” to at least one ACS question were classified differently by the two question sets: they were considered to have a disability according to the ACS definition, while they were not considered to have a disability based on the dichotomized WG-SS measure. This discrepancy accounts for the higher estimate of disability prevalence using the ACS version of the questions, relative to the WG-SS version. In their report, Weeks et al. (2021) concluded that the population with disability defined by the ACS questions is more heterogeneous in functional level than that defined by the WG-SS questions.

#### **1.3.1.4 NCHS-Directed Cognitive Interviews of WG-SS in ACS Environment**

Two important limitations of the earlier testing of WG-SS warrant discussion. The first limitation relates to the mode of survey administration. The WG-SS has been tested extensively in the context of interviewer-administered data collection and less so in self-response environments (although the WG-SS has been included in censuses and surveys that use various modes of administration, including self-response). ACS data collection includes two self-administered versions of the questionnaire: a self-administered paper instrument and a self-administered internet instrument. In recent years, the internet instrument has become increasingly popular, highlighting the importance of directly evaluating whether the WG-SS performs well within the context of a self-administered questionnaire. The second limitation regards the use of within-household proxy responses. Although the WG-SS has been used in censuses where information about all members of the household is obtained from a single household respondent, during

initial testing of the WG-SS, individuals were only asked about difficulties they themselves experience with activities, such as seeing, hearing, self-care, and communication. The ACS, on the other hand, is designed for both self-report and proxy responses; that is, the individual completing the survey is not only asked about their own difficulties with activities, but also about difficulties other members of the household experience. This highlights the importance of directly testing performance of the WG-SS in proxy response contexts.

Prior to the current field test, cognitive interviewing was conducted by NCHS to evaluate the performance of the WG-SS in an ACS-like environment that included self-response and within-household proxy reporting. Between August 2019 and February 2020, NCHS conducted 43 cognitive interviews. Respondents completed an abbreviated, self-administered paper version of the ACS instrument, with half of the respondents receiving a form that included a version of the ACS-6 Disability question set, and the other half receiving a form with a version of the WG-SS. The two versions of the ACS survey instrument contained minimal non-Disability content, to allow analysis to focus on the Disability section. All non-Disability content present in the instruments was identical across the two versions. To assess differences between the two question sets in proxy reporting, respondents not only answered the questions for themselves, but also on behalf of other household members.

In order to focus testing on the effect of the answer categories, the WG-SS was modified slightly for use in these cognitive interviews. Specifically, the wording of the ambulatory difficulty and self-care difficulty questions was adjusted to match the wording used in the ACS-6 question set (i.e., “bathing” was used in lieu of “washing all over” and “climbing stairs” was used instead of “climbing steps”) and the communication question was removed. Since the question about independent living difficulty in the ACS-6 question set is not in the WG-SS, it was omitted from the ACS-6 question set for this evaluation (resulting in a set of ACS questions modified to include only five questions). As such, the biggest difference between the two Disability question sets tested in the NCHS evaluation concerned answer categories. Respondents who received the version of the ACS instrument with the modified WG-SS could choose among four response options for the Disability questions (“no difficulty”; “some difficulty”; “a lot of difficulty”; “cannot do at all”), while those who received the version containing the modified ACS-6 were presented with dichotomous response options (“yes”; “no”). After completing the ACS-style form, respondents were probed to better understand how they interpreted the questions, as well as how they decided among the response categories. Respondents who received the version of the ACS instrument containing the modified WG-SS were also asked how they would have responded if they had instead received the version with the modified ACS-6, and vice versa.

Results from the NCHS cognitive interviews suggest that there is variability across individuals in the way respondents evaluate their level of difficulty with activities, both for the (modified) ACS-6 and the (modified) WG-SS. Two respondents who experience a similar level of difficulty seeing, for example, may not provide the same answer to the question about vision difficulty.

This issue appeared to be more pronounced, however, for the (modified) ACS-6. NCHS noted that some respondents who experienced minor difficulty with an activity responded “no” to the relevant (modified) ACS-6 question, while others responded “yes.” This situation resulted in a high degree of heterogeneity in the functional abilities of people falling into the “yes” category for the (modified) ACS-6, as well as some heterogeneity among those who responded “no.”

In their report on the NCHS cognitive testing, Miller et al. (2020) find that respondents who reported “a lot of difficulty” or “cannot do at all” when presented with the (modified) WG-SS version were a more homogeneous group, relative to those who responded “yes” to the (modified) ACS-6. Specifically, respondents who selected “a lot of difficulty” reported experiencing difficulty frequently and in numerous contexts, and those who reported “cannot do at all” said they could not perform the activity in any context. Overall, findings from the NCHS cognitive interviews suggest that the dichotomous WG-SS measure of Disability captures a population with higher and more homogeneous levels of functional difficulties (i.e., in terms of frequency, intensity, and impact) than the population captured by the ACS-6. Miller et al. (2020) also present evidence that the WG-SS is better able to depict the range of functioning with more consistency, compared to the ACS-6. Finally, in terms of proxy reporting, the cognitive interviews suggested that respondents draw on similar information when responding to either the (modified) ACS-6 or the (modified) WG-SS on behalf of other household members. Specifically, they drew on their own observations of household members to assess frequency and severity of difficulty with activities, or on teachers’ or doctors’ evaluations if they felt their own observations were inadequate.

#### **1.3.1.5 Summary and Motivation for Inclusion of Disability in the Content Test**

Based on the results of the analysis of the NHIS data and the NCHS-directed cognitive interview study, researchers from NCHS concluded that the WG-SS was preferable to the ACS-6. The WG-SS provides more granular responses that may better describe the functional characteristics of the population. Rather than measuring disability as a dichotomy, the WG-SS is designed to obtain information on the broader range of difficulties in each domain. Thus, it may be possible to disaggregate equity and other measures by an overall indicator of disability, as well as by level of difficulty in each of the functional domains, which may increase the policy relevance of the information.

The 2022 ACS Content Test extends the evaluation of the performance of the WG-SS, relative to the ACS-6. As previously mentioned, the ACS heavily relies on within-household proxy reporting as well as self-response modes, such as mail and internet. As such, the main goal for the field test was to evaluate how the WG-SS performs in the ACS context, specifically in self-response modes and situations when the respondent is not answering about themselves.

#### **1.3.2 Census-Directed Cognitive Testing for Disability**

As explained in Section 1.2, the Census Bureau routinely conducts cognitive interviews as part of the process of pre-testing new and modified survey questions. In preparation for the field

test portion of the 2022 ACS Content Test, revisions to the ACS Disability question set (ACS-6) were tested along with changes in the health insurance and educational attainment questions. While separate evaluations were written for each topic area, all interviewees had the three topic areas included in the interview. This process was different from the NCHS-directed testing which focused exclusively on Disability.

A total of 115 cognitive interviews were conducted in two rounds (45 interviews in Round 1; 70 interviews in Round 2). In Round 1 of cognitive interviewing, two versions of the instrument were tested across two different modes of survey administration. Respondents were randomly assigned to either Version 1 or Version 2 of the instrument and to either the paper or CAI (Computer-Assisted Interviewing) mode of administration. The two versions of the instrument were then revised in response to findings from Round 1 and underwent further testing in Round 2.<sup>5</sup> For more information on the cognitive testing procedures and results from Round 1 and Round 2, see RTI International (2022a).

In terms of testing Disability content, the cognitive interviews were designed to evaluate different versions of the WG-SS, as well as to assess how respondents decide between the four response categories (“no difficulty”; “some difficulty”; “a lot of difficulty”; “cannot do at all”). In Round 1, two versions of the WG-SS were compared: a version that used the same wording as the question set developed by the Washington Group (Version 2), and a version with some wording modifications to reduce differences with the ACS-6 (Version 1). The two versions of the WG-SS were then revised after Round 1 and underwent further testing in Round 2.

In some cases, the cognitive interviews supported the original WG-SS wording, while in other cases, the modified wording performed better. In measuring ambulatory difficulty, for example, Round 1 cognitive interviews indicated that there was some variation in the interpretation of the original WG-SS wording, which asked about “difficulty climbing steps,” while the modified WG-SS wording, “difficulty climbing stairs,” was interpreted in the same way by nearly all respondents (RTI International, 2022a). In Round 2, the Disability Subcommittee sought to obtain additional information about the performance of the original WG-SS version of the question (“difficulty climbing steps”) by only including this version of the question in the Round 2 materials. Overall, most respondents understood the question as intended, leading the subcommittee to recommend that the original WG-SS wording be used in the field test.

The two rounds of cognitive interviews were also an opportunity to evaluate the performance of the WG-SS question about communication difficulty, which is not included in the ACS-6. Overall, most respondents who completed the English-language ACS questionnaire understood the communication question as intended, including both monolingual and bilingual individuals. A few response problems were found for the Spanish translation of the communication

---

<sup>5</sup> A third round of cognitive interviews was conducted with respondents in Puerto Rico and group quarters, but the results from those interviews were not available until design of the field test was complete, and thus did not affect the field test. For the results of that testing, see RTI International (2022b).

question in Round 1. Some monolingual Spanish speakers misinterpreted the question as asking about difficulty they experienced communicating with non-Spanish speakers (RTI International, 2022a). A new translation of the question was proposed and tested in Round 2. Results from Round 2 indicated that the revised Spanish translation was generally understood as intended. Based on these results, the Disability Subcommittee moved forward with the original WG-SS version of the communication question for the English-language instrument and the revised Spanish translation of the question for the Spanish-language instrument.

In addition, cognitive interviews were used to assess the performance of two versions of the ACS-6 question about independent living difficulty when asked alongside the WG-SS (since the WG-SS does not include a question about independent living difficulty). In the ACS-6, the question about independent living difficulty (“difficulty doing errands alone...”) is preceded by text intended to reduce reports of independent living difficulty that are due to issues other than disability. Specifically, respondents are asked to only report difficulty “due to a physical, mental, or emotional condition,” as opposed to difficulty due to transportation issues, language barriers, or other out-of-scope causes. Since no other questions in the WG-SS include a preamble, in Round 1 a version of the independent living difficulty without a preamble was compared to the ACS-6 version. Overall, respondents who received the version without a preamble appeared more likely to misinterpret the question. This was particularly true among monolingual Spanish speakers, some of whom reported having difficulty doing errands alone because they are not native English speakers, which is out-of-scope (RTI International, 2022a). In Round 2, only the version of the question with a preamble was included in the survey instruments. Results indicated that most respondents understood the question as intended. As such, the Disability Subcommittee recommended that the field test version of the question about independent living difficulty include a preamble.

Finally, the two rounds of cognitive interviews provided an additional opportunity to assess how respondents choose among the four response categories in the WG-SS. Consistent with findings from cognitive interviews and data analyses conducted by NCHS (Weeks et al., 2021), most respondents selected “some difficulty” or “no difficulty” when answering the WG-SS questions, while reports of “a lot of difficulty” and “cannot do at all” were relatively uncommon and restricted to individuals with higher levels of functional difficulty. Notably, the cognitive interviews suggested that the “some difficulty” category captured a more heterogeneous group in terms of functional ability, ranging from individuals with outdated eyeglass prescriptions to adults who currently receive disability benefits.

Overall, these results provide further evidence that the WG-SS measure of disability, when dichotomized based on NCHS recommendations to define disability as “a lot of difficulty” or “cannot do at all”, will likely capture a population with a higher level of functional difficulty than the population captured by the current ACS Disability measure (ACS-6). The WG-SS measure appears to primarily capture individuals who perceive themselves as having an impairment that significantly curtails their ability to complete activities on a daily basis, while

generally excluding individuals who feel their impairment limits their activities less frequently, to a lesser degree, or only in certain contexts. While this approach is expected to result in lower estimates of disability prevalence in the U.S., it may be more consistent with the social model of disability, which conceptualizes disability as arising only if an individual's impairment leads them to encounter barriers to participating in society (Davis, 2006).

### 1.3.3 Question Content

For the field test component of the 2022 ACS Content Test, different versions of the ACS instrument were fielded. This included a Test treatment containing a version of the WG-SS (selected based on the Census Bureau-directed cognitive interviews) and a Control treatment containing the current, production version of the ACS Disability measure (ACS-6). Figure 1 and Figure 2 show the Control and Test versions of the Disability question set as they appeared on the paper questionnaire for the field test. Automated versions of the questionnaire had the same content formatted accordingly for each mode.

**Figure 1. Control Version of the Disability Questions (Paper)**

**18** a. Is this person deaf or does he/she have serious difficulty hearing?

☐ Yes

☐ No

**1** Answer question 20 if this person is 15 years old or over. Otherwise, SKIP to the questions for Person 2 on page 19.

**20** Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?

☐ Yes

☐ No

**H** Answer questions 19a – c if this person is 5 years old or over. Otherwise, SKIP to the questions for Person 2 on page 19.

**19** a. Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?

☐ Yes

☐ No

b. Does this person have serious difficulty walking or climbing stairs?

☐ Yes

☐ No

c. Does this person have difficulty dressing or bathing?

☐ Yes

☐ No



**Figure 2. Test Version of the Disability Questions (Paper)**

<p><b>18</b> a. Does this person have difficulty seeing, even if wearing glasses?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p> <p>b. Does this person have difficulty hearing, even if using a hearing aid?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p> <p><b>H</b> Answer questions 19a – d if this person is 5 years old or over. Otherwise, SKIP to the questions for Person 2 on page 19.</p> <p><b>19</b> a. Does this person have difficulty walking or climbing steps?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p>	<p>b. Does this person have difficulty remembering or concentrating?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p> <p>c. Does this person have difficulty with self-care, such as washing all over or dressing?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p> <p>d. Using his or her usual language, does this person have difficulty communicating, for example understanding or being understood?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p> <p><b>I</b> Answer question 20 if this person is 15 years old or over. Otherwise, SKIP to the questions for Person 2 on page 19.</p> <p><b>20</b> Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?</p> <p><input type="checkbox"/> No difficulty</p> <p><input type="checkbox"/> Some difficulty</p> <p><input type="checkbox"/> A lot of difficulty</p> <p><input type="checkbox"/> Cannot do at all</p>
---	---

The ACS-6 and WG-SS have a number of differences between them, including question wording, answer choices, question ordering, and whether they ask about communication or independent living difficulty. As a result, the Test and Control question sets have the following differences:

- Differences in wording among the questions that capture the same concepts, such as omitting the word “serious” in the Test question set.
  - VISION
    - Control Version: Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?
    - Test Version: Does this person have difficulty seeing, even if wearing glasses?
  - HEARING
    - Control Version: Is this person deaf or does he/she have serious difficulty hearing?
    - Test Version: Does this person have difficulty hearing, even if using a hearing aid?

- COGNITION
  - Control Version: Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?
  - Test Version: Does this person have difficulty remembering or concentrating?
- AMBULATION
  - Control Version: Does this person have serious difficulty walking or climbing stairs?
  - Test Version: Does this person have difficulty walking or climbing steps?
- SELF-CARE
  - Control Version: Does this person have difficulty dressing or bathing?
  - Test Version: Does this person have difficulty with self-care, such as washing all over or dressing?
- Differences in answer choices
  - Control Version: Yes/no
  - Test Version: No difficulty/Some difficulty/A lot of difficulty/Cannot do at all
- An additional question in Test Version asking about communication difficulty,
- Differences in question order:

<b>Control Version</b>	<b>Test Version</b>
1. Hearing	1. Vision
2. Vision	2. Hearing
3. Cognition	3. Ambulation
4. Ambulation	4. Cognition
5. Self-care	5. Self-care
6. Independent living	6. Communication
	7. Independent living

As shown above, the Control version has six questions, while the Test version has seven questions. The ACS question on Independent Living was retained within the Test version, as it meets specific needs of ACS data users. The wording of the Test question on Independent Living is identical to that of the Control version, so that only the answer categories differ across treatment.

Note that Disability question items differ in age universe. Table 1 presents the age universe of each question item. These do not differ between Control and Test versions.

**Table 1. Age Universes by Disability Question**

Question	Ages 0 to 4	Ages 5 to 14	Ages 15+
Hearing	Yes	Yes	Yes
Vision	Yes	Yes	Yes
Cognition	No	Yes	Yes
Ambulation	No	Yes	Yes
Self-Care	No	Yes	Yes
Communication	No	Yes	Yes
Independent Living	No	No	Yes

### 1.3.4 Research Questions

The research questions considered during the field test are presented below.

1. *How do the item missing data rates differ between the Test treatment and the Control treatment (overall, by mode, and by proxy/self-report)?<sup>6</sup>*
2. *How do the estimated proportions of persons with any Disability differ between the Test treatment and the Control treatment (overall and by mode)?*
3. *How do the estimated proportions of persons with each Disability type differ between the Test treatment and the Control treatment (overall and by mode)?*
4. *In the Test treatment, what is the distribution of responses for each Disability question (overall and by mode)?*
5. *How do the measures of response reliability differ between the Test treatment and the Control treatment for overall Disability (overall, by mode of original interview, and by same vs. different CFU respondent)?*
6. *How do the measures of response reliability differ between the Test treatment and the Control treatment by Disability type (overall, by mode of original interview, and by same vs. different CFU respondent)?*
7. *What is the response reliability for the Test treatment, looking at movement between the four response categories (L-fold Index of Inconsistency) for each Disability type?<sup>7</sup>*

---

<sup>6</sup> The responses for the person identified as the household respondent are considered to be self-responses, while the responses for all other household members are considered to be proxy responses. The household respondent is the person who provides data for all members of a household. For the internet and CAPI instruments, the household respondent is identified during the interview and questions are tailored to them by filling “you” or another household member’s name in the question. For the paper questionnaire, the household respondent is the person listed as Person 1 on the questionnaire.

<sup>7</sup> This research question originally read, “How do the measures of response reliability in the Test treatment differ by response category (“no difficulty”; “some difficulty”; “a lot of difficulty”; “cannot do at all”) (overall and by mode of original interview)?” However, during analysis we determined that it would not be informative to

8. *Using internet and CAPI paradata, how does respondent burden differ between the Test treatment and the Control treatment?*

## 2 METHODOLOGY

### 2.1 Sample Design

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

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

In the sample selection process, we selected an initial sample of 40,000 addresses, then selected the two nearest neighbors for each initially selected address. If possible, we selected nearest neighbors that were in both the same content test sampling stratum as well as the same state, county, and sub-county area as the initially selected address. In total, three samples were selected, one for the Control treatment and two for the two test treatments. These three treatments are shown in Table 2.

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

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

---

compare measures of reliability between the individual response categories, because the measures would be distorted by the relative rarity of some response categories (e.g., “cannot do at all”) compared to others. The research question was revised to reflect a measure of response reliability across all four categories.

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.<sup>8</sup> For the Disability analyses contained in this report, we only used data from the Control and Test treatments, so as not to confound any variations within the Roster Test survey with the Control Disability question performance.

**Table 2. 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 <sup>†</sup>
Labor Force	Production	Test Version 1	Test Version 2
Income	Production	Test Version 1	Test Version 2

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

## 2.2 Data Collection

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

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

<sup>8</sup> We examined differences in key household and person characteristics among the Control and Roster Test treatments to explore any indication of bias in the Health Insurance Coverage, Labor Force, and Income analyses. See Spiers et al. (2023) for more information.

(CAPI) operation, where Census Bureau field representatives attempt to complete a survey for a sub-sample of the remaining nonresponding addresses.

