

April 17, 2024

ACS Research & Evaluation Analysis Plan (REAP)

**Sexual Orientation and Gender
Identity Test**

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. BACKGROUND.....	1
2.1 Cognitive Testing.....	1
2.2 Data Collection and Field Testing.....	2
2.3 Content Follow-Up.....	4
2.4 Question Content for SOGI Field Test.....	6
2.4.1 Guidelines and Recommendations for Measuring SOGI.....	6
2.4.2 Question Content for Gender Identity and Sexual Orientation.....	7
2.4.3 Internet Write-In Box Display.....	9
2.4.4 Additional Questionnaire and Instrument Changes.....	11
2.4.5 Qualitative Questions.....	13
3. LITERATURE REVIEW.....	14
3.1 SOGI Questions in Federal Surveys.....	15
3.2 Additional SOGI Testing.....	20
4. RESEARCH QUESTIONS AND METHODOLOGY.....	21
4.1 Sample Design.....	21
4.2 Experimental Design.....	22
4.3 Research Questions.....	23
4.4 Analysis Metrics.....	31
4.4.1 Unit-Level Analysis.....	32
4.4.2 Benchmarks.....	33
4.4.3 Item Missing Data Rates.....	36
4.4.4 Response Distributions.....	36
4.4.5 Response Reliability.....	37
4.5 Weighting.....	41
5. DECISION CRITERIA.....	41
6. ASSUMPTIONS AND LIMITATIONS.....	45
6.1 Assumptions.....	45
6.2 Limitations.....	45
7. POTENTIAL CHANGES TO ACS.....	46
8. REFERENCES.....	47

1. INTRODUCTION

In December of 2022, the U.S. Census Bureau requested funding to research the effect of including questions about Sexual Orientation and Gender Identity (SOGI) on the American Community Survey (ACS). The request was made amid growing interest in the federal government and survey research community to have more detailed information about sexual and gender minority populations. The Census Bureau received funding as part of the fiscal year 2023 Omnibus Appropriations bill to examine the feasibility of collection SOGI information on the ACS.

The purpose of the 2024 ACS SOGI Test is to test the feasibility of adding SOGI questions to the ACS. This test will include cognitive testing and field testing. We will test the wording, format, and placement of the SOGI questions, investigate translation issues for Spanish questionnaires, and evaluate potential issues with proxy reporting.

The test will be conducted in two parts: first focusing on self-response and later focusing on interviewer-administered data collection. This analysis plan is specific to the self-response portion of the field test; the interviewer-administered portion of the field test will be detailed in a separate document.

2. BACKGROUND

The ACS, conducted by the Census Bureau, is the premiere source for detailed population and housing information about our nation. Currently, the ACS does not collect information regarding sexual orientation or gender identity status.

The Office of Management and Budget's (OMB) Standards and Guidelines for Statistical Surveys (OMB, 2006) and the Census Bureau's Statistical Quality Standards (U.S. Census Bureau, 2022a) require that any proposed changes to the ACS first undergo cognitive testing and field testing.

2.1 Cognitive Testing

Cognitive testing of the SOGI questions will focus on three main areas:

- Spanish-language translation
- Multi-mode measurement
- Proxy response

About half of the total respondents recruited will be straight and have gender identities that match their sex assigned at birth. The other half of respondents recruited will be a sexual

and/or gender minority.¹ Since proxy reporting is a key research area, we will aim for roughly 80 percent of the respondents to be from a multi-person household.

We will also aim for demographic diversity on the following characteristics:

- Age and education
- Race/ethnicity (English interviews only)
- Hispanic origin (Spanish stateside interviews only)
- Geographic diversity in region of the country
- Household composition (e.g., number of people, relationships of household members, and whether there are children)

The cognitive testing interviews will be conducted by RTI International and partner Research Support Services (RSS). RSS will conduct 218 cognitive interviews—64 stateside English interviews, 80 stateside Spanish interviews, and 74 Puerto Rican Spanish interviews. Interviewing is expected to start mid-January and run through May of 2024.

Normally, the results from cognitive testing help fine-tune the content or mail materials used in field testing. However, the cognitive testing results will not be available in time for the 2024 SOGI field test and will serve as a resource for future SOGI research and for translation of the SOGI questions into Spanish.

2.2 Data Collection and Field Testing

The 2024 ACS SOGI Test will mirror the production ACS data collection process as much as possible, with some slight variations.

Data Collection for ACS Production

The monthly ACS production sample consists of approximately 295,000 housing unit addresses, which we refer to as a panel. Data collection for each panel occurs over three months. During the first two months, respondents are asked to self-respond by internet or paper questionnaire. In the third month, a Computer-Assisted Personal Interviewing (CAPI) nonresponse follow-up operation begins. There is also a Telephone Questionnaire Assistance (TQA) option throughout the data collection period.

Data Collection for the 2024 ACS SOGI Test

The 2024 ACS SOGI Test will occur in parallel with data collection activities for the August and September 2024 ACS production panels. The test will use a nationally representative sample, independent of the ACS production sample, distributed evenly among four treatments. The

¹ Sexual and gender minority populations include, but are not limited to, those who identify as lesbian, gay, bisexual, transgender, queer, intersex, Two-Spirit, nonbinary, and/or asexual.

treatments are defined by their unique combination of question wording for the Current Gender question and internet write-in box display for the Current Gender and Sexual Orientation questions. More information about the sample is provided in Section 4.1, and more information about the treatments is provided in Section 4.2. Data collection will follow the same protocols as ACS production, except where noted.

The data collection protocols that will be the same are:

- In the first and second month of both panels in data collection, sampled addresses will be mailed up to five mailings to encourage self-response by internet or paper questionnaire.
- A subsample of nonresponding addresses after the second month of data collection will be selected for the CAPI nonresponse follow-up