The following data collection protocols for the 2022 ACS Content Test remained the same as production ACS:

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

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

- There were no paper versions of the 2022 ACS Content Test questionnaires in Spanish.<sup>9</sup>
- 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.<sup>10</sup> 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.<sup>11</sup>
- 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).

---

<sup>9</sup> 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.

<sup>10</sup> 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.

<sup>11</sup> 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.

## **2.3 Content Follow-Up Operation**

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

### **2.3.1 Content Test Follow-Up Protocol**

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

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

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

The CFU collected an independent household roster by re-asking the Household Roster questions along with Relationship, Sex, Age, and Date of Birth. The remaining CFU questions were only asked of the original household roster members. Only the Control and Roster Test panels collected an independent household roster. The Test panel used the original household roster to ask housing and detailed person questions.<sup>12</sup>

---

<sup>12</sup> 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.

### 2.3.2 Content Test Follow-Up for Disability

For the CFU reinterview, the Disability question series was re-asked of respondents that completed the original interview (or if the original respondent was not available, was re-asked to another household member). Each treatment path of the CFU instrument had the CAPI version of the original interview questions, which was almost identical to the corresponding internet and paper versions. The only difference in the CAPI version is the inclusion of an introductory phrase before the first question in the Disability series informing the respondent that the next questions ask about difficulties that they or a household member may have doing certain activities. Though the phrase is slightly different between treatments, the introduction is present in both Control and Test treatments.<sup>13</sup>

## 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 Disability 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 Disability 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.<sup>14</sup>

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 nonresponse.<sup>15</sup>

Comparisons between the Control and Test versions of Disability were conducted using a two-tailed t-test at the  $\alpha=0.1$  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

---

<sup>13</sup> In the Control version, the CAPI instrument uses the introductory phrase currently used in the ACS: "I am now going to ask some questions about difficulty (you/Name) may have with ordinary daily activities." For the Test version, the CAPI instrument uses the introductory phrase recommended by the Washington Group: "The next questions ask about difficulties (you/Name) may have doing certain activities."

<sup>14</sup> 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.

<sup>15</sup> The Content Test weight creation process does not include all the steps followed in the ACS, including the noninterview adjustment for the original interview and calibration to housing unit and population controls (see U.S. Census Bureau, 2022b, Chapter 11). For more information on the 2022 Content Test weighting procedure, see Risley and Oliver (2022) and Keathley (2022).



(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.<sup>16</sup> 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).<sup>17</sup>

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).<sup>18</sup>

### 2.4.2 Topic-Level Analysis

To evaluate the changes to Disability, we calculated a variety of metrics, presented in Sections 2.4.2.1 through 2.4.2.5.

#### 2.4.2.1 Benchmarks

There is no official source of disability statistics that can serve as a true benchmark for overall disability prevalence or for the prevalence of disability types. A number of surveys collect information on disability. Some use the WG-SS, some use the ACS-6, and others use varieties of questions tailored to the specific population of interest. It is well-established in the disability literature that different question sets, or the same questions in different surveys, particularly those focusing on specific topics, result in different prevalence estimates (Albrecht et al., 2001; Altman & Bernstein, 2008). In addition, as disability exists on a continuum, the choice of a cut point – i.e., the level of difficulty one must experience to be classified as having a disability – will greatly affect prevalence estimates. As such, we had no way to evaluate the field test

<sup>16</sup> See Section 2.4.2.4 for the definition of Gross Difference Rate.

<sup>17</sup> Use the MULTTEST Procedure in SAS®.

<sup>18</sup> As part of the 2022 ACS Content Test, we also analyzed respondent burden. The results of this analysis are contained in Virgile et al. (2023).

results in the context of their consistency (or inconsistency) with benchmarks for “true” disability prevalence. While we did compare estimates of Disability prevalence, both overall and by Disability type, between the Test and Control treatments (see Section 2.4.2.3), this was for informational purposes and was not part of the decision criteria.

Prior analyses by NCHS demonstrated that the WG-SS and the ACS-6 produce different estimates of disability prevalence, and that the WG-SS captures a smaller population with a higher level of functional difficulty, relative to the ACS-6 Disability measure, with an estimate about 50 percent lower when using the recommended Definition 1 (for dichotomizing the graded answer scale).

#### **2.4.2.2 Item Missing Data Rates**

To measure nonresponse to Disability, 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.

We first compared the percentage of eligible people who did not provide a response to any of the Disability questions (i.e., all items are missing) in the Control treatment with the corresponding percentage in the Test treatment. Notably, the Test treatment included seven Disability questions, while the Control treatment included six Disability questions, which may have affected the overall item missing data rate comparison.

We next compared the percentage of eligible people who did not provide a response for each of the Disability questions separately. We compared the item missing data rates by each Disability type instead of by question number, since there are differences in question order between the Control and Test treatments. Ideally, we wanted to examine if the wording differences affected item nonresponse, regardless of question order.

We compared item missing data rates overall, as well as by mode and by household respondent status.<sup>19</sup> As mentioned in Section 1.3.1.5, it was important to evaluate the Test treatment in the self-response modes and situations when the respondent was not answering about themselves because the WG-SS has not been tested extensively in those contexts. Additionally, including the break-down by mode was important, as we have historically observed different patterns of item missingness across modes in the ACS, with higher proportions of select item missingness in self-administered modes (e.g., paper).

We compared item missing data rates via two-tailed t-tests. For all comparisons, we excluded persons that had all detailed person questions missing or persons with early breakoffs (i.e., stopped answering the questionnaire before reaching the Disability questions).

---

<sup>19</sup> For a definition of household respondent, see footnote 6 on page 23.

### 2.4.2.3 Response Distributions

To assess how changes to the Disability question series affected Disability estimates, we compared the response distributions, both overall and for each Disability type, between the Control and Test treatments. We calculated the response distributions as the proportion of valid responses in a category out of all valid responses.

To compare estimates of disability prevalence, we needed to define the category of having a disability. For the Control treatment, a person had a disability if they answered “yes” to the relevant Disability-related question. Respondents who answered “yes” to the question about hearing difficulty, for example, were considered to have a hearing disability, while respondents who answered “no” to the hearing difficulty question did not have this type of disability.

For the Test treatment, a person had four response options on an ordinal response scale. The Washington Group recommends dichotomizing disability status as follows: a person has a disability if they answered “a lot of difficulty” or “cannot do at all” to the relevant Disability item and does not have a disability otherwise. Respondents who answered “a lot of difficulty” or “cannot do at all” to the question about hearing difficulty, for example, would be considered to have a hearing disability, while respondents who answered “no difficulty” or “some difficulty” would not have this type of disability. For the purposes of understanding the field test data, we examined two different definitions of Disability with the expectation that neither categorization of the Test treatment would ‘match’ the yes/no options of the Control treatment. These two definitions are described in Table 3.

**Table 3. Definitions of a Disability in the Control and Test Treatments**

Definition of a Disability	Disability Category	Control Treatment	Test Treatment
<b>Definition 1</b>	With a disability	Percentage of people with “yes”	Percentage of people with “a lot of difficulty” or “cannot do at all”
	Without a disability	Percentage of people with “no”	Percentage of people with “no difficulty” or “some difficulty”
<b>Definition 2<sup>20</sup></b>	With a disability	Percentage of people with “yes”	Percentage of people with “some difficulty”; “a lot of difficulty”; or “cannot do at all”
	Without a disability	Percentage of people with “no”	Percentage of people with “no difficulty”

<sup>20</sup> It should be noted that NCHS also makes use of this second definition, but to identify members of the population with “functional limitations” (as opposed to the population with disabilities, which NCHS identifies using the first

We also constructed an overall measure of Disability. As in current ACS production, a person was defined as having any disability (i.e., of any type) if they were categorized as having one or more of the individual disability types. A person was defined as not having a disability if they were determined not to have any of the individual disability types. Note that the Test treatment included one more Disability question than the Control treatment, which may have affected the comparison of the overall measure of Disability.

We calculated response distributions for each Disability question, overall and by mode and household respondent status. These comparisons were made using two-tailed t-tests. Response distributions comparisons between the Test and Control treatments were for informational purposes and were not a formal part of the decision criteria.

#### **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 Disability, 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:

- Gross difference rate (GDR)
- Index of inconsistency (IOI)
- L-fold index of inconsistency (IOI<sub>L</sub>)

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 4, “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.

---

definition in Table 3). For more information, see the following webpage:

<https://www.cdc.gov/nchs/hus/topics/functional-limitation.htm#:~:text=Functional%20limitation%20is%20defined%20by,cognition%2C%20and%20self%2Dcare>.

**Table 4. 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. For each individual Disability question item, for example, the Control treatment had two response categories (yes/no) while the Test treatment has four response categories. As such, the response categories for the Test questions were collapsed into yes/no using either the first or second definition of a disability as described in Table 3. Then GDR and IOI were calculated as discordance in yes/no between the original interview and reinterview.

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 also measured the overall response reliability of a question using the L-fold index of inconsistency, IOI<sub>L</sub>. We looked at the difference in IOI<sub>L</sub> between treatments and tested for significance as with the single category measures.

Suppose a question has L response categories. Let X<sub>ij</sub> 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 5 shows a cross-tabulation of the original interview and CFU results for a generic analysis topic. Note that if L = 2, then Table 5 is equivalent to Table 4.

**Table 5. Cross-Tab of Original Interview and CFU Results: Questions with Response Categories**

		Original Interview categories						CFU totals
		1	2	...	j	...	L	
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

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

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):

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

The response reliability measures described in this section (GDR, IOI) were calculated for both overall Disability and for each Disability type, and for the three original interview modes individually as well as overall. We also calculated IOI<sub>L</sub> for each Disability type in the Test treatment.

While not part of the decision criteria, we also compared measures of response reliability between the two Test treatment definitions of Disability to better examine how each definition functioned in the context of the ACS instrument. That is, whether response reliability in the Test treatment is higher when we use one definition of Disability over another.

Finally, we calculated reliability measures separately by whether or not the CFU responses were reported by the same individual who completed the original interview or a new respondent.

#### 2.4.2.5 Other Metrics

Changes to survey instruments should not increase respondent burden unnecessarily, since high levels of respondent burden can lead to measurement error. To consider all available information in assessing the impact on respondent burden, we made use of paradata from the internet and CAPI instruments.

Greater usage of the help text on the internet and CAPI instruments may be considered an indicator of respondent burden. That is, respondents who are confused by a question may be more likely to click on help text than are respondents who understand a question as initially asked. Thus, we compared the percentage of people using help text (on any screen within the Disability section) on the internet version of the Test treatment to the corresponding percentage for the internet version of the Control treatment.<sup>21</sup>

In addition, respondents who are confused by a question or unsure of how to answer may spend a greater amount of time on that question. We thus compared the median completion time for the Disability questions between treatments, for the internet mode.<sup>22</sup>

---

<sup>21</sup> The original research plan included comparing help text use for the CAPI instrument as well as internet. However, we only performed this analysis for the internet instrument because use of help text in CAPI may be an effect of interviewer behavior more than respondent behavior.

<sup>22</sup> The original research plan included other measures of respondent burden: switching answer choices, moving back and forth between screens, and completion times for each individual Disability question. Due to time



Finally, if a respondent finds a particular question too confusing or overly sensitive, it may cause them to stop answering the questionnaire entirely, i.e., a breakoff. If significantly more or fewer respondents broke off during the Test Disability questions as compared to the Control questions, that could indicate a difference in respondent burden. To check for this, we compared the breakoff rate—that is, the proportion of respondents that logged into the internet instrument but did not reach the “Submit” action, and whose final action in the instrument took place on a screen within the Disability series—between the Test and Control treatments.

For all research questions, statistical significance between treatments was determined using a two-tailed t-test. For all comparisons, we excluded persons that had all detailed person questions missing or persons with early breakoffs (i.e., stopped answering the questionnaire before reaching the Disability questions). Note that these additional evaluations of respondent burden were not part of the formal decision criteria.

### **3 DECISION CRITERIA**

Before field testing Disability, a team of subject matter experts identified and prioritized which of the research questions presented in Section 1.3.4, would determine which version of Disability would be recommended for inclusion in the ACS. The decision criteria for Disability are presented in Table 6.

---

constraints in the analysis phase of this test, we determined that these metrics would be less informative than the other measures of respondent burden and did not analyze them at this time.

**Table 6. Decision Criteria for Disability**

Priority	Research Question*	Decision Criteria
1	5, 6	<p>A lower GDR or IOI indicates higher reliability and is preferable. Because we expect more variability between individual responses in the Test treatment (four choices offer more opportunity for switching), the decision criteria will rely on a comparison of the GDR/IOI for the summary measure of overall Disability– “with a disability” or “without a disability”. Although a lower GDR or IOI is preferable, lower reliability in the Test treatment may be acceptable when considering the documented cognitive testing validity of the Test treatment.</p> <p>Because the self-response option was not available in earlier tests of the proposed questions (Test treatment), it is particularly important that the magnitude of the difference in GDR and IOI between the self-response modes (internet and paper) and the personal interview mode (CAPI) be similar across the Test and Control treatments.</p>
2	1	<p>In general, lower item missing data rates are preferable. However, the Test treatment could have higher item missing data rates than the Control treatment simply because of the more complex response choices and the addition of the question about communication difficulty (more items, more chance of missing data). Thus, a small increase in item missing data rates is an acceptable consequence.</p> <p>Because the self-response option was not available in earlier tests of the proposed questions (Test treatment), it is particularly important that the magnitude of the difference in item missing data rates between the self-response modes (internet and paper) and the personal interview mode (CAPI) be similar across the Test and Control treatments.</p>

\*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. The unit analysis results

indicate that this assumption was met for overall response. Details can be found in Section 5.

- There was no difference between treatments in mail delivery timing or subsequent response time. The treatments had the same sample size and used the same postal sort and mailout procedures. Previous research indicated that postal procedures alone could cause a difference in response rates at a given point in time between experimental treatments of different sizes, with response for the smaller treatments lagging (Heimel, 2016).
- We assume that the frequency of real changes in answers due to a change in life circumstances between the original interview and CFU reinterview were similar between treatments.

## **4.2 Limitations**

- 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 the 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 (SE) = 0.2) of households from the original Control interviews, 2.5 percent (SE = 0.2) of households from the original Test interviews, and 3.0 percent (SE = 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 (SE = 0.4) of CFU Control cases, 9.4 percent (SE = 0.5) of CFU Test cases, and 8.5 percent (SE = 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. The sample for the test also over-sampled addresses in low response areas. As a result, any estimates derived from the Content Test data did not provide the same level of inference as the production ACS and cannot be compared to production estimates.

- Comparisons between the Test and Control treatments were limited in some cases by differences in question order and in the number of items in the Disability question series (six questions in the Control treatment, compared to seven in the Test treatment). Differences in the number of items in the question series, for example, could impact estimates of the proportion of people who were missing for all Disability questions, as well as estimates of the proportion of people with any disability. As such, we were cautious in our interpretation of any observed differences in these metrics.
- All estimates based on the construct of self-report versus proxy response are limited due to potential error in the identification of the respondent. For example, we do not know whether the person filling out the internet form is answering for other people or recording answers provided by the other household members. Due to these limitations, analyses regarding self-report versus proxy responses were only used to understand the data and were not used in the decision criteria.
- A limitation to comparisons by mode of survey administration is that respondents self-selected into mode. As such, differences by mode may arise due to not only mode effects, but also differences in respondent characteristics. It is not possible to disentangle these two potential causes.

## 5 RESULTS

This section of the report presents the results of various metrics used to evaluate Disability. The comparisons presented assume comparable samples, as evidenced by the homogeneity of the unit response rates for the three treatments, prior to the field test. We tested this assumption via unit-level (i.e., address level) analyses. The full results are presented in Spiers et al. (2023); the salient results for Disability are summarized here.

Table 7 presents the unit response rates in the original interview, overall and by mode. When comparing the Control and Test treatments in the original interview, there were no statistically significant differences in unit response rates, either for all modes combined or each mode separately. We are confident that there are no underlying response rate concerns that would impact topic-specific original interview comparisons between treatments.

**Table 7. Original Interview Unit Response Rates for Control and Test Treatments, Overall and Distributed by Mode**

Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Difference	Adj. P-value
Overall	19,000	83.3 (0.4)	19,000	83.6 (0.4)	-0.2 (0.5)	0.66
Self-Response	15,000	59.2 (0.4)	15,000	59.6 (0.4)	-0.3 (0.6)	0.66
Internet	12,000	48.4 (0.4)	11,500	47.7 (0.4)	0.7 (0.6)	0.66
Mail	3,500	12.3 (0.3)	3,700	13.1 (0.3)	-0.8 (0.4)	0.35
CAPI	4,200	24.4 (0.4)	4,100	24.2 (0.4)	0.2 (0.5)	0.66

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0058

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

We also compared the distributions of demographic and socioeconomic characteristics (age, sex, Hispanic origin, race, building tenure, and language of response) between treatments. There were no statistically significant differences between treatments for any of the examined variables, both for all modes combined and self-response. When looking at CAPI responses, the distribution of race was significantly different between the Test and Control treatments ( $p$ -value = 0.02). When t-tests were performed on the individual response categories, the only significant difference was in the “Some Other Race” category. The proportion of respondents in this category was 3.6 percentage points lower (SE = 1.6,  $p$ -value = 0.09) in Test than Control. This difference should be kept in mind when interpreting the CAPI results.

Aside from race in the CAPI mode, the results of the unit analysis do not present any concerns for analysis of the Disability data.

### 5.1 Benchmark Results for Disability

As discussed in Section 2.4.2.1, there were no benchmarks used in the Disability analysis.

### 5.2 Item Missing Data Rate Results for Disability

Given expected differences in response distributions between the Test and Control treatments and the absence of a benchmark for disability, item missing data rates were a key criterion for evaluating the Test version of Disability. For details of the item missing data rate calculations, see Section 2.4.2.2.

*RQ 1. How do the item missing data rates differ between the Test treatment and the Control treatment (overall, by mode, and by proxy/self-report)?*

#### 5.2.1 Item Missing Data Rates

When evaluating the two versions of the Disability question set, we considered a measure with lower item missing data rates preferable. However, the Test treatment had more response categories, as well as one more question item, than the Control. It was expected that these differences could lead to higher item missing data rates in the Test treatment. As such, it was

established a priori that a small increase in item missing data rates in the Test treatment might be acceptable.

The subsections that follow present the item missing data rates overall, by mode, and for each Disability type.

#### **5.2.1.1 Item Missing Data Rates for All Disability Questions**

The item missing data rate for all Disability questions (Table 8) measures the proportion of respondents who skipped *all* Disability questions for which they were in-universe.<sup>23</sup> For simplicity, we will sometimes refer to this rate as the item missing data rate for Disability overall (as opposed to the item missing data rate for specific Disability types). When all response modes were combined, the item missing data rate for Disability overall was significantly higher in the Test treatment than the Control treatment. The magnitude of the difference in item missing data rates between the Test treatment and Control was 1.0 percentage point (SE = 0.4; 6.1 percent (SE = 0.3) for the Test treatment vs. 5.1 percent (SE = 0.2) for the Control).

We next looked at the item missing data rate for Disability overall by mode. The item missing data rate was significantly higher in the Test treatment compared to the Control treatment for self-response overall (i.e., mail and internet combined) (6.1 percent (SE = 0.3) in Test vs. 5.3 percent (SE = 0.2) in Control) and for internet specifically (7.0 percent (SE = 0.3) in Test vs. 6.0 percent (SE = 0.3) in Control), although again these differences were 1.0 percentage point or less. There were no significant item missing data rate differences between Test and Control in the CAPI or mail modes.

Since the WG-SS was initially developed and tested in interviewer-administered surveys, while the ACS is most frequently completed via paper or internet, it was important to assess whether the Test treatment (based on the WG-SS) performed better or worse in self-response modes, relative to its performance in CAPI. Nominally, the gap in item missing data rates between Test and Control for Disability overall did not appear to be any larger in self-response modes (6.1 percent (SE = 0.3) vs. 5.3 percent (SE = 0.2), adjusted p-value = 0.05), relative to the gap between Test and Control in CAPI (6.0 percent (SE = 0.8) vs. 4.4 percent (SE = 0.7), adjusted p-value = 0.23).

---

<sup>23</sup> Note that respondents who exited the survey early, before reaching the Disability section, were not considered in-universe.

**Table 8. Item Missing Data Rates: All Disability Questions**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.1 (0.3)	5.1 (0.2)	1.0 (0.4)	0.04*
Self-Response	6.1 (0.3)	5.3 (0.2)	0.8 (0.3)	0.05*
Internet	7.0 (0.3)	6.0 (0.3)	1.0 (0.4)	0.04*
Mail	1.5 (0.3)	2.0 (0.3)	-0.5 (0.4)	0.23
CAPI	6.0 (0.8)	4.4 (0.7)	1.6 (1.1)	0.23

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### 5.2.1.2 Item Missing Data Rates for Individual Disability Questions

Table 9 through Table 15 present item missing data rates for the individual Disability questions.

The item missing data rate for the Hearing question was significantly higher in the Test treatment than in Control when all response modes were combined (6.6 percent (SE = 0.3) vs. 5.5 percent (SE = 0.2)), as well as when self-response modes were combined (6.8 percent (SE = 0.3) vs. 5.8 percent (SE = 0.2)) (Table 9).<sup>24</sup> When each response mode was considered individually, however, only within the internet mode were item missing data rates for Hearing significantly higher in Test, relative to Control (no differences between treatments were detected within the mail or CAPI modes).

For the Vision (Table 10), Ambulation (Table 12), Self-Care (Table 13), and Independent Living (Table 14) questions, no significant differences were detected between treatments when all modes were combined. When analysis was broken down by mode, however, significant differences were detected within the mail mode. For all four Disability types, the differences favored the Test treatment. That is, within mail mode, the Test treatment had significantly lower item missing dates than Control (Vision: 2.4 percent (SE = 0.3) vs. 4.6 percent (SE = 0.4); Ambulation: 3.2 percent (SE = 0.3) vs. 5.2 percent (SE = 0.4); Self-Care: 3.1 percent (SE = 0.4) vs. 5.2 percent (SE = 0.5); Independent Living: 4.0 percent (SE = 0.4) vs. 6.2 percent (SE = 0.6)).

Finally, item missing data rates for the Cognition question did not significantly differ between Test and Control when all modes were combined (Table 11). Within individual modes, however, a couple of significant differences were detected, one favoring the Test treatment and the other favoring Control. Specifically, in the internet mode, the Test treatment had a significantly higher item missing data rate than the Control treatment (8.0 percent (SE = 0.3) vs. 6.9 percent (SE = 0.3)), while in the mail mode, the Test treatment had a significantly lower item missing data rate than Control (3.2 percent (SE = 0.4) vs. 5.4 percent (SE = 0.4)).

<sup>24</sup> As described in Section 2.4.2.2, the Hearing question was ordered as first within the Control treatment, and as second within the Test treatment.



**Table 9. Item Missing Data Rates: Hearing**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.6 (0.3)	5.5 (0.2)	1.1 (0.4)	0.02*
Self-Response	6.8 (0.3)	5.8 (0.2)	1.0 (0.4)	0.02*
Internet	7.6 (0.3)	6.2 (0.3)	1.3 (0.4)	0.01*
Mail	2.7 (0.3)	3.6 (0.4)	-0.9 (0.5)	0.13
CAPI	6.1 (0.8)	4.5 (0.7)	1.7 (1.1)	0.13

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 10. Item Missing Data Rates: Vision**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.4 (0.3)	5.7 (0.2)	0.7 (0.4)	0.25
Self-Response	6.5 (0.3)	6.2 (0.2)	0.4 (0.4)	0.28
Internet	7.3 (0.3)	6.5 (0.3)	0.9 (0.4)	0.13
Mail	2.4 (0.3)	4.6 (0.4)	-2.2 (0.5)	<0.01*
CAPI	6.1 (0.8)	4.4 (0.7)	1.7 (1.1)	0.25

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 11. Item Missing Data Rates: Cognition**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.9 (0.3)	6.1 (0.2)	0.8 (0.4)	0.15
Self-Response	7.2 (0.3)	6.6 (0.2)	0.5 (0.4)	0.27
Internet	8.0 (0.3)	6.9 (0.3)	1.1 (0.4)	0.04*
Mail	3.2 (0.4)	5.4 (0.4)	-2.2 (0.6)	<0.01*
CAPI	6.1 (0.8)	4.5 (0.6)	1.6 (1.1)	0.27

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 12. Item Missing Data Rates: Ambulation**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.8 (0.3)	6.1 (0.3)	0.7 (0.4)	0.24
Self-Response	7.0 (0.3)	6.6 (0.2)	0.4 (0.4)	0.25
Internet	7.8 (0.3)	6.9 (0.3)	0.9 (0.4)	0.15
Mail	3.2 (0.3)	5.2 (0.4)	-2.0 (0.6)	<0.01*
CAPI	6.2 (0.8)	4.5 (0.7)	1.7 (1.1)	0.25

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 13. Item Missing Data Rates: Self-Care**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	6.9 (0.3)	6.2 (0.3)	0.7 (0.4)	0.31
Self-Response	7.1 (0.3)	6.8 (0.3)	0.4 (0.4)	0.33
Internet	7.9 (0.3)	7.1 (0.3)	0.8 (0.4)	0.19
Mail	3.1 (0.4)	5.2 (0.5)	-2.1 (0.6)	<0.01*
CAPI	6.1 (0.8)	4.5 (0.7)	1.6 (1.1)	0.31

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 14. Item Missing Data Rates: Independent Living**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	7.5 (0.3)	6.8 (0.3)	0.7 (0.4)	0.20
Self-Response	7.9 (0.3)	7.5 (0.3)	0.4 (0.4)	0.40
Internet	8.7 (0.3)	7.8 (0.3)	0.9 (0.5)	0.20
Mail	4.0 (0.4)	6.2 (0.6)	-2.3 (0.7)	0.01*
CAPI	6.1 (0.8)	4.3 (0.6)	1.8 (1.1)	0.20

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 15. Item Missing Data Rates: Communication (Test Version Only)**

Mode	Test Percent
Overall	6.9 (0.3)
Self-Response	7.1 (0.3)
Internet	7.9 (0.3)
Mail	3.2 (0.4)
CAPI	6.1 (0.8)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

## 5.2.2 Item Missing Data Rates for Proxy versus Self-Report Responses

Finally, we compared item missing rates between Test and Control for proxy responses versus self-report responses.<sup>25</sup> Proxy responses are provided when the household respondent answers questions on behalf of other household members, rather than for themselves. The WG-SS (on which the Test treatment is based) has not been as extensively tested in the proxy response context. As such, it was important to assess whether the Test treatment performed better or worse for proxy responses, relative to self-reports.

### 5.2.2.1 Item Missing Data Rates for All Disability Questions, Proxy versus Self-Report Responses

Table 16 and Table 17 present the item missing data rates for Disability overall, by type of response and mode. When people responded for themselves (self-report), there was no difference between Test and Control in the percentage of respondents who were missing data on all Disability questions. This was true when all modes were combined, as well as within each mode. In contrast, when people answered questions about other people in their household (proxy responses), the item missing data rate for Disability overall was significantly higher in the Test treatment when all modes were combined (7.5 percent (SE = 0.4) vs. 6.1 percent (SE = 0.3)) and when all self-response modes were combined. When each mode was considered individually, however, only within the internet mode were item missing data rates significantly higher in Test, relative to Control.

<sup>25</sup> For definitions of proxy versus self-report responses, see footnote 6 on page 23.

**Table 16. Item Missing Data Rates for All Disability Questions: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	4.2 (0.2)	3.7 (0.2)	0.5 (0.3)	0.41
Self-Response	4.2 (0.2)	3.8 (0.2)	0.4 (0.3)	0.41
Internet	4.7 (0.3)	4.2 (0.2)	0.5 (0.4)	0.41
Mail	1.9 (0.3)	2.1 (0.4)	-0.2 (0.5)	0.66
CAPI	4.3 (0.6)	3.3 (0.5)	1.1 (0.8)	0.41

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 17. Item Missing Data Rates for All Disability Questions: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	7.5 (0.4)	6.1 (0.3)	1.4 (0.5)	0.03*
Self-Response	7.7 (0.3)	6.5 (0.3)	1.2 (0.5)	0.04*
Internet	8.8 (0.4)	7.2 (0.3)	1.5 (0.5)	0.03*
Mail	1.1 (0.3)	1.9 (0.4)	-0.7 (0.5)	0.15
CAPI	7.2 (1.1)	5.1 (0.8)	2.0 (1.4)	0.15

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### 5.2.2.2 Item Missing Data Rates for Individual Disability Questions, Proxy versus Self-Report Responses

Finally, we examined item missing data rates for the individual Disability items by proxy responses versus self-report responses. Results are presented in Table 18 through Table 31.

Among self-report responses, there were very few significant differences between the Test treatment and Control when analysis was broken down by disability type. The differences that were detected were all within the mail mode and favored the Test treatment over the Control. Specifically, in mail mode, item missing data rates were lower in Test (relative to Control) for Vision, Cognition, Self-Care, and Independent Living.

Results looked different for proxy responses. While Test had lower item missing data rates than Control for four disability types in mail mode, in many other cases Test had higher item missing data rates than Control. For example, Test had higher item missing data rates than Control for Hearing, Vision, and Cognition when all response modes were combined, as well as within internet mode. Test also had higher item missing data rates than Control for Hearing when all self-response modes were combined and for Ambulation within internet mode.

**Table 18. Item Missing Data Rates for Hearing: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	4.8 (0.3)	4.1 (0.2)	0.7 (0.3)	0.22
Self-Response	4.9 (0.2)	4.3 (0.2)	0.6 (0.3)	0.24
Internet	5.2 (0.3)	4.5 (0.2)	0.7 (0.4)	0.22
Mail	3.4 (0.4)	3.6 (0.5)	-0.1 (0.6)	0.83
CAPI	4.5 (0.6)	3.4 (0.5)	1.1 (0.9)	0.45

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 19. Item Missing Data Rates for Hearing: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	8.0 (0.4)	6.5 (0.3)	1.5 (0.5)	0.02*
Self-Response	8.3 (0.4)	7.0 (0.3)	1.3 (0.5)	0.02*
Internet	9.4 (0.4)	7.6 (0.4)	1.8 (0.6)	0.01*
Mail	1.9 (0.4)	3.7 (0.5)	-1.7 (0.7)	0.02*
CAPI	7.2 (1.1)	5.1 (0.8)	2.1 (1.4)	0.13

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 20. Item Missing Data Rates for Vision: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	4.6 (0.2)	4.5 (0.2)	0.1 (0.3)	0.80
Self-Response	4.6 (0.2)	4.8 (0.2)	-0.2 (0.3)	0.80
Internet	5.0 (0.3)	4.7 (0.3)	0.3 (0.4)	0.80
Mail	2.9 (0.3)	5.4 (0.6)	-2.4 (0.6)	<0.01*
CAPI	4.5 (0.6)	3.4 (0.5)	1.1 (0.9)	0.80

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 21. Item Missing Data Rates for Vision: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	7.9 (0.4)	6.7 (0.3)	1.2 (0.5)	0.08*
Self-Response	8.1 (0.3)	7.2 (0.3)	0.9 (0.5)	0.13
Internet	9.1 (0.4)	7.8 (0.3)	1.3 (0.5)	0.06*
Mail	1.9 (0.4)	3.8 (0.5)	-1.9 (0.6)	0.02*
CAPI	7.3 (1.1)	5.1 (0.8)	2.1 (1.4)	0.13

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 22. Item Missing Data Rates for Cognition: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	5.1 (0.2)	4.8 (0.2)	0.3 (0.3)	0.81
Self-Response	5.2 (0.2)	5.2 (0.2)	0.1 (0.4)	0.88
Internet	5.6 (0.3)	5.0 (0.3)	0.6 (0.4)	0.60
Mail	3.8 (0.4)	5.9 (0.6)	-2.1 (0.7)	0.02*
CAPI	4.4 (0.6)	3.4 (0.5)	1.1 (0.8)	0.60

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 23. Item Missing Data Rates for Cognition: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	8.5 (0.4)	7.3 (0.3)	1.2 (0.6)	0.09*
Self-Response	8.9 (0.4)	7.9 (0.3)	1.0 (0.5)	0.12
Internet	9.9 (0.4)	8.4 (0.4)	1.5 (0.6)	0.03*
Mail	2.6 (0.5)	4.9 (0.6)	-2.3 (0.8)	0.02*
CAPI	7.4 (1.1)	5.4 (0.8)	2.0 (1.5)	0.17

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 24. Item Missing Data Rates for Ambulation: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	5.0 (0.3)	4.7 (0.2)	0.3 (0.4)	0.89
Self-Response	5.1 (0.2)	5.1 (0.2)	0.1 (0.4)	0.89
Internet	5.4 (0.3)	5.0 (0.3)	0.4 (0.4)	0.84
Mail	4.0 (0.4)	5.5 (0.6)	-1.6 (0.8)	0.27
CAPI	4.5 (0.6)	3.5 (0.5)	1.0 (0.9)	0.84

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 25. Item Missing Data Rates for Ambulation: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	8.4 (0.4)	7.2 (0.3)	1.1 (0.6)	0.13
Self-Response	8.7 (0.4)	7.9 (0.3)	0.8 (0.5)	0.13
Internet	9.8 (0.4)	8.5 (0.4)	1.3 (0.6)	0.10*
Mail	2.3 (0.4)	4.8 (0.5)	-2.5 (0.7)	<0.01*
CAPI	7.4 (1.1)	5.2 (0.8)	2.2 (1.5)	0.13

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 26. Item Missing Data Rates for Self-Care: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	5.0 (0.2)	4.8 (0.2)	0.1 (0.3)	0.87
Self-Response	5.1 (0.2)	5.2 (0.2)	-0.1 (0.3)	0.87
Internet	5.5 (0.3)	5.1 (0.3)	0.5 (0.4)	0.87
Mail	3.4 (0.4)	5.7 (0.6)	-2.2 (0.8)	0.02*
CAPI	4.4 (0.6)	3.6 (0.6)	0.8 (0.9)	0.87

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 27. Item Missing Data Rates for Self-Care: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	8.5 (0.4)	7.4 (0.4)	1.1 (0.6)	0.15
Self-Response	8.9 (0.4)	8.1 (0.4)	0.8 (0.5)	0.16
Internet	9.9 (0.4)	8.7 (0.4)	1.2 (0.6)	0.15
Mail	2.7 (0.6)	4.8 (0.5)	-2.0 (0.8)	0.05
CAPI	7.4 (1.1)	5.2 (0.8)	2.2 (1.5)	0.16

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 28. Item Missing Data Rates for Independent Living: Self-Report**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	5.2 (0.3)	5.1 (0.2)	0.1 (0.3)	0.88
Self-Response	5.4 (0.3)	5.6 (0.3)	-0.2 (0.4)	0.88
Internet	5.7 (0.3)	5.3 (0.3)	0.4 (0.4)	0.88
Mail	4.4 (0.4)	7.0 (0.7)	-2.6 (0.7)	<0.01*
CAPI	4.4 (0.6)	3.4 (0.5)	1.0 (0.8)	0.86

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 29. Item Missing Data Rates for Independent Living: Proxy**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	9.9 (0.4)	8.5 (0.4)	1.4 (0.6)	0.12
Self-Response	10.5 (0.5)	9.5 (0.4)	1.0 (0.6)	0.12
Internet	11.9 (0.5)	10.4 (0.5)	1.5 (0.7)	0.12
Mail	3.5 (0.6)	5.3 (0.7)	-1.8 (1.0)	0.12
CAPI	7.9 (1.2)	5.2 (0.8)	2.7 (1.5)	0.12

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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



**Table 30. Item Missing Data Rates for Communication: Self-Report (Test Version Only)**

Mode	Test Percent
<b>Overall</b>	5.0 (0.2)
Self-Response	5.1 (0.2)
Internet	5.5 (0.3)
Mail	3.5 (0.4)
CAPI	4.4 (0.6)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 31. Item Missing Data Rates for Communication: Proxy (Test Version Only)**

Mode	Test Percent
<b>Overall</b>	8.5 (0.4)
Self-Response	8.8 (0.4)
Internet	9.8 (0.4)
Mail	2.9 (0.7)
CAPI	7.4 (1.1)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

### 5.2.3 Summary of Item Missing Data Rate Results

In summary, the percentage of respondents who were missing data on all Disability questions was significantly higher in the Test treatment, relative to the Control treatment, when all modes (and proxy vs. self-report responses) were combined. However, the magnitude of this difference was small (1 percentage point) and results varied when analysis was broken down by mode, Disability type, and proxy vs. self-report responses.

In terms of analysis by mode, the percentage of respondents missing data on all Disability questions was significantly higher in Test, relative to Control, within the internet mode and when all self-response modes were combined, but not within the mail or CAPI modes. When each Disability type was considered on its own, some questions had higher item missing rates in the Test treatment compared with Control, particularly in the internet mode, while others had *lower* item missing data rates in Test, relative to Control, in the mail mode.

In particular, the item missing data rate for the Hearing question was significantly higher in the Test treatment than in Control for all response modes combined, for the self-response modes combined, and in the internet mode.<sup>26</sup> The item missing data rate for Cognition was also higher in the Test treatment than in the Control treatment, though only in the internet mode. On the other hand, in the mail mode, five of the seven Disability questions had significantly *lower* item

<sup>26</sup> As described in Section 2.4.2.2, the Hearing question was ordered as first within the Control treatment, and as second within the Test treatment.

missing data rates in the Test treatment than in Control. These questions included Vision, Cognition, Ambulation, Self-Care, and Independent Living.

Given that the WG-SS has not been extensively tested within proxy response contexts, we next examined results for proxy responses versus self-report responses. Notably, gaps between the Test treatment and Control were more frequently significant for proxy responses than for self-report responses. Among self-report responses, no significant difference was detected between Test and Control for Disability overall, regardless of mode. When analysis was broken down by Disability type, the only significant differences detected were in the mail mode and they favored the Test treatment (i.e., lower item missing rates in Test, relative to Control).

In contrast, among proxy responses, the item missing data rate for Disability overall was significantly *higher* in Test, relative to Control, when all modes were combined, when all self-response modes were combined, and within the internet mode. When item missing data rates among proxy responses were broken down by Disability type, three Disability types had higher item missing data rates in Test than Control when all modes were combined. That being said, when analysis was disaggregated by both Disability type and mode, results were more mixed. While item missing data rates were higher in Test, relative to Control, for four Disability types within the internet mode (Hearing, Vision, Cognition, and Ambulation), within the mail mode, these four Disability types had *lower* item missing data rates in Test, compared to Control. Notably, among proxy responses, when item missing data rates were higher in Test than Control, they were generally no more than 1 or 2 percentage points higher.

### **5.3 Response Distribution Results for Disability**

Disability within a population is a complex phenomenon that can be measured in a variety of ways, and different subpopulations may be captured depending on the Disability measure that is used. Based on existing research, the Test version of the Disability question set is expected to produce different estimates of disability prevalence compared to the Control version. This section of the report examines differences in response distributions between the Test treatment and Control to demonstrate the ways in which the measures diverge. The response distribution results are intended to provide insight into how the two treatments differ in disability measurement, as well as how the proposed change would affect the ACS Disability time series. The response distribution results are not part of the decision criteria for Disability.

#### **5.3.1 Response Distributions for Overall Disability**

*RQ 2. How do the estimated proportions of persons with any disability differ between the Test treatment and the Control treatment (overall and by mode)?*

Prior comparisons of the WG-SS (on which the Test treatment is based) and the current Disability measure in ACS (ACS-6) (used in the Control treatment) found that, when using the five questions with overlapping domains (and thus excluding the WG-SS question on Communication and the ACS-6 question on Independent Living), the estimated proportion of

persons with any disability as measured by the WG-SS was about 50 percent lower than the corresponding ACS estimate (Weeks et al., 2021). As such, we expected prevalence estimates in the Test treatment using the definition of Disability recommended by the Washington Group (Definition 1) to be about half the size of the estimates in the Control treatment.

In addition to the definition of Disability recommended by the Washington Group, the graded response categories in the Test treatment allow for other definitions of Disability. In the subsections that follow, the Test treatment is also compared with the Control treatment using a second, broader definition of Disability (Definition 2) (see Table 3).

#### **5.3.1.1 Distributions for Overall Disability Using Definition 1 of Disability**

Table 32 presents the distribution of persons with a disability according to Definition 1. Under Definition 1, respondents in the Control treatment were considered to have a disability if they answered “yes” to at least one question in the series. Respondents in the Test treatment were considered to have a disability if they answered “a lot of difficulty” or “cannot do at all” to at least one question in the series.

When Definition 1 of Disability was used, the percentage of respondents with a disability was significantly lower in the Test treatment (8.1 percent (SE = 0.2)) compared with the Control treatment (13.9 percent (SE = 0.3)). In every mode, the Disability estimate in the Test treatment was significantly lower than in the Control. This result is consistent with previous research, given the differences in how the two treatments measure disability. In the prior NCHS evaluation of the WG-SS question set, the WG-SS estimate of disability prevalence was about half the size of the ACS-6 estimate; in the 2022 ACS Content Test, the Test estimate (which is based on the WG-SS) was about 40 percent lower than the Control (ACS-6) estimate. As such, ACS Content Test results are nominally in line with previous results.

As previously mentioned, the WG-SS has not been extensively tested in self-response modes. Results from this field test suggest that when Definition 1 of Disability is employed, the difference between Test and Control in the percentage of respondents with a disability is in the same direction and of a nominally similar magnitude in self-response modes, relative to CAPI. In other words, the difference between Test and Control estimates of overall disability prevalence does not appear to be larger in self-response modes, relative to within the interviewer-administered mode (CAPI).

**Table 32. Distribution of Overall Disability: Definition 1**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	8.1 (0.2)	13.9 (0.3)	-5.8 (0.4)	<0.01*
Self-Response	7.7 (0.2)	13.0 (0.2)	-5.3 (0.3)	<0.01*
Internet	7.0 (0.2)	11.4 (0.2)	-4.5 (0.3)	<0.01*
Mail	11.5 (0.9)	20.4 (0.9)	-8.9 (1.0)	<0.01*
CAPI	9.2 (0.7)	16.8 (0.7)	-7.6 (0.9)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### 5.3.1.2 Distributions for Overall Disability Using Definition 2 of Disability

Table 33 presents the distribution of persons with a disability according to Definition 2 of Disability. For the Control treatment, Definition 2 was identical to Definition 1. For the Test treatment, a person was counted as having a disability under Definition 2 if they answered “some difficulty,” “a lot of difficulty,” or “cannot do at all” to at least one question in the series. It should be noted that NCHS also makes use of this second definition, but to identify members of the population with “functional limitations” (as opposed to the population with disabilities, which NCHS identifies using Definition 1).

When using Definition 2 and when all modes were combined, the percentage of people with a disability was significantly higher in Test than in Control. In other words, regardless of the definition of Disability employed, the Test and Control estimates of disability prevalence significantly differed from each other: the Test estimate was lower than Control (-5.8 percentage points (SE = 0.4)) when Definition 1 was used, and the Test estimate was higher than Control (17.8 percentage points (SE = 0.5)) when Definition 2 was used.

Looking at differences by mode, when Definition 2 was used, differences between Test and Control estimates ranged from 9.9 percentage points (SE = 1.3) in the CAPI mode to 25.0 percentage points (SE = 1.6) in the mail mode. The differences by mode using Definition 1 (Table 32) had ranged from -4.5 percentage points (SE = 0.3) to -8.9 percentage points (SE = 1.0). In other words, in every mode, significantly more people were identified as having a disability in Test than in Control when Definition 2 of Disability was used, while significantly fewer people were identified as having a disability in Test, relative to Control, when Definition 1 was used.

**Table 33. Distribution of Overall Disability: Definition 2**

Mode	Test Percent	Control Percent	Difference	Adj. P-value
<b>Overall</b>	31.7 (0.4)	13.9 (0.3)	17.8 (0.5)	<0.01*
Self-Response	33.4 (0.5)	13.0 (0.2)	20.3 (0.6)	<0.01*
Internet	30.9 (0.5)	11.4 (0.2)	19.4 (0.6)	<0.01*
Mail	45.3 (1.4)	20.4 (0.9)	25.0 (1.6)	<0.01*
CAPI	26.7 (1.1)	16.8 (0.7)	9.9 (1.3)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### 5.3.2 Response Distributions by Disability Type

*RQ 3. How do the estimated proportions of persons with each Disability type differ between the Test treatment and the Control treatment (overall and by mode)?*

Table 34 through Table 40 present the percentage of people with each disability type, by treatment and mode. In the Control treatment, a respondent was considered to have a specific disability type if they answered “yes” to the question about that disability type. Respondents who said “yes” to the Hearing question in the Control, for example, were considered to have hearing disability. In the Test treatment, whether a respondent was considered to have a certain disability type depended on the definition of Disability employed. Under Definition 1 of Disability, a respondent had a particular disability type in the Test treatment if they responded “a lot of difficulty” or “cannot do at all” to the relevant question. For Definition 2 of Disability, the respondent was considered to have a particular disability type in the Test treatment if they responded “some difficulty,” “a lot of difficulty” or “cannot do at all” to the relevant question.

Overall, when Definition 1 of Disability was used, Disability estimates were significantly lower in the Test treatment, relative to Control, for all Disability types across all modes, with the exception of Vision. For Vision, Test had a significantly lower estimate than Control when all response modes were combined, but when broken down by mode, the percent with vision disability was only significantly higher in Test, relative to Control, within CAPI mode.

In contrast, when Definition 2 of Disability was used, the proportion of respondents with a disability was significantly higher in Test than Control for every Disability type in every mode. Notably, the absolute difference between Test and Control was nominally larger when using Definition 2 of Disability, relative to Definition 1, for every Disability type in every mode with only one exception (Self-Care in mail mode).

In the remainder of this section, we discuss each individual Disability question after the tables presenting the results of that question.

**Table 34. Distribution of Disability: Hearing**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	1.7 (0.1)	11.7 (0.3)	4.0 (0.1)	-2.2 (0.2)	<0.01*	7.8 (0.3)	<0.01*
Self-Response	1.7 (0.1)	12.0 (0.3)	3.9 (0.1)	-2.1 (0.2)	<0.01*	8.1 (0.3)	<0.01*
Internet	1.5 (0.1)	10.7 (0.3)	3.3 (0.1)	-1.8 (0.2)	<0.01*	7.3 (0.3)	<0.01*
Mail	2.7 (0.4)	18.2 (0.9)	6.5 (0.5)	-3.8 (0.6)	<0.01*	11.7 (1.0)	<0.01*
CAPI	1.8 (0.2)	11.1 (0.7)	4.4 (0.4)	-2.6 (0.4)	<0.01*	6.7 (0.8)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

Table 34 (above) compares response distributions for Hearing within the Test treatment and Control, by definition of Disability and mode.

Differences between Test and Control in the estimated percentage of people with hearing disability may be due to differences in the answer categories and the threshold used for Disability in the two treatments. That being said, the differences may also be due to differences in question text between Test and Control. The Control version used the ACS-6 question text and asked, “Is this person deaf or does he/she have serious difficulty hearing?” In contrast, the Test version, which is based on the WG-SS, did not include the word “deaf” and asked about hearing difficulty when using a hearing aid. Specifically, the Test treatment asked, “Does this person have difficulty hearing, even if using a hearing aid?”

An estimated 4.0 percent (SE = 0.1) of respondents said “yes” when asked the Hearing question in the Control. In the Test treatment, 1.7 percent (SE = 0.1) of respondents said, “a lot of difficulty” or “cannot do at all,” to the Hearing question (Definition 1 of Disability), while 11.7 percent (SE = 0.3) reported “some difficulty”, “a lot of difficulty” or “cannot do at all” in the Test treatment (Definition 2 of Disability). For both definitions of Disability, the gap between Test and Control estimates appeared to be nominally larger in the mail mode compared to the other modes.

**Table 35. Distribution of Disability: Vision**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	1.7 (0.1)	16.3 (0.4)	2.6 (0.1)	-0.9 (0.1)	<0.01*	13.7 (0.4)	<0.01*
Self-Response	1.8 (0.1)	17.2 (0.4)	2.0 (0.1)	-0.2 (0.2)	0.37	15.2 (0.4)	<0.01*
Internet	1.6 (0.1)	15.7 (0.4)	1.8 (0.1)	-0.2 (0.2)	0.39	13.9 (0.4)	<0.01*
Mail	2.7 (0.3)	24.4 (1.2)	3.1 (0.3)	-0.4 (0.4)	0.39	21.4 (1.4)	<0.01*
CAPI	1.6 (0.2)	13.4 (0.8)	4.4 (0.3)	-2.8 (0.4)	<0.01*	9.0 (0.9)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

Like the Hearing question, the version of the Vision question in the Test treatment included changes to the question text as well as to the answer categories.

Specifically, the Control treatment used the ACS-6 question text and asked, “Is this person blind or does he/she have serious difficulty seeing, even if wearing glasses?” In contrast, the Test treatment does not include the word “blind” and instead asked, “Does this person have difficulty seeing, even if wearing glasses?”

An estimated 2.6 percent (SE = 0.1) of respondents said “yes” when asked the Vision question in Control. In the Test treatment, 1.7 percent (SE = 0.1) of respondents said “a lot of difficulty” or “cannot do at all,” corresponding to having a disability under Definition 1, while 16.3 percent (SE = 0.4) said “some difficulty,” “a lot of difficulty,” or “cannot do at all,” corresponding to disability under Definition 2. Compared to other modes, the gap between Test and Control estimates appeared to be nominally larger in the CAPI mode for Definition 1, and nominally larger in the mail mode for Definition 2.

**Table 36. Distribution of Disability: Cognition**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	2.3 (0.1)	15.3 (0.3)	5.2 (0.2)	-2.9 (0.2)	<0.01*	10.1 (0.4)	<0.01*
Self-Response	2.2 (0.1)	15.7 (0.3)	4.6 (0.2)	-2.4 (0.2)	<0.01*	11.1 (0.4)	<0.01*
Internet	2.1 (0.1)	14.5 (0.3)	4.2 (0.2)	-2.2 (0.2)	<0.01*	10.3 (0.4)	<0.01*
Mail	3.1 (0.4)	14.7 (1.1)	6.4 (0.4)	-3.4 (0.6)	<0.01*	14.7 (1.1)	<0.01*
CAPI	2.4 (0.3)	14.2 (0.8)	7.1 (0.5)	-4.7 (0.7)	<0.01*	7.1 (1.1)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

The text of the Cognition question in the Test treatment differed from Control in a number of key ways. The Control version of the question included a preamble qualifying the conditions that might cause difficulty and included reference to “making decisions.” Specifically, the

question asked, “Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?” In contrast, the Test version simply asked, “Does this person have difficulty remembering or concentrating?”

In the Control treatment, an estimated 5.2 percent (SE = 0.2) of respondents answered “yes” to the Cognition question. In the Test treatment, 2.3 percent (SE = 0.1) of people reported “a lot of difficulty” or “cannot do at all,” corresponding to having a disability under Definition 1. When Definition 2 of Disability was used, the percent with cognitive difficulty increased to 15.3 (SE = 0.3). Compared to other modes, the gap between Test and Control estimates appeared to be nominally larger in the CAPI mode for Definition 1, and nominally smaller in the CAPI mode for Definition 2.

**Table 37. Distribution of Disability: Ambulation**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	4.3 (0.2)	16.0 (0.3)	7.4 (0.2)	-3.1 (0.3)	<0.01*	8.6 (0.4)	<0.01*
Self-Response	4.0 (0.2)	16.0 (0.4)	6.6 (0.2)	-2.6 (0.3)	<0.01*	9.4 (0.4)	<0.01*
Internet	3.6 (0.2)	14.0 (0.4)	5.3 (0.2)	-1.7 (0.2)	<0.01*	8.8 (0.4)	<0.01*
Mail	5.9 (0.4)	25.0 (1.0)	12.7 (0.7)	-6.8 (0.8)	<0.01*	12.3 (1.2)	<0.01*
CAPI	5.4 (0.5)	16.3 (0.9)	10.0 (0.6)	-4.6 (0.8)	<0.01*	6.3 (1.1)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

The question text for Ambulation was similar across the two treatments. The only difference was that the Test version of the question excluded the word “serious” before “difficulty” and replaced the word “stairs” with “steps.” Yet, when Definition 1 of Disability was used, the percentage with ambulatory difficulty in Test was significantly lower than the corresponding estimate in Control (4.3 percent (SE = 0.2) vs. 7.4 percent (SE = 0.2)) and when Definition 2 of Disability was used, the percentage with ambulatory difficulty in Test was significantly higher than Control (16.0 percent (SE = 0.3) vs. 7.4 percent (SE = 0.2)). For both definitions of Disability, the gap between Test and Control estimates appeared to be nominally larger in the mail mode compared to the other modes.



**Table 38. Distribution of Disability: Self-Care**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	1.4 (0.1)	5.7 (0.2)	2.5 (0.1)	-1.1 (0.1)	<0.01*	3.1 (0.2)	<0.01*
Self-Response	1.2 (0.1)	4.8 (0.2)	2.2 (0.1)	-1.0 (0.1)	<0.01*	2.6 (0.2)	<0.01*
Internet	1.1 (0.1)	4.3 (0.2)	1.7 (0.1)	-0.6 (0.1)	<0.01*	2.6 (0.2)	<0.01*
Mail	1.7 (0.3)	7.3 (0.5)	4.6 (0.3)	-2.9 (0.4)	<0.01*	2.6 (0.6)	<0.01*
CAPI	1.9 (0.3)	8.2 (0.6)	3.5 (0.3)	-1.6 (0.5)	<0.01*	4.7 (0.7)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

The text of the Self-Care question also varied between the Test and Control treatments. In the Control treatment, the question asked, “Does this person have difficulty dressing or bathing?” In the Test treatment, the term “bathing” was replaced with “washing all over” and respondents were also prompted to consider other forms of “self-care.” Specifically, the Test question asks, “Does this person have difficulty with self-care, such as washing all over or dressing?”

In the Control treatment, 2.5 percent (SE = 0.1) of respondents said “yes” to the Self-Care question. In the Test treatment, 1.4 percent (SE = 0.1) of people responded “a lot of difficulty” or “cannot do at all” to the question, which corresponds to Definition 1 of Disability, while 5.7 percent (SE = 0.2) responded, “some difficulty,” “a lot of difficulty,” or “cannot do at all,” corresponding to Definition 2 of Disability. Compared to other modes, the gap between Test and Control estimates appeared to be nominally larger in the mail mode for Definition 1, and nominally larger in the CAPI mode for Definition 2.

**Table 39. Distribution of Disability: Independent Living**

Mode	Test Def. 1 Percent	Test Def. 2 Percent	Control Percent	Definition 1 Difference	Adj. P- value	Definition 2 Difference	Adj. P- value
<b>Overall</b>	3.4 (0.1)	9.6 (0.2)	5.9 (0.2)	-2.4 (0.2)	<0.01*	3.7 (0.3)	<0.01*
Self-Response	3.1 (0.1)	9.2 (0.3)	5.5 (0.2)	-2.4 (0.2)	<0.01*	3.6 (0.3)	<0.01*
Internet	2.8 (0.2)	8.3 (0.3)	4.8 (0.2)	-2.0 (0.2)	<0.01*	3.5 (0.3)	<0.01*
Mail	4.6 (0.5)	12.7 (0.7)	8.5 (0.5)	-3.9 (0.7)	<0.01*	4.3 (0.8)	<0.01*
CAPI	4.4 (0.4)	11.1 (0.7)	7.1 (0.5)	-2.6 (0.6)	<0.01*	4.1 (0.8)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

In both Test and Control, the question about independent living difficulty asked, “Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor’s office or shopping?” The text did not differ between the Test

treatment and Control because this item is not part of the WG-SS. As such, the only difference between Test and Control for this Disability question was in the answer categories (two response options in Control; four response options in Test).

The results show that in the Control treatment, 5.9 percent (SE = 0.2) of people answered “yes” to the Independent Living question. In the Test treatment, 3.4 percent (SE = 0.1) of people reported “a lot of difficulty” or “cannot do at all” (Definition 1 of Disability) and 9.6 percent (SE = 0.2) of people reported “some difficulty,” “a lot of difficulty,” or “cannot do at all” (Definition 2 of Disability). In other words, regardless of the definition of Disability employed, the percentage of people with independent living difficulty in the Test significantly differed from Control. The gap between Test and Control estimates appeared to be of nominally similar magnitude across modes, regardless of the Disability definition used.

**Table 40. Distribution of Disability: Communication (Test Version Only)**

Mode	Test Def. 1 Percent	Test Def. 2 Percent
<b>Overall</b>	0.9 (0.1)	5.5 (0.2)
Self-Response	0.8 (0.1)	4.8 (0.2)
Internet	0.8 (0.1)	4.5 (0.2)
Mail	1.2 (0.2)	6.1 (0.5)
CAPI	1.1 (0.2)	7.7 (0.6)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

The Test version also included a question about communication difficulty. Table 40 presents response distributions for this question for each definition of Disability. Since the Communication question did not appear in the Control treatment, it is not possible to compare the Test and Control distributions for this Disability type.

### 5.3.3 Response Distributions of Test Treatment Response Categories

*RQ 4. In the Test treatment, what is the distribution of responses for each Disability question (overall and by mode)?*

For informational purposes, we also computed the distribution of responses among the four response categories in the Test treatment. Table 41 through Table 47 show the distributions for each Disability type, both overall and by mode.

Response distributions looked similar across the Disability questions. Given the low prevalence of Disability in the Test treatment, it is not surprising that the most common response categories across all Disability types and modes were “No Difficulty” and “Some Difficulty.”

**Table 41. Distribution of Test Treatment Response Categories: Hearing**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
<b>Overall</b>	88.3 (0.3)	10.0 (0.2)	1.7 (0.1)	0.1 (<0.1)
Self-Response	88.0 (0.3)	10.2 (0.2)	1.7 (0.1)	0.1 (<0.1)
Internet	89.4 (0.3)	9.1 (0.2)	1.4 (0.1)	0.1 (<0.1)
Mail	81.8 (0.9)	15.5 (0.7)	2.7 (0.4)	<0.1 (<0.1)
CAPI	88.9 (0.7)	9.3 (0.6)	1.7 (0.3)	0.2 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 42. Distribution of Test Treatment Response Categories: Vision**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
<b>Overall</b>	83.7 (0.4)	14.5 (0.3)	1.6 (0.1)	0.2 (<0.1)
Self-Response	82.8 (0.4)	15.4 (0.4)	1.6 (0.1)	0.2 (<0.1)
Internet	84.3 (0.4)	14.1 (0.4)	1.4 (0.1)	0.2 (<0.1)
Mail	75.6 (1.2)	21.7 (1.2)	2.6 (0.3)	0.1 (<0.1)
CAPI	86.6 (0.8)	11.8 (0.8)	1.4 (0.2)	0.2 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 43. Distribution of Test Treatment Response Categories: Cognition**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
<b>Overall</b>	84.7 (0.3)	13.0 (0.3)	2.1 (0.1)	0.2 (<0.1)
Self-Response	84.3 (0.4)	13.4 (0.3)	2.1 (0.1)	0.1 (<0.1)
Internet	85.5 (0.3)	12.5 (0.3)	1.9 (0.1)	0.1 (<0.1)
Mail	78.8 (1.1)	18.1 (0.9)	3.0 (0.4)	0.1 (<0.1)
CAPI	85.8 (0.8)	11.8 (0.7)	2.1 (0.3)	0.3 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 44. Distribution of Test Treatment Response Categories: Ambulation**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
<b>Overall</b>	84.0 (0.3)	11.7 (0.3)	3.7 (0.1)	0.6 (0.1)
Self-Response	84.1 (0.4)	12.0 (0.3)	3.6 (0.2)	0.4 (<0.1)
Internet	85.9 (0.4)	10.4 (0.3)	3.2 (0.2)	0.3 (<0.1)
Mail	75.0 (1.0)	19.1 (0.9)	5.1 (0.4)	0.8 (0.2)
CAPI	83.7 (0.9)	10.9 (0.7)	4.2 (0.4)	1.2 (0.2)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 45. Distribution of Test Treatment Response Categories: Self-Care**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
Overall	94.3 (0.2)	4.3 (0.1)	1.0 (0.1)	0.4 (<0.1)
Self-Response	95.1 (0.2)	3.6 (0.1)	1.0 (0.1)	0.3 (<0.1)
Internet	95.7 (0.2)	3.2 (0.2)	0.9 (0.1)	0.2 (<0.1)
Mail	92.7 (0.5)	5.5 (0.4)	1.4 (0.2)	0.3 (0.1)
CAPI	91.8 (0.6)	6.3 (0.5)	1.2 (0.2)	0.6 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 46. Distribution of Test Treatment Response Categories: Independent Living**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
Overall	90.4 (0.2)	6.2 (0.2)	1.8 (0.1)	1.6 (0.1)
Self-Response	90.9 (0.3)	6.0 (0.2)	1.8 (0.1)	1.3 (0.1)
Internet	91.6 (0.3)	5.5 (0.2)	1.7 (0.1)	1.1 (0.1)
Mail	87.3 (0.7)	8.1 (0.5)	2.4 (0.3)	2.2 (0.4)
CAPI	88.9 (0.7)	6.7 (0.5)	1.9 (0.3)	2.6 (0.3)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 47. Distribution of Test Treatment Response Categories: Communication**

Mode	No Difficulty	Some Difficulty	A Lot of Difficulty	Cannot Do at All
Overall	94.5 (0.2)	4.6 (0.2)	0.7 (0.1)	0.2 (<0.1)
Self-Response	95.2 (0.2)	3.9 (0.2)	0.7 (0.1)	0.1 (<0.1)
Internet	95.5 (0.2)	3.7 (0.2)	0.7 (0.1)	0.1 (<0.1)
Mail	93.9 (0.5)	4.9 (0.4)	1.0 (0.2)	0.3 (0.1)
CAPI	92.3 (0.6)	6.6 (0.6)	0.7 (0.1)	0.4 (0.1)

Source: U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

### 5.3.4 Summary of Response Distribution Results

In summary, as expected, using Definition 1 of Disability (the standard/recommended definition), the overall Disability rate in the Test treatment (8.1 percent (SE = 0.2)) was 5.8 percentage points lower (SE = 0.4) than the corresponding estimate in Control (13.9 percent (SE = 0.3)). This is a decline of about 40 percent in estimated Disability prevalence when moving from the Control measure to the Test measure of Disability, which is comparable to the 50 percent decline observed in prior NCHS comparisons of the WG-SS to the ACS-6 (see Section 5.3.1). A significantly lower rate of Disability in the Test treatment, relative to Control, was observed across all Disability types and modes, with the exceptions of Vision in self-response modes.

Definition 1 of Disability identifies a population with higher levels of functional difficulty than Definition 2, in that some respondents who report difficulty with activities in the Test treatment (i.e., “some difficulty”) are classified as not having a disability under Definition 1. As such, we also explored the option of using a broader definition of Disability (Definition 2). When Definition 2 of Disability is employed, individuals who report *any* difficulty (“some difficulty,” “a lot of difficulty” or “cannot do at all”) are classified as having a disability.

Using Definition 2 of Disability, the overall Disability rate in Test (31.7 percent (SE = 0.4)) was 17.8 percentage points *higher* (SE = 0.5) than the corresponding estimate in Control (13.9 percent (SE = 0.3)). In other words, the Disability rate in the Test treatment was more than two times as large as the Disability rate in Control when using Definition 2. This difference is in the opposite direction of the difference observed between Test and Control when using Definition 1 of Disability and appears to be of a nominally larger magnitude. When Definition 2 of Disability was used, the Test treatment had higher Disability rates, relative to Control, across all Disability types and modes. Prior NCHS research based on 2011-2012 NHIS data found that the majority of respondents who reported “some difficulty” to the WG-SS had reported “no” to the ACS-6 question set earlier in the survey (Weeks et al., 2021). Thus, the finding that the Test treatment yields disability estimates that are significantly higher than Control when Definition 2 of Disability is used is consistent with earlier evaluations.

It should be noted that the magnitude of the difference between Test and Control estimates of Disability may vary by Disability type. Overall, the percentage of people with any disability in Test was about 40 percent lower than the corresponding estimate in Control when Definition 1 of Disability was used (8.1 percent (SE = 0.2) vs. 13.9 percent (SE = 0.3)). When broken down by type, however, some gaps were nominally larger than others. For example, the percentage of people with seeing difficulty was about 35 percent lower in the Test treatment, relative to Control (1.7 percent (SE = 0.1) with seeing difficulty in Test vs. 2.6 percent (SE = 0.1) in Control). Meanwhile, estimates of the prevalence of hearing and cognitive difficulty were roughly 55 percent lower in Test than Control (1.7 percent (SE = 0.1) vs. 4.0 percent (SE = 0.1) with hearing difficulty; 2.3 percent (SE = 0.1) vs. 5.2 percent (SE = 0.2) with cognitive difficulty). The magnitude of the difference between Test and Control estimates of Disability may also vary by Disability type when Definition 2 of Disability is used. Under Definition 2, the percentage of people with any disability in Test was about 130 percent higher in Test than the corresponding estimate in Control (31.7 percent (SE = 0.4) vs. 13.9 percent (SE = 0.3)). However, the percentage of people with independent living difficulty was only about 60 percent higher in Test, relative to Control (9.6 percent (SE = 0.2) with independent living difficulty in Test vs. 5.9 percent (SE = 0.2) in Control), while the percentage with seeing difficulty in Test was about 530 percent higher in Test than Control (16.3 percent (SE = 0.4) with seeing difficulty in Test vs. 2.6 percent (SE = 0.1) in Control).

Again, it is important to understand that estimates of disability prevalence are not part of the decision criteria for the reasons stated above (see Section 2.4.2.1). Analyses of response

distributions by treatment are provided for informational purposes only, as it is incumbent upon the Census Bureau to prepare data users for this question change by providing as much information as available on implications of this change. Overall, adopting the Test version of Disability is expected to result in a significant change in ACS estimates of the number and percentage of people in the U.S. with a disability, relative to prior years. If Definition 1 of Disability is employed, we can expect a significant decrease in estimated disability prevalence in the U.S. If Definition 2 of Disability is employed, in contrast, we can expect a significant *increase* in estimated disability prevalence.

#### 5.4 Response Reliability Results for Disability

Response reliability represents one of the key criteria for evaluating the Test version of Disability, relative to Control. The main reliability measures, Gross Difference Rate (GDR) and Index of Inconsistency (IOI), are based upon responses to the Content Follow-Up interview (CFU). See Section 2.3 for more information about the CFU.

A full analysis of CFU response rates is available in Spiers et al. (2023). Table 48 presents the CFU unit response rates for the Control and Test treatments, overall and by mode of original interview. Although not all respondents completed a CFU interview, there was no statistically significant difference between Test and Control in CFU response rates. The CFU response rate was 32.9 percent (SE = 0.5) for the Control treatment and 34.2 percent (SE = 0.5) for the Test treatment.<sup>27</sup> That being said, we did note that the Disability rate was higher, on average, among respondents who completed a CFU interview, relative to those who did not (Table 49). This difference in Disability rates between CFU respondents and CFU nonrespondents was observed in both the Control and Test treatment and regardless of the definition of Disability used (see Table 3 for a discussion of Definitions 1 and 2 of Disability).

**Table 48. CFU Reinterview Unit Response Rates for Control and Test Treatments, Overall and by Mode**

Original Interview Mode	Test CFU Interviews	Test Percent	Control CFU Interviews	Control Percent	Difference	Adj. P-value
<b>Overall</b>	5,600	34.2 (0.5)	5,500	32.9 (0.5)	1.3 (0.7)	0.18
Self-Response	4,900	35.0 (0.5)	4,800	33.6 (0.5)	1.4 (0.7)	0.18
Internet	3,700	33.7 (0.5)	3,500	32.2 (0.6)	1.5 (0.8)	0.18
Mail	1,400	40.1 (1.2)	1,300	38.2 (1.1)	1.8 (1.5)	0.44
CAPI	700	31.0 (1.3)	700	30.2 (1.2)	0.8 (1.8)	0.66

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0058

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

<sup>27</sup> The CFU response rate is the proportion of households that were eligible for CFU who completed a CFU reinterview (see Section 2.3 for details about CFU eligibility).

**Table 49. Distribution of Overall Disability: CFU Respondents vs. Nonrespondents**

Treatment	CFU Respondents	CFU Non-respondents	Difference	Adj. P-value
Control	16.3 (0.5)	12.9 (0.3)	3.5 (0.6)	<0.01*
Test (Definition 1)	9.1 (0.4)	7.6 (0.3)	1.6 (0.5)	<0.01*
Test (Definition 2)	34.9 (0.8)	30.3 (0.5)	4.6 (1.0)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

This section uses data from respondents with a CFU interview to estimate response reliability for overall Disability, as well as by Disability type, in the Control and Test treatments.

#### 5.4.1 Response Reliability for Overall Disability

*RQ 5. How do the measures of response reliability differ between the Test treatment and the Control treatment for overall Disability (overall, by mode of original interview, and by same vs. different CFU respondent)?*

When evaluating response reliability, higher reliability (and conversely, lower response variance) is generally preferred. In the case of Disability and this study, however, lower response reliability in the Test treatment *may* be acceptable when considering the previously documented validity of the WG-SS. Prior NCHS evaluations (Madans et al., 2011; Miller et al., 2020), including cognitive testing, indicated that the Test treatment is a valid measure of functional disability with potential advantages over the ACS-6.

Although the WG-SS has been used in various surveys that are administered using a variety of modes, most of the prior testing of the WG-SS primarily relied on interviewer administration of the survey, rather than self-response. The ACS relies heavily on mail and internet response. As such, it was particularly important to evaluate response reliability of the Test treatment within self-response modes (mail; internet), relative to interviewer-administered modes (CAPI).

The Control treatment has only two response categories while the Test treatment has four response categories, and thus the response categories for the Test treatment were collapsed into yes/no using either the first or second definition of Disability described in Table 3. Then GDR and IOI were calculated as discordance in yes/no between the original interview and reinterview.

##### 5.4.1.1 Response Reliability for Overall Disability Using Definition 1 of Disability

Table 50 and Table 51 present the GDR and IOI results for overall Disability status using Definition 1 of Disability, for all persons that answered at least one Disability question in both the original interview and the CFU interview. Higher GDR and IOI scores indicate lower response reliability.

As described in Section 2.4.2.4, before calculating IOI, we calculated the NDR to check the assumption of parallel measures. For each table of IOI results, if at least one NDR for that table was found to be significantly different from zero, we used the adjusted IOI for that entire table. In the results below, the header for each table indicates whether it uses adjusted IOI.

As shown in Table 50, when using Definition 1 of Disability, the GDR for overall Disability was significantly higher for Control than for the Test treatment in every mode except CAPI. In other words, based on GDR, the Test treatment had better response reliability than Control.

On the other hand, IOI results sometimes indicated lower response reliability in the Test treatment, relative to the Control. For overall Disability using Definition 1, IOI was significantly higher for Test than Control in all modes combined and CAPI.

Here it should be noted that GDR is a simple ratio of the number of responses that switched between having a disability and not having a disability (in any direction) over the total number of responses. By contrast, IOI is a more complex calculation that accounts for the size of the relevant category, in this case, the proportion of people with a disability. The Test version of the Disability series, when using Definition 1, counts fewer people as having a disability than the Control version (see Section 5.3); such differences may influence whether responses remain consistent at follow-up, and thus are accounted for within the IOI. GDR does not account for differences between Test and Control in the proportion with a disability, which may help explain why when using Definition 1 of Disability, GDR indicated that overall Disability had better reliability in the Test treatment, relative to Control, while IOI results were in the opposite direction, indicating better response reliability in the Control treatment.

In terms of scale, a rule of thumb used by the Census Bureau states that IOI values below 20 indicate low response inconsistency, 20 to 50 indicate moderate inconsistency, and over 50 indicate high inconsistency (Singer & Ennis, 2003). When using Definition 1 of Disability in Test, across most modes the Test and Control treatments had IOI scores for overall Disability that were in the moderate inconsistency range. The one exception was the CAPI mode, in which the Test treatment had a high level of inconsistency (54.9 (SE = 4.7)).<sup>28</sup>

---

<sup>28</sup> As discussed in Section 4.2, a limitation of the field test was that CFU reinterviews were only conducted by phone, while the original interviews were conducted by internet, mail, or personal interview. The changes in mode between the two interviews may have had an effect on the measured inconsistency by mode.



**Table 50. GDR for Definition 1 of Disability: Overall Disability**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	6.6 (0.4)	9.9 (0.5)	-3.3 (0.5)	<0.01*
Self-Response	5.6 (0.3)	9.3 (0.5)	-3.7 (0.5)	<0.01*
Internet	4.8 (0.3)	7.9 (0.5)	-3.1 (0.6)	<0.01*
Mail	8.8 (0.9)	14.8 (1.2)	-6.0 (1.6)	<0.01*
CAPI	10.4 (1.1)	12.1 (1.2)	-1.8 (1.6)	0.25

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 51. Adjusted IOI for Definition 1 of Disability: Overall Disability**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	43.1 (1.8)	35.8 (1.5)	7.4 (2.3)	0.01*
Self-Response	38.9 (1.9)	36.3 (1.7)	2.6 (2.6)	0.60
Internet	38.3 (2.5)	36.6 (2.0)	1.6 (3.1)	0.60
Mail	41.1 (3.3)	37.6 (3.1)	3.4 (5.0)	0.60
CAPI	54.9 (4.7)	35.1 (3.2)	19.7 (5.9)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

Although not part of the decision criteria, we also calculated GDR and IOI separately by whether or not the CFU responses were reported by the same individual who completed the original interview. We expected that there would be more inconsistency between reporting in initial versus follow-up interview when different respondents report. No meaningful insights were gained from this analysis. Since this comparison was not part of the decision criteria, we do not include it in this section of the report. Tables comparing GDR and IOI between households with the same versus different CFU respondents can be found in Appendix A.<sup>29</sup>

In summary, when using Definition 1 of Disability, response reliability results for overall Disability were mixed. The Test treatment performed better than Control according to one metric (GDR), while it sometimes performed worse than the Control according to another metric (IOI). It should be noted, however, that higher IOI for the Test treatment was only observed in one mode: CAPI. No significant differences in IOI were detected between Test and Control treatments in either of the two self-response modes (i.e., mail; internet). Finally, the IOI scores for both treatments were generally in the moderate inconsistency range, with the exception of the Test treatment in CAPI mode, which was in the high inconsistency range.

<sup>29</sup> The universe of households with a different CFU respondent was relatively small (8.4 percent (SE = 0.3) of all CFU cases).

#### 5.4.1.2 Response Reliability for Overall Disability Using Definition 2 of Disability

Definition 1 of Disability is consistent with the Washington Group’s recommendation for how to define disability status. Analyses were also conducted using a second definition of Disability that considers a broader range of functional difficulty. Table 52 and Table 53 give the response reliability results for overall disability status using Definition 2 of Disability.<sup>30</sup>

When using Definition 2, GDR for overall Disability was significantly higher in Test than in Control in every mode. IOI for overall Disability was also significantly higher in Test than in Control, in every mode except mail. In other words, when Definition 2 of Disability was used, overall Disability in the Test treatment generally had lower response reliability than it did in the Control, regardless of the metric used (GDR or IOI). However, in terms of the magnitude of response inconsistency in Test and Control, when Definition 2 of Disability was used, IOI scores for overall Disability were in the moderate inconsistency range for both treatments. This was true when all modes were combined and within each mode.

**Table 52. GDR for Definition 2 of Disability: Overall Disability**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	19.3 (0.6)	9.9 (0.5)	9.4 (0.7)	<0.01*
Self-Response	18.8 (0.6)	9.3 (0.5)	9.6 (0.8)	<0.01*
Internet	18.2 (0.6)	7.9 (0.5)	10.3 (0.8)	<0.01*
Mail	21.2 (1.4)	14.8 (1.2)	6.4 (1.7)	<0.01*
CAPI	20.9 (1.4)	12.1 (1.2)	8.7 (1.6)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 53. Adjusted IOI for Definition 2 of Disability: Overall Disability**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	43.5 (1.2)	35.8 (1.5)	7.7 (1.8)	<0.01*
Self-Response	41.5 (1.1)	36.3 (1.7)	5.2 (1.9)	0.02*
Internet	42.2 (1.2)	36.6 (2.0)	5.6 (2.4)	0.04*
Mail	41.5 (2.4)	37.6 (3.1)	3.9 (3.6)	0.28
CAPI	48.2 (2.8)	35.1 (3.2)	13.1 (3.7)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

<sup>30</sup> For Definition 2, we did not perform additional analyses for same versus different CFU respondents.

#### **5.4.2 Response Reliability by Disability Type**

In addition to examining response reliability for overall Disability, we compared response reliability between Test and Control treatments by Disability type.

*RQ 6. How do the measures of response reliability differ between the Test treatment and the Control treatment by Disability type (overall, by mode of original interview, and by same vs. different CFU respondent)?*

##### **5.4.2.1 Response Reliability by Disability Type Using Definition 1 of Disability**

Table 54 through Table 67 present the GDR and IOI for each Disability type using Definition 1 of Disability. When Definition 1 was used, GDR was significantly lower for the Test treatment than Control for every Disability type and in every mode, except for Vision in the internet mode, where it was not significantly different between treatments. As such, the GDR results suggested higher response reliability in the Test treatment.

The IOI results by Disability type using Definition 1, on the other hand, were mixed. For some Disability types (Cognition, Self-Care, Independent Living), no significant difference was detected between the Test treatment and Control. For others (Vision, Ambulation), IOI was significantly higher in Test than Control, but only in one mode (CAPI for Vision difficulty) or only if all modes were combined. The one exception was Hearing difficulty, which had significantly higher IOI in the Test treatment across all individual modes and for all modes combined.

When Definition 1 of Disability was used, the magnitude of response inconsistency in the Test and Control treatments was in the moderate range for Cognition, Ambulation, and Self-Care. This was true across all modes except CAPI, which sometimes had IOI levels in the high range for these three Disability types, particularly in the Test treatment. In contrast to Cognition, Ambulation, and Self-Care, response inconsistency levels were often high for Hearing and Vision, though in most cases only in the Test treatment. In the Control treatment, Hearing and Vision generally had IOI scores in the moderate inconsistency range. Nominally, Hearing and Vision for the Test treatment had among the highest IOI values among the Disability types, both overall and for most modes.

Response reliability for the Communication difficulty question could not be compared between treatments since this question was only asked in Test treatment. We can note, however, that all IOI scores were in either in the moderate inconsistency range or relatively close to it with the one exception of CAPI, which had a high IOI score.

**Table 54. GDR for Definition 1 of Disability: Hearing**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.0 (0.2)	4.5 (0.3)	-2.5 (0.3)	<0.01*
Self-Response	1.9 (0.2)	4.1 (0.3)	-2.2 (0.3)	<0.01*
Internet	1.7 (0.2)	3.3 (0.3)	-1.6 (0.4)	<0.01*
Mail	2.6 (0.5)	7.6 (0.9)	-5.1 (1.0)	<0.01*
CAPI	2.5 (0.5)	5.7 (0.7)	-3.2 (1.0)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 55. Adjusted IOI for Definition 1 of Disability: Hearing**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	65.0 (3.4)	45.7 (2.5)	19.3 (4.0)	<0.01*
Self-Response	63.2 (4.2)	45.3 (2.7)	18.0 (5.4)	<0.01*
Internet	60.8 (4.8)	47.1 (3.9)	13.7 (6.8)	0.07*
Mail	70.6 (6.7)	43.3 (4.4)	27.3 (8.4)	<0.01*
CAPI	70.6 (10.8)	46.8 (6.4)	23.7 (13.2)	0.07*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 56. GDR for Definition 1 of Disability: Vision**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.9 (0.2)	2.8 (0.2)	-1.0 (0.3)	<0.01*
Self-Response	1.7 (0.2)	2.4 (0.2)	-0.7 (0.3)	0.06*
Internet	1.7 (0.2)	2.0 (0.3)	-0.3 (0.3)	0.34
Mail	1.7 (0.3)	4.0 (0.6)	-2.3 (0.6)	<0.01*
CAPI	2.4 (0.6)	4.2 (0.6)	-1.8 (0.9)	0.07*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 57. Adjusted IOI for Definition 1 of Disability: Vision**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	64.4 (5.2)	45.4 (3.1)	19.0 (6.5)	0.01*
Self-Response	60.3 (5.6)	48.3 (3.8)	12.0 (7.2)	0.26
Internet	64.6 (6.3)	52.1 (5.3)	12.5 (8.3)	0.26
Mail	47.3 (10.4)	41.9 (5.7)	5.5 (12.1)	0.65
CAPI	77.8 (9.4)	41.0 (5.8)	36.8 (11.5)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 58. GDR for Definition 1 of Disability: Cognition**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.2 (0.2)	5.0 (0.3)	-2.8 (0.4)	<0.01*
Self-Response	2.0 (0.2)	4.3 (0.3)	-2.3 (0.4)	<0.01*
Internet	1.9 (0.2)	3.9 (0.4)	-2.0 (0.4)	<0.01*
Mail	2.4 (0.5)	6.0 (0.7)	-3.7 (0.8)	<0.01*
CAPI	2.8 (0.6)	7.1 (0.9)	-4.4 (1.1)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 59. Adjusted IOI for Definition 1 of Disability: Cognition**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	49.9 (4.0)	44.8 (2.6)	5.2 (4.8)	0.76
Self-Response	48.6 (4.6)	41.5 (2.9)	7.2 (5.4)	0.76
Internet	49.1 (5.3)	43.1 (3.8)	6.0 (6.7)	0.76
Mail	47.3 (8.4)	38.0 (4.2)	9.2 (8.7)	0.76
CAPI	53.6 (8.5)	53.8 (4.8)	-0.2 (9.7)	0.98

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 60. GDR for Definition 1 of Disability: Ambulation**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	3.7 (0.2)	5.9 (0.3)	-2.1 (0.4)	<0.01*
Self-Response	3.1 (0.2)	5.0 (0.3)	-1.9 (0.4)	<0.01*
Internet	2.5 (0.2)	3.7 (0.3)	-1.2 (0.4)	<0.01*
Mail	5.7 (0.6)	10.1 (0.9)	-4.4 (1.1)	<0.01*
CAPI	5.9 (0.8)	8.9 (1.0)	-3.1 (1.2)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 61. Adjusted IOI for Definition 1 of Disability: Ambulation**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	42.6 (2.4)	34.8 (1.3)	7.9 (2.6)	0.01*
Self-Response	39.4 (2.6)	33.6 (1.7)	5.7 (3.2)	0.22
Internet	38.3 (3.5)	33.7 (2.5)	4.5 (4.3)	0.29
Mail	41.7 (4.8)	35.8 (2.9)	6.0 (5.4)	0.29
CAPI	50.4 (5.1)	37.8 (3.5)	12.6 (6.2)	0.16

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 62. GDR for Definition 1 of Disability: Self-Care**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.1 (0.1)	2.4 (0.2)	-1.3 (0.3)	<0.01*
Self-Response	0.9 (0.1)	2.0 (0.2)	-1.1 (0.2)	<0.01*
Internet	0.8 (0.1)	1.5 (0.2)	-0.7 (0.2)	<0.01*
Mail	1.4 (0.4)	4.0 (0.7)	-2.5 (0.8)	<0.01*
CAPI	1.8 (0.5)	3.9 (0.6)	-2.1 (0.8)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 63. Adjusted IOI for Definition 1 of Disability: Self-Care**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	42.6 (5.1)	41.3 (3.3)	1.3 (5.7)	0.98
Self-Response	36.8 (5.4)	37.0 (3.5)	-0.2 (6.2)	0.98
Internet	36.8 (6.3)	41.4 (5.1)	-4.6 (8.0)	0.98
Mail	37.0 (10.3)	32.8 (5.0)	4.3 (10.9)	0.98
CAPI	58.7 (12.9)	51.8 (7.4)	6.8 (14.5)	0.98

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 64. GDR for Definition 1 of Disability: Independent Living**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.6 (0.2)	5.1 (0.3)	-2.5 (0.4)	<0.01*
Self-Response	2.1 (0.2)	4.5 (0.4)	-2.3 (0.4)	<0.01*
Internet	1.8 (0.2)	3.5 (0.4)	-1.7 (0.4)	<0.01*
Mail	3.3 (0.6)	8.1 (0.8)	-4.8 (1.0)	<0.01*
CAPI	4.2 (0.7)	7.3 (0.9)	-3.1 (1.2)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 65. Adjusted IOI for Definition 1 of Disability: Independent Living**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	34.8 (2.7)	40.7 (2.2)	-5.9 (3.4)	0.26
Self-Response	31.8 (2.7)	39.1 (2.5)	-7.3 (3.6)	0.17
Internet	32.8 (3.2)	37.8 (3.3)	-5.0 (4.3)	0.48
Mail	30.2 (4.9)	42.7 (3.9)	-12.4 (6.1)	0.17
CAPI	42.4 (6.8)	44.6 (4.4)	-2.2 (8.5)	0.79

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 66. GDR for Definition 1 of Disability: Communication (Test Version Only)**

Mode	Test GDR
<b>Overall</b>	0.9 (0.1)
Self-Response	0.8 (0.1)
Internet	0.7 (0.1)
Mail	1.2 (0.3)
CAPI	1.4 (0.4)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 67. Adjusted IOI for Definition 1 of Disability: Communication (Test Version Only)**

Mode	Test IOI
<b>Overall</b>	54.0 (6.2)
Self-Response	48.9 (5.8)
Internet	53.1 (6.2)
Mail	41.2 (10.1)
CAPI	69.3 (15.2)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

We also compared response reliability for households that had the same respondent in the original interview and the reinterview, versus households that had a different respondent in the CFU. No meaningful insights were gained from this analysis. Since this comparison was not part of the decision criteria, we do not include it in the main text of this report, although results can be found in Appendix A.

In summary, the response reliability results by Disability type were mixed when using Definition 1 of Disability. The Test treatment performed better than the Control when GDR was used as a metric. When IOI was used, the Test treatment performed similarly to the Control for certain Disability types, while it performed worse than Control for others (including Hearing in particular).

#### **5.4.2.2 Response Reliability by Disability Type Using Definition 2 of Disability**

Table 68 through Table 81 present the response reliability results by Disability type, using Definition 2 of Disability. When Definition 2 was used, GDR was significantly higher (indicating lower response reliability) in the Test treatment, relative to Control, for every Disability type in every mode, except for Self-Care and Independent Living in the mail mode.

The IOI results by Disability type, on the other hand, were more mixed. When Definition 2 of Disability was used, IOI scores for Vision were significantly higher in Test, relative to Control, when all modes were combined and across all modes except internet. IOI for Independent Living difficulty was also significantly higher in Test compared to Control when all modes were combined (though not in any individual modes). There were no significant differences in IOI



between Test and Control for any of the other Disability types (Hearing, Cognition, Ambulation, and Self-Care).

Regarding the magnitude of response inconsistency, when Definition 2 of Disability was used, most Disability types had IOI scores in the moderate inconsistency range. This was true regardless of treatment. Exceptions to this general pattern were IOI scores within the CAPI mode, which often fell in the high inconsistency range, and IOI scores for Vision, which were in the high inconsistency range for all modes in Test and for the internet mode in Control.

For Communication, comparisons between treatments were not possible because the Control treatment did not include this question. We can note that under Definition 2, the IOI scores for Communication were all in the high inconsistency range in Test, with a particularly high IOI score for the CAPI mode.

**Table 68. GDR for Definition 2 of Disability: Hearing**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
Overall	9.5 (0.4)	4.5 (0.3)	5.0 (0.5)	<0.01*
Self-Response	8.8 (0.4)	4.1 (0.3)	4.7 (0.5)	<0.01*
Internet	7.9 (0.5)	3.3 (0.3)	4.7 (0.6)	<0.01*
Mail	12.4 (0.9)	7.6 (0.9)	4.7 (1.3)	<0.01*
CAPI	11.9 (1.1)	5.7 (0.7)	6.2 (1.2)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 69. Adjusted IOI for Definition 2 of Disability: Hearing**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
Overall	48.7 (1.9)	45.7 (2.5)	3.0 (2.7)	0.93
Self-Response	44.9 (1.8)	45.3 (2.7)	-0.3 (3.2)	0.93
Internet	43.9 (2.3)	47.1 (3.9)	-3.2 (4.5)	0.93
Mail	48.3 (3.1)	43.3 (4.4)	5.0 (5.2)	0.93
CAPI	62.3 (5.1)	46.8 (6.4)	15.5 (7.3)	0.17

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 70. GDR for Definition 2 of Disability: Vision**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	14.6 (0.5)	2.8 (0.2)	11.7 (0.6)	<0.01*
Self-Response	14.3 (0.5)	2.4 (0.2)	11.8 (0.6)	<0.01*
Internet	13.5 (0.5)	2.0 (0.3)	11.4 (0.6)	<0.01*
Mail	17.4 (1.4)	4.0 (0.6)	13.5 (1.5)	<0.01*
CAPI	15.7 (1.3)	4.2 (0.6)	11.5 (1.5)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 71. Adjusted IOI for Definition 2 of Disability: Vision**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	59.7 (1.7)	45.4 (3.1)	14.3 (3.5)	<0.01*
Self-Response	57.3 (1.7)	48.3 (3.8)	9.0 (4.3)	0.07*
Internet	58.0 (1.9)	52.1 (5.3)	5.9 (5.8)	0.31
Mail	55.9 (3.0)	41.9 (5.7)	14.1 (6.7)	0.07*
CAPI	66.1 (4.5)	41.0 (5.8)	25.1 (7.1)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 72. GDR for Definition 2 of Disability: Cognition**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	12.9 (0.6)	5.0 (0.3)	7.9 (0.7)	<0.01*
Self-Response	12.0 (0.6)	4.3 (0.3)	7.6 (0.7)	<0.01*
Internet	11.2 (0.7)	3.9 (0.4)	7.2 (0.8)	<0.01*
Mail	15.1 (1.4)	6.0 (0.7)	9.1 (1.6)	<0.01*
CAPI	16.3 (1.4)	7.1 (0.9)	9.2 (1.6)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 73. Adjusted IOI for Definition 2 of Disability: Cognition**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	49.4 (1.9)	44.8 (2.6)	4.7 (3.0)	0.47
Self-Response	46.1 (1.9)	41.5 (2.9)	4.6 (3.2)	0.47
Internet	45.4 (2.4)	43.1 (3.8)	2.3 (4.4)	0.61
Mail	48.6 (3.7)	38.0 (4.2)	10.6 (5.7)	0.33
CAPi	60.2 (4.1)	53.8 (4.8)	6.4 (6.2)	0.61

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 74. GDR for Definition 2 of Disability: Ambulation**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	10.1 (0.5)	5.9 (0.3)	4.2 (0.6)	<0.01*
Self-Response	8.9 (0.4)	5.0 (0.3)	3.9 (0.6)	<0.01*
Internet	7.9 (0.5)	3.7 (0.3)	4.2 (0.6)	<0.01*
Mail	12.7 (0.9)	10.1 (0.9)	2.6 (1.4)	0.07*
CAPi	14.2 (1.3)	8.9 (1.0)	5.2 (1.5)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 75. Adjusted IOI for Definition 2 of Disability: Ambulation**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	35.3 (1.5)	34.8 (1.3)	0.6 (2.0)	0.88
Self-Response	32.4 (1.4)	33.6 (1.7)	-1.2 (2.2)	0.88
Internet	33.3 (1.8)	33.7 (2.5)	-0.5 (3.0)	0.88
Mail	32.6 (2.6)	35.8 (2.9)	-3.2 (4.1)	0.88
CAPi	44.1 (3.6)	37.8 (3.5)	6.3 (4.7)	0.88

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 76. GDR for Definition 2 of Disability: Self-Care**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	5.4 (0.3)	2.4 (0.2)	2.9 (0.4)	<0.01*
Self-Response	4.1 (0.3)	2.0 (0.2)	2.1 (0.3)	<0.01*
Internet	3.7 (0.3)	1.5 (0.2)	2.3 (0.3)	<0.01*
Mail	5.4 (0.7)	4.0 (0.7)	1.5 (0.9)	0.12
CAPI	9.9 (1.1)	3.9 (0.6)	5.9 (1.3)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 77. Adjusted IOI for Definition 2 of Disability: Self-Care**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	48.5 (2.5)	41.3 (3.3)	7.2 (4.0)	0.36
Self-Response	41.3 (2.5)	37.0 (3.5)	4.3 (4.4)	0.75
Internet	42.6 (3.3)	41.4 (5.1)	1.3 (6.1)	0.84
Mail	38.4 (3.9)	32.8 (5.0)	5.6 (6.4)	0.75
CAPI	65.4 (6.0)	51.8 (7.4)	13.6 (8.9)	0.51

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 78. GDR for Definition 2 of Disability: Independent Living**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	8.3 (0.5)	5.1 (0.3)	3.2 (0.6)	<0.01*
Self-Response	7.2 (0.4)	4.5 (0.4)	2.7 (0.5)	<0.01*
Internet	6.7 (0.4)	3.5 (0.4)	3.2 (0.5)	<0.01*
Mail	8.9 (0.9)	8.1 (0.8)	0.8 (1.3)	0.53
CAPI	12.7 (1.3)	7.3 (0.9)	5.4 (1.7)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 79. Adjusted IOI for Definition 2 of Disability: Independent Living**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	46.9 (1.9)	40.7 (2.2)	6.3 (2.7)	0.10*
Self-Response	42.8 (1.8)	39.1 (2.5)	3.7 (2.9)	0.41
Internet	44.5 (2.4)	37.8 (3.3)	6.8 (3.5)	0.17
Mail	39.0 (3.7)	42.7 (3.9)	-3.7 (5.2)	0.48
CAPI	59.1 (5.1)	44.6 (4.4)	14.5 (6.9)	0.14

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 80. GDR for Definition 2 of Disability: Communication (Test Version Only)**

Mode	Test GDR
<b>Overall</b>	6.4 (0.3)
Self-Response	5.1 (0.3)
Internet	4.8 (0.4)
Mail	6.1 (0.8)
CAPI	11.0 (1.2)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 81. IOI for Definition 2 of Disability: Communication (Test Version Only)**

Mode	Test IOI
<b>Overall</b>	61.9 (2.2)
Self-Response	55.5 (3.0)
Internet	55.4 (3.4)
Mail	56.1 (6.2)
CAPI	77.1 (5.1)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

### 5.4.3 Response Reliability by Response Category in the Test Treatment

*RQ 7. What is the response reliability for the Test treatment, looking at movement between the four response categories (L-fold Index of Inconsistency) for each Disability type?*

For informational purposes, for the Test treatment we also examined IOI<sub>L</sub>, the L-fold index of inconsistency, which is based on the four individual response categories rather than a dichotomized yes/no measure. The IOI<sub>L</sub> for each Disability type can be found in Table 82. According to the IOI<sub>L</sub> scores, most of the Disability questions had a level of response inconsistency in the high range, with the exception of Ambulation. Communication and Vision Difficulty were among the Disability types with a higher IOI<sub>L</sub> score. The high level of response

inconsistency may be due to the use of an expanded list of response options in the Test treatment.

**Table 82. IOI<sub>L</sub> by Disability Type (Test Treatment Only)**

Disability Type	Test IOI <sub>L</sub>
Hearing	54.9 (1.9)
Vision	65.0 (1.7)
Cognition	55.1 (1.8)
Ambulation	46.2 (1.5)
Self-Care	56.2 (2.8)
Independent Living	58.3 (1.8)
Communication	67.8 (2.2)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

**Note:** Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

#### 5.4.4 Summary of Response Reliability Results

In summary, response reliability results did not consistently favor one version of the Disability question set over another.

Response reliability results varied depending on the definition of Disability used. We first summarize results for Definition 1, which is the definition recommended by the Washington Group. When GDR was used as the metric, the Test treatment generally had better response reliability than the Control. When IOI was used, the Test treatment had worse response reliability than the Control for overall Disability, but results varied when broken down by mode and Disability type. In terms of mode, IOI scores for overall Disability were significantly higher in Test within the CAPI mode. No significant differences between Test and Control were detected within any of the other modes for overall Disability. Additionally, there is evidence that the higher IOI score for overall Disability in the Test treatment was largely driven by specific Disability types, particularly Hearing, Vision, and Ambulation. No IOI differences between Test and Control were detected for Cognition, Self-Care, and Independent Living.

In most cases, the IOI scores for the Test treatment remained within the same range as those of the Control treatment, with most IOI scores falling between 20 and 50. In other words, Disability generally had only a moderate level of response inconsistency, regardless of treatment. There were, however, a few exceptions. First, IOI scores were often in the high inconsistency range within the CAPI mode, particularly in the Test treatment. Second, most IOI scores for two Disability types – Hearing and Vision – were in the high inconsistency range in the Test treatment. Vision also had IOI scores in the high range for the Control treatment in the internet mode.

We also evaluated the performance of an alternative definition of Disability (Definition 2). When Definition 2 of Disability was used, the Test treatment had consistently higher GDR scores compared to Control, indicating lower response reliability. As a reminder, GDR does not

account for differences between Test and Control in the proportion with a disability. When Definition 2 was used in Test, a much larger proportion of people were defined as having a disability compared to Control, which may explain why GDR was much higher in Test than Control, while the opposite was true when Definition 1 was used. GDR scores for overall Disability were higher in Test than Control when all modes were combined, as well as for each mode and Disability type. When IOI was used as the response reliability metric, however, the picture was less clear. The IOI score for overall Disability was higher in Test than in Control when all modes were combined, but results by mode and Disability type varied. In some modes (internet; CAPI), the Test treatment had significantly higher IOI scores than Control for overall Disability, while in others (mail), no significant difference between the treatments was detected. Similarly, for two Disability types (Vision; Independent Living), IOI scores were generally significantly higher in Test than Control (though results varied by mode). No significant differences were detected between Test and Control for any other Disability types (Hearing; Cognition; Ambulation; Self-Care), regardless of mode.

Notably, fewer IOI scores for the Test treatment fell within the high inconsistency range when Definition 2 of Disability was used, compared to Definition 1. Like Definition 1, IOI scores in Test were often in the high inconsistency range within the CAPI mode. However, while IOI scores for Hearing and Vision were often in the high inconsistency range for the Test treatment when Definition 1 of Disability was used, when Definition 2 was used, only Vision had multiple high IOI values in Test.

In summary, GDR results generally favored the Test treatment when Definition 1 of Disability was used, while they favored Control when Definition 2 of Disability was used. IOI results, on the other hand, were more mixed. When a significant difference was detected between the treatments, it always favored Control. However, in many cases there were no significant differences between the Test and Control IOI scores. This was true regardless of the definition of Disability employed. In other words, we did not find clear, consistent evidence that one treatment had significantly better response reliability than the other.

## **5.5 Other Metrics: Respondent Burden Results for Disability**

*RQ 8. Using internet and CAPI paradata, how does respondent burden differ between the Test treatment and the Control treatment?*

The full results of the respondent burden analysis can be found in Virgile et al. (2023). This section summarizes the relevant results for Disability.

The median completion time looked at the amount of time respondents who completed the survey spent on the Disability question screens. In the internet mode, the median completion time for Disability was 57 seconds for Control and 1 minute and 19 seconds for Test, for a difference of 22 seconds. In the CAPI mode, it was 40 seconds for Control and 1 minute and 11

seconds for Test, for a difference of 31 seconds.<sup>31</sup> Here it should be recalled that the Control treatment contained six Disability questions while the Test treatment contained seven. The Test treatment's longer completion time is likely a combination of respondents requiring more time per question and having more questions to answer. In addition, the Control questions had two response options per question while the Test questions had four options, all of which had to be read out loud by the interviewer in the CAPI mode. This likely contributed to the longer completion time for the Test treatment in CAPI.

We also examined breakoff rates and help screen access rates for the Disability questions in the internet mode. There were no significant differences in either rate between the Test and Control treatments.

## **6 CONCLUSIONS AND RECOMMENDATIONS**

The ACS is a critical source of national and sub-national disability estimates in the United States. While other federal surveys, such as the NHIS, collect more detailed data on disability, the ACS represents the primary source of state- and county-level disability estimates. Local, state, tribal, and federal agencies use ACS Disability data to plan and fund programs for people with disabilities. ACS Disability data are also used to evaluate other government programs and policies to ensure that they fairly and equitably serve the needs of all groups, as well as enforce laws, regulations, and policies against discrimination.

Since 2008, disability has been measured in the ACS using a standardized 6-item question set, known as the ACS-6. The ACS-6 was developed by the ACS Subcommittee on Disability Measurement, which was set up in 2003 under the auspices of OMB's Interagency Committee for the ACS. Efforts to develop a standardized Disability measure for inclusion in federal surveys emerged in response to data user concerns about Disability questions in the 2000 Census and in the ACS.

The WG-SS, which is the basis for the Test version of Disability in the current ACS Content Test, was developed around the same time as the ACS-6. Like the ACS-6, the WG-SS emerged in response to concerns about disability measurement across national surveys and censuses. The question set was developed by the Washington Group on Disability Statistics, which is a City Group created by the UN Statistical Commission, with its Secretariat located at the National Center for Health Statistics (NCHS).

In 2018, the American Community Survey sought input from 25 federal agencies regarding the survey's content. The National Center for Health Statistics proposed that the Disability measure be replaced with the WG-SS. Like all proposals, the request was reviewed to ensure that it met a statutory or regulatory need for data at small geographic levels or for small populations.

---

<sup>31</sup> Because these are median times, there are no standard errors or significance tests for these results.



Continuity in disability measurement allows for analysis of disability trends over time. A change to the ACS Disability question series would represent a break in series and thus warrants careful evaluation. A break in series may mean that the ACS would not publish 5-year estimates for any period where some respondents were asked the old questions and others the new questions. Therefore, it is possible that after 2024 (i.e., the 2020-2024 5-year estimates), the next 5-year estimates for Disability would not be until 2029 (i.e., the 2025-2029 5-year estimates). Regardless, there would continue to be 1-year estimates for the nation, states, and other areas of 65,000 people or more, with Disability estimates using the new questions starting with the 2025 data release.

Given the implications of a change to Disability, to accept the proposed change, the questions should a) represent an improvement to the Disability measure; and b) perform adequately in the ACS environment. The WG-SS has been extensively tested and validated within the United States and in a number of other countries. By Washington Group tallies, this question set has been included in over 90 national surveys and censuses worldwide (Miller et al., 2020). The WG-SS question set is recommended by the National Center for Health Statistics, the federal government's principal health statistics agency, and has already been incorporated into key health surveys in the United States, including the National Health Interview Survey (NHIS) and the National Survey of Family Growth (NSFG), as well as other federal surveys, such as the Census Bureau's experimental Household Pulse Survey. It is also the disability measure recommended by the United Nations, including the UN Statistical Division and the UN Economic Commission for Europe's Conference of European Statisticians. Overall, the WG-SS represents a validated measure of disability. Further, the WG-SS has potential advantages over the ACS-6, including international comparability and the use of a graded answer scale, which may more accurately reflect the continuum of functional difficulty and could provide the public with more granular data about disability status (provided disclosure avoidance and other Census Bureau standards for public release are met).<sup>32</sup>

Consequently, the criteria for whether the ACS should adopt the tested version of the WG-SS as its Disability measure primarily concerned the performance of the question set in the ACS environment. Two key decision criteria were employed in the field test for Disability: a) the response reliability of the Disability measure in the Test treatment (modified WG-SS), relative to the Control (ACS-6); and b) the item missing data rates for the Disability measure in the Test treatment versus Control.

Both the Control and Test versions of Disability in the field test consisted of a series of questions that asked respondents about difficulty they have with certain activities. A key difference between the two is that the Test (modified WG-SS) employed graded response

---

<sup>32</sup> It is important to note that the total measures and data products that would come from this question change are not yet known. Specifically, whether the PUMS data file and/or published tables would include the granularity of the four individual graded responses, as well as both Definition 1 and Definition 2, is dependent upon whether it meets disclosure avoidance and any other Census standards.

options, while the Control (ACS-6) used a dichotomized “yes”/“no” response option. Disability could thus be defined in more than one way in the Test treatment, depending on which response option was used as a cut-off for inclusion in the population with disabilities. The definition of Disability recommended by the Washington Group, Definition 1, considers someone to have a disability if they report “a lot of difficulty” or “cannot do at all” when asked about difficulty with certain activities. We also evaluated the performance of a second, broader definition of Disability (Definition 2) that considers someone to have a disability if they report any difficulty at all (e.g., “some difficulty,” “a lot of difficulty,” “cannot do at all”).

In terms of response reliability, the a priori criterion was that a lower Gross Difference Rate (GDR) and lower Index of Inconsistency (IOI) are generally preferred, as this would indicate better response reliability.<sup>33</sup> Results were mixed when Definition 1 of Disability was used. The Test treatment had lower GDR than Control in almost all cases, indicating better response reliability according to this metric. In terms of IOI, however, the Test treatment sometimes performed worse than Control. For example, IOI was higher in Test for Overall Disability when all modes were combined, and it was also higher in Test for certain Disability types and modes. It should be noted, however, that higher IOI for the Test treatment was primarily driven by one mode (CAPI), with the exception of Hearing difficulty, which was higher in the Test treatment across all modes. When Definition 2 of Disability was employed, the Test treatment performed consistently worse than Control when GDR was used as the response reliability metric. When using IOI scores instead, the results were mixed. In some cases, IOI was significantly higher in Test, while in others, no significant difference was detected between Test and Control.

Regardless of the definition of Disability employed, most IOI scores for Test and Control were in the moderate inconsistency range, though some IOI scores fell into the high inconsistency range, particularly in the Test treatment. Notably, IOI scores for the Test treatment were sometimes nominally lower under Definition 2 of Disability, compared to under Definition 1 of Disability. The variation in the proportion identified as having a disability contributed to differences in response reliability results between the two metrics (GDR and IOI), as well as differences in reliability results between definitions of Disability in the Test treatment (Definition 1 and Definition 2).<sup>34</sup> Overall, when all response reliability results are considered together, they present a mixed picture of the response reliability of the Test treatment. The decision criteria, however, called for weighing the results against the previously documented validity of the Washington Group Short Set. In consultation with the Disability Subcommittee, including NCHS, it was determined that the Test treatment met the first decision criterion.

---

<sup>33</sup> An important limitation of this field test is that only about 30 percent of the original respondents completed the Content Follow-up interview, leading to smaller groups for analysis of reliability.

<sup>34</sup> The percent identified as having a disability was 8.1 percent (SE = 0.2) in the Test treatment using Definition 1 of Disability, while it was 13.9 percent (SE = 0.3) in the Control treatment and 31.7 percent (SE = 0.4) in the Test treatment using Definition 2 of Disability.

Regarding the second decision criteria, lower item missing data rates were generally considered preferable. It was established a priori, however, that a small increase in item missing data rates was an acceptable consequence of switching from a binary response option (yes/no in Control) to a more complex graded response (no difficulty/some difficulty/a lot of difficulty/cannot do at all in the Test treatment), as well as the addition of another question (Communication difficulty). Field test results indicated that when all modes were combined, the Test treatment had higher item missing data rates than Control for Disability overall and for Hearing. Differences between Test and Control varied when analysis was broken down by mode. When item missing data rates were higher in Test, they were only higher by about 1 percentage point. In consultation with the Disability Subcommittee, it was determined that the Test treatment met the second decision criterion.

Overall, it was concluded that the Test treatment met both decision criteria. This suggests that the proposed question set performs adequately in the ACS environment based on the two pre-defined criteria (of response reliability and item missing data rates), including within self-response modes. Thus, the recommendation of the Disability Subcommittee is to move forward with the Test version of the Disability question set, which is based on the Washington Group Short Set and also includes Independent Living difficulty. Further, it is recommended that the standard definition of Disability used by the Washington Group (Definition 1) be used as the primary Census measure, as field test results did not indicate that an alternative approach (Definition 2) performed better. Whether estimates based on Definition 2 of Disability will also be released in data productions will depend on results of subsequent analyses.

It is important for ACS data users to understand that the revised Disability measure will not be comparable to the current ACS measure. Response distribution results from the field test suggest that the proposed revision to the Disability question series would result in a significant change in ACS estimates of the number and percentage of people in the U.S. with a disability. This is consistent with prior NCHS evaluations. The ACS estimate of functional disability prevalence in the U.S., which is used by various agencies for a number of different programs, is expected to be around 8 percent when using the revised measure (and the recommended definition of Disability), while the current (2022) ACS estimate is 13 percent. It should be noted that the magnitude of the difference between the current and revised Disability measure may also vary nontrivially by Disability type.<sup>35</sup> Cognitive interviews conducted prior to the field test suggest that the revised Disability measure identifies a smaller population and a population with a higher level of functional difficulty, relative to current ACS measure. These differences are important for other government agencies to understand, so that they can address the implications of the Disability measure change for their own surveys and programmatic uses.

---

<sup>35</sup> Further research is needed to evaluate whether/how the magnitude of difference between the current and revised Disability measure varies by sociodemographic characteristics such as age, race/ethnicity, and socioeconomic status.

## 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.

We would also like to thank the following individuals for their contributions to the testing of the Disability topic:

- Laryssa Mykyta for contributions to evaluation of the cognitive testing and development of the field test research plan.
- Joan Hill, Kelly Holder, Elizabeth Poehler, David Raglin, and Leann Weyl for contributions to critical review of the final report.

## 8 REFERENCES

- Albrecht, G. L., Seelman, K., & Bury, M. (Eds.). (2001). *Handbook of disability studies*. SAGE Publications.
- Altman, B. M., & Bernstein, A. (2008). *Disability and health in the United States, 2001-2005*. National Center for Health Statistics. <https://stacks.cdc.gov/view/cdc/6983>
- Altman, B. M. (Ed.). (2016). *International measurement of disability: Purpose, method, and application*. Springer.
- Biemer, P. (2011). *Latent class analysis of survey error*. John Wiley & Sons, Inc.
- Brault, M. W. (2009). *Review of changes to the measurement of disability in the 2008 American Community Survey*. U.S. Census Bureau. Retrieved February 24, 2022, from <https://www.census.gov/library/working-papers/2009/demo/brault-01.html>
- Brault, M., Stern, S., & Raglin, D. (2007). *Evaluation report covering disability (2006 American Community Survey Content Test Report P.4)*. U.S. Census Bureau. Retrieved February 24, 2022, from [https://www.census.gov/library/working-papers/2007/acs/2007\\_Brault\\_01.html](https://www.census.gov/library/working-papers/2007/acs/2007_Brault_01.html)
- Davis, L. J. (2006). *The disability studies reader*. Routledge.
- 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. Retrieved January 17, 2017, from <https://doi.org/10.2307/2336325>
- Keathley, D. (2022). *Specifications for selecting and weighting the 2022 American Community Survey Content Test sample*. U.S. Census Bureau. DSSD 2022 American Community Survey Memorandum Series #ACS22-S-21.
- Miller, K., Vickers, J. B., & Scanlon, P. (2020). *Comparison of American Community Survey and Washington Group disability questions*. National Center for Health Statistics. <https://wwwn.cdc.gov/QBank/Report.aspx?1216>.
- Madans, J. H., Loeb, M. E., & Altman, B. M. (2011). Measuring disability and monitoring the UN Convention on the Rights of Persons with Disabilities: The work of the Washington Group on Disability Statistics. *BMC Public Health*, 11(S4). <https://doi.org/10.1186/1471-2458-11-S4-S4>

- Madans, J. H. (2017, July 14). *Mainstreaming disability data: The Washington Group on Disability Statistics* [PowerPoint presentation]. UNSD Expert Group Meeting on Guidelines and Principles for the Development of Disability Statistics, New York, NY, United States.
- Office of Management and Budget (2006). *Standards and guidelines for statistical surveys*. Retrieved February 24, 2022, from [https://www.whitehouse.gov/wpcontent/uploads/2021/04/standards\\_stat\\_surveys.pdf](https://www.whitehouse.gov/wpcontent/uploads/2021/04/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). *2022 ACS Content Test: Round 1 and Round 2 Cognitive Testing*. Retrieved July 19, 2022, from [https://www.census.gov/content/dam/Census/library/working-papers/2022/acs/2022\\_Wilson\\_01.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2022/acs/2022_Wilson_01.pdf)
- RTI International (2022b). *Cognitive Testing for the 2022 ACS Content Test: Round 3 Briefing*. Retrieved August 29, 2022 from [https://www.census.gov/library/working-papers/2022/acs/2022\\_Wilson\\_02.html](https://www.census.gov/library/working-papers/2022/acs/2022_Wilson_02.html)
- Singer, P., & Ennis, S. (2003). *Census 2000 Content Reinterview Survey: Accuracy of data for selected population and housing characteristics as measured by reinterview*. U.S. Census Bureau. Retrieved June 7, 2023, from [https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/txe-program/B\\_5.pdf](https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/txe-program/B_5.pdf)
- 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*. U.S. Census Bureau. American Community Survey Memorandum Series #ACS23-RER-03. [https://www.census.gov/content/dam/Census/library/working-papers/2023/acs/2023\\_Spiers\\_01.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2023/acs/2023_Spiers_01.pdf)
- United Nations. (2017). *Principles and recommendations for population and housing censuses: Revision 3*. [https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles\\_and\\_Recommendations/Population-and-Housing-Censuses/Series\\_M67rev3-E.pdf](https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf)

United Nations Economic Commission for Europe. (2015). *Conference of European Statisticians recommendations for the 2020 census of population and housing*.

[https://unece.org/DAM/stats/publications/2015/ECECES41\\_EN.pdf](https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf)

United Nations General Assembly (2007, January 24). *Convention on the Rights of Persons with Disabilities: Resolution / adopted by the General Assembly (A/RES/61/106)*.

<https://www.refworld.org/docid/45f973632.html>

United Nations General Assembly (2015, October 21). *Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1)*.

<https://www.refworld.org/docid/57b6e3e44.html>

United Nations Economic and Social Commission for Asia and the Pacific (2012). *Incheon strategy to “make the right real” for persons with disabilities in Asia and the Pacific*.

<https://www.unescap.org/sites/default/d8files/knowledge-products/Incheon%20Strategy%20%28English%29.pdf>

U.S. Department of Health and Human Services. (2011). *U.S. Department of Health and Human Services implementation guidance on data collection standards for race, ethnicity, sex, primary language, and disability status*. Retrieved September 22, 2023, from <https://aspe.hhs.gov/sites/default/files/private/pdf/76331/index.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>

Virgile, M., Weng, C. F., Mills, G., & Spiers, S. (2023). *2022 American Community Survey Content Test evaluation report: Respondent burden*. U.S. Census Bureau. American Community Survey Memorandum Series #ACS23-RER-04. [https://www.census.gov/library/working-papers/2023/acs/2023\\_Virgile\\_01.html](https://www.census.gov/library/working-papers/2023/acs/2023_Virgile_01.html)

Washington Group on Disability Statistics (2020). *An introduction to the Washington Group on Disability Statistics questions sets*. <https://www.washingtongroup-disability.com/fileadmin/uploads/wg/Documents/primer.pdf>

Weeks, J. D., Dahlhamer, J. M., Madans, J. H., & Maitland, A. (2021). *Measuring disability: An examination of differences between the Washington Group Short Set on Functioning and the American Community Survey disability questions* (National Health Statistics Reports No. 161). National Center for Health Statistics. <https://www.cdc.gov/nchs/data/nhsr/nhsr161-508.pdf>

DRB Clearance Number—CDRB-FY23-ACSO003-B0069

World Health Organization (2001). *International Classification of Functioning, Disability and Health*. <https://icd.who.int/dev11/l-icf/en>



## APPENDIX A. Response Reliability Results for Same vs. Different CFU Respondent

### Response Reliability for Overall Disability: Same CFU Respondent

**Table 83. GDR for Definition 1 of Disability for Same CFU Respondent: Overall Disability**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
Overall	6.5 (0.4)	9.9 (0.5)	-3.4 (0.6)	<0.01*
Self-Response	5.5 (0.3)	9.3 (0.5)	-3.8 (0.6)	<0.01*
Internet	4.7 (0.3)	7.9 (0.5)	-3.2 (0.6)	<0.01*
Mail	8.9 (1.0)	15.0 (1.3)	-6.1 (1.6)	<0.01*
CAPI	10.5 (1.2)	12.0 (1.3)	-1.5 (1.7)	0.38

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 84. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Overall Disability**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
Overall	43.0 (2.0)	36.4 (1.6)	6.6 (2.6)	0.05*
Self-Response	39.6 (2.1)	37.2 (1.8)	2.4 (2.9)	0.80
Internet	39.1 (2.8)	38.1 (2.2)	1.0 (3.7)	0.80
Mail	41.9 (3.7)	37.6 (3.1)	4.3 (5.3)	0.80
CAPI	52.7 (5.2)	35.1 (3.4)	17.6 (6.5)	0.03*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### Response Reliability for Overall Disability: Different CFU Respondent

**Table 85. GDR for Definition 1 of Disability for Different CFU Respondent: Overall Disability**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
Overall	7.7 (1.3)	10.3 (1.3)	-2.6 (1.8)	0.36
Self-Response	6.6 (1.4)	9.2 (1.5)	-2.5 (2.3)	0.36
Internet	5.9 (1.7)	7.8 (1.2)	-1.9 (2.1)	0.36
Mail	8.3 (2.2)	13.3 (4.5)	-5.1 (5.3)	0.36
CAPI	9.7 (2.6)	13.2 (2.8)	-3.4 (3.6)	0.36

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 86. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Overall Disability**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	43.8 (4.9)	31.3 (4.0)	12.5 (6.5)	0.23
Self-Response	34.6 (6.0)	29.4 (4.8)	5.2 (8.1)	0.94
Internet	33.0 (7.4)	26.1 (4.7)	6.9 (8.5)	0.94
Mail	37.0 (9.1)	38.3 (12.5)	-1.3 (16.3)	0.94
CAPI	67.8 (9.1)	35.2 (7.2)	32.7 (12.7)	0.05*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

### Response Reliability by Disability Type: Same CFU Respondent

**Table 87. GDR for Definition 1 of Disability for Same CFU Respondent: Hearing**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.9 (0.2)	4.4 (0.3)	-2.4 (0.4)	<0.01*
Self-Response	1.8 (0.2)	4.1 (0.3)	-2.3 (0.4)	<0.01*
Internet	1.6 (0.2)	3.2 (0.3)	-1.6 (0.4)	<0.01*
Mail	2.4 (0.5)	7.8 (0.9)	-5.4 (1.1)	<0.01*
CAPI	2.6 (0.6)	5.4 (0.8)	-2.8 (1.1)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 88. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Hearing**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	64.8 (3.8)	46.2 (2.7)	18.6 (4.4)	<0.01*
Self-Response	64.0 (4.5)	46.1 (3.1)	17.9 (5.8)	<0.01*
Internet	60.9 (5.2)	47.6 (4.3)	13.2 (7.5)	0.10*
Mail	74.7 (7.7)	44.7 (4.8)	30.0 (9.8)	<0.01*
CAPI	67.3 (11.7)	46.6 (6.7)	20.7 (14.0)	0.14

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 89. GDR for Definition 1 of Disability for Same CFU Respondent: Vision**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.7 (0.2)	2.7 (0.2)	-1.0 (0.3)	<0.01*
Self-Response	1.5 (0.2)	2.4 (0.3)	-0.9 (0.3)	0.01*
Internet	1.5 (0.2)	2.0 (0.3)	-0.5 (0.3)	0.16
Mail	1.4 (0.3)	4.3 (0.6)	-2.8 (0.6)	<0.01*
CAPI	2.3 (0.6)	3.8 (0.6)	-1.5 (0.9)	0.16

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 90. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Vision**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	61.4 (5.7)	45.1 (3.5)	16.3 (7.1)	0.09*
Self-Response	57.5 (6.0)	50.2 (4.2)	7.3 (7.9)	0.72
Internet	60.9 (6.7)	52.9 (5.7)	8.0 (8.7)	0.72
Mail	45.7 (11.6)	45.7 (6.7)	<0.1 (13.5)	1.00
CAPI	74.7 (10.5)	37.0 (5.9)	37.7 (12.6)	0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 91. GDR for Definition 1 of Disability for Same CFU Respondent: Cognition**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.3 (0.2)	4.9 (0.3)	-2.6 (0.4)	<0.01*
Self-Response	2.1 (0.2)	4.4 (0.3)	-2.3 (0.4)	<0.01*
Internet	2.0 (0.2)	3.9 (0.4)	-1.9 (0.5)	<0.01*
Mail	2.4 (0.6)	6.3 (0.8)	-3.9 (0.9)	<0.01*
CAPI	3.1 (0.7)	6.7 (0.9)	-3.6 (1.2)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 92. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Cognition**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	52.2 (4.4)	46.4 (2.4)	5.7 (5.0)	0.77
Self-Response	51.6 (5.2)	44.4 (2.9)	7.2 (6.0)	0.77
Internet	52.2 (5.8)	47.9 (3.8)	4.3 (7.3)	0.87
Mail	49.6 (9.9)	38.0 (4.6)	11.6 (10.2)	0.77
CAPI	53.7 (8.7)	52.1 (5.1)	1.6 (10.2)	0.87

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 93. GDR for Definition 1 of Disability for Same CFU Respondent: Ambulation**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	3.6 (0.2)	5.6 (0.3)	-2.2 (0.4)	<0.01*
Self-Response	3.0 (0.2)	5.0 (0.3)	-2.0 (0.3)	<0.01*
Internet	2.3 (0.2)	3.7 (0.3)	-1.4 (0.3)	<0.01*
Mail	5.8 (0.7)	10.2 (1.0)	-4.4 (1.2)	<0.01*
CAPI	6.1 (0.9)	8.7 (1.1)	-2.6 (1.3)	0.04*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 94. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Ambulation**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	41.7 (2.6)	34.5 (1.4)	7.2 (2.9)	0.06*
Self-Response	38.9 (2.7)	33.9 (1.8)	5.1 (3.3)	0.38
Internet	37.5 (3.4)	34.1 (2.7)	3.4 (4.2)	0.43
Mail	42.2 (5.1)	35.8 (2.9)	6.4 (5.6)	0.43
CAPI	48.6 (5.9)	36.5 (3.9)	12.1 (7.0)	0.33

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 95. GDR for Definition 1 of Disability for Same CFU Respondent: Self-Care**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.1 (0.2)	2.2 (0.2)	-1.1 (0.3)	<0.01*
Self-Response	0.9 (0.1)	1.9 (0.2)	-1.0 (0.2)	<0.01*
Internet	0.8 (0.1)	1.4 (0.2)	-0.7 (0.2)	0.01*
Mail	1.4 (0.4)	3.9 (0.7)	-2.5 (0.8)	<0.01*
CAPI	2.0 (0.6)	3.3 (0.7)	-1.3 (0.9)	0.15

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 96. IOI for Definition 1 of Disability for Same CFU Respondent: Self-Care**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	42.2 (5.6)	40.1 (3.7)	2.1 (6.2)	0.89
Self-Response	35.9 (5.4)	36.7 (3.7)	-0.8 (6.0)	0.89
Internet	36.3 (5.9)	41.3 (5.5)	-5.0 (8.0)	0.89
Mail	34.9 (11.3)	32.4 (5.3)	2.5 (11.8)	0.89
CAPI	60.7 (15.5)	49.4 (8.8)	11.3 (17.3)	0.89

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 97. GDR for Definition 1 of Disability for Same CFU Respondent: Independent Living**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.6 (0.2)	4.9 (0.4)	-2.4 (0.5)	<0.01*
Self-Response	2.1 (0.2)	4.4 (0.4)	-2.3 (0.4)	<0.01*
Internet	1.8 (0.2)	3.5 (0.4)	-1.6 (0.4)	<0.01*
Mail	3.1 (0.7)	8.0 (0.8)	-5.0 (1.0)	<0.01*
CAPI	4.5 (0.8)	6.9 (1.0)	-2.3 (1.4)	0.09*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 98. Adjusted IOI for Definition 1 of Disability for Same CFU Respondent: Independent Living**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	36.1 (3.0)	40.6 (2.4)	-4.6 (3.9)	0.75
Self-Response	33.2 (3.0)	39.3 (2.7)	-6.1 (4.1)	0.55
Internet	34.9 (3.8)	38.2 (3.5)	-3.3 (4.9)	0.91
Mail	30.1 (5.6)	42.5 (4.0)	-12.4 (6.7)	0.33
CAPI	43.2 (7.7)	44.3 (4.9)	-1.1 (9.5)	0.91

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 99. GDR for Definition 1 of Disability for Same CFU Respondent: Communication (Test Version Only)**

Mode	Test GDR
<b>Overall</b>	0.9 (0.1)
Self-Response	0.8 (0.1)
Internet	0.6 (0.1)
Mail	1.2 (0.3)
CAPI	1.4 (0.5)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 100. IOI for Definition 1 of Disability for Same CFU Respondent: Communication (Test Version Only)**

Mode	Test IOI
<b>Overall</b>	51.2 (6.8)
Self-Response	45.1 (6.2)
Internet	47.3 (6.3)
Mail	41.0 (11.4)
CAPI	73.5 (18.8)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Response Reliability by Disability Type: Different CFU Respondent****Table 101. GDR for Definition 1 of Disability for Different CFU Respondent: Hearing**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.6 (0.6)	5.6 (1.0)	-3.0 (1.1)	0.04*
Self-Response	3.0 (0.8)	4.9 (1.2)	-1.9 (1.5)	0.37
Internet	2.8 (1.1)	4.4 (1.3)	-1.6 (1.8)	0.37
Mail	3.4 (1.0)	6.4 (1.9)	-3.0 (2.2)	0.37
CAPI	1.8 (1.0)	7.3 (2.3)	-5.6 (2.5)	0.11

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 102. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Hearing**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	66.2 (10.0)	42.3 (8.2)	23.9 (12.5)	0.22
Self-Response	59.8 (11.9)	39.0 (8.6)	20.8 (15.4)	0.36
Internet	60.4 (15.6)	43.3 (10.0)	17.1 (18.6)	0.36
Mail	58.5 (15.3)	32.9 (11.7)	25.6 (19.9)	0.36
CAPI	100.6 (0.8) <sup>36</sup>	48.6 (19.5)	51.9 (19.5)	0.04*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 103. GDR for Definition 1 of Disability for Different CFU Respondent: Vision**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	3.3 (0.8)	3.8 (0.8)	-0.5 (1.1)	0.64
Self-Response	3.5 (1.1)	2.4 (0.6)	1.1 (1.2)	0.64
Internet	3.7 (1.5)	2.7 (0.8)	1.1 (1.7)	0.64
Mail	2.9 (1.0)	1.7 (0.8)	1.3 (1.3)	0.64
CAPI	2.8 (1.2)	7.0 (2.1)	-4.2 (2.4)	0.43

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

<sup>36</sup> Typically, IOI should fall between 0 – 100. However, in this case, due to having few CAPI cases in which a different household member responded in CFU, *a* (the number of persons who had a hearing disability in both interviews) was equal to 0 and therefore IOI came out to slightly over 100. This also occurred for Communication in the internet mode (Table 114Table 114). These results should be interpreted with caution.

**Table 104. IOI for Definition 1 of Disability for Different CFU Respondent: Vision**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	78.8 (8.8)	47.8 (8.0)	30.9 (12.5)	0.05*
Self-Response	74.6 (11.3)	34.5 (8.7)	40.1 (14.1)	0.02*
Internet	87.5 (10.3)	46.6 (13.4)	40.8 (17.2)	0.05*
Mail	52.4 (18.3)	14.8 (8.8)	37.6 (21.4)	0.16
CAPI	90.7 (11.9)	70.9 (11.4)	19.8 (16.6)	0.23

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 105. GDR for Definition 1 of Disability for Different CFU Respondent: Cognition**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.2 (0.4)	5.8 (1.1)	-4.6 (1.2)	<0.01*
Self-Response	1.2 (0.4)	3.9 (1.0)	-2.7 (1.2)	0.04*
Internet	0.8 (0.5)	3.9 (1.2)	-3.1 (1.3)	0.04*
Mail	2.0 (0.9)	3.7 (1.5)	-1.7 (1.7)	0.32
CAPI	1.2 (0.6)	10.2 (2.4)	-8.9 (2.4)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 106. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Cognition**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	31.1 (8.5)	35.3 (8.3)	-4.2 (12.7)	0.91
Self-Response	25.4 (8.2)	23.6 (7.2)	1.8 (11.6)	0.91
Internet	19.5 (11.2)	21.2 (7.2)	-1.7 (14.6)	0.91
Mail	35.8 (11.5)	38.9 (16.8)	-3.1 (22.1)	0.91
CAPI	53.5 (22.1)	63.5 (13.2)	-10.0 (26.6)	0.91

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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



**Table 107. GDR for Definition 1 of Disability for Different CFU Respondent: Ambulation**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	4.6 (1.0)	6.8 (1.2)	-2.2 (1.5)	0.57
Self-Response	4.3 (1.0)	5.1 (1.2)	-0.7 (1.6)	0.76
Internet	4.2 (1.2)	3.7 (0.8)	0.4 (1.5)	0.76
Mail	4.7 (1.8)	9.3 (3.3)	-4.5 (3.9)	0.73
CAPI	5.1 (2.4)	10.9 (2.8)	-5.8 (3.3)	0.39

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 108. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Ambulation**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	48.3 (6.5)	37.3 (4.7)	10.9 (7.5)	0.54
Self-Response	42.1 (7.6)	31.8 (5.9)	10.3 (9.3)	0.54
Internet	43.5 (10.2)	29.3 (6.3)	14.2 (12.2)	0.54
Mail	39.1 (11.1)	36.4 (12.4)	2.6 (15.6)	0.87
CAPI	62.7 (11.1)	46.5 (9.2)	16.1 (14.7)	0.54

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 109. GDR for Definition 1 of Disability for Different CFU Respondent: Self-Care**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	1.0 (0.3)	4.3 (0.9)	-3.4 (0.9)	<0.01*
Self-Response	1.0 (0.5)	2.7 (0.7)	-1.7 (0.9)	0.16
Internet	0.7 (0.7)	2.2 (0.9)	-1.6 (1.1)	0.29
Mail	1.6 (0.9)	4.1 (2.5)	-2.5 (2.6)	0.33
CAPI	1.0 (0.5)	8.1 (2.1)	-7.1 (2.2)	<0.01*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 110. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Self-Care**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	46.8 (12.6)	49.3 (8.6)	-2.5 (15.5)	0.98
Self-Response	47.2 (20.7)	39.6 (8.7)	7.6 (22.7)	0.98
Internet	43.1 (36.6)	42.3 (14.1)	0.8 (38.4)	0.98
Mail	52.0 (23.9)	36.3 (18.4)	15.7 (25.5)	0.98
CAPI	46.8 (12.6)	49.3 (8.6)	-2.5 (15.5)	0.98

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 111. GDR for Definition 1 of Disability for Different CFU Respondent: Independent Living**

Mode	Test GDR	Control GDR	Difference	Adj. P-value
<b>Overall</b>	2.7 (0.6)	6.9 (1.1)	-4.1 (1.2)	<0.01*
Self-Response	2.7 (0.7)	5.3 (1.0)	-2.6 (1.3)	0.12
Internet	1.9 (0.6)	4.0 (1.0)	-2.1 (1.2)	0.17
Mail	4.5 (1.4)	8.9 (2.9)	-4.4 (3.2)	0.17
CAPI	2.7 (1.3)	10.2 (2.4)	-7.5 (2.6)	0.02*

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 112. Adjusted IOI for Definition 1 of Disability for Different CFU Respondent: Independent Living**

Mode	Test IOI	Control IOI	Difference	Adj. P-value
<b>Overall</b>	28.0 (5.3)	41.5 (5.2)	-13.5 (7.4)	0.34
Self-Response	25.0 (6.4)	37.9 (5.1)	-12.9 (8.4)	0.49
Internet	20.7 (8.1)	34.3 (8.2)	-13.6 (10.6)	0.60
Mail	31.1 (9.4)	43.1 (10.4)	-12.0 (13.6)	0.62
CAPI	36.5 (14.9)	46.6 (11.2)	-10.1 (20.3)	0.62

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

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

**Table 113. GDR for Definition 1 of Disability for Different CFU Respondent: Communication (Test Version Only)**

Mode	Test GDR
<b>Overall</b>	1.4 (0.5)
Self-Response	1.5 (0.7)
Internet	1.7 (1.0)
Mail	1.0 (0.6)
CAPI	1.3 (0.6)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

**Table 114. IOI for Definition 1 of Disability for Different CFU Respondent: Communication (Test Version Only)**

Mode	Test GDR
<b>Overall</b>	69.4 (11.9)
Self-Response	78.8 (15.9)
Internet	100.8 (0.6) <sup>37</sup>
Mail	42.1 (13.5)
CAPI	54.8 (17.8)

**Source:** U.S. Census Bureau, 2022 American Community Survey Content Test | DRB No. CBDRB-FY23-ACSO003-B0069

Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses.

<sup>37</sup>Typically, IOI should fall between 0 – 100. However, in this case, due to having few CAPI cases in which a different household member responded in CFU, *a* (the number of persons who had a communication disability in both interviews) was equal to 0 and therefore IOI came out to slightly over 100. This also occurred for Hearing in CAPI (Table 102Table 102). These results should be interpreted with caution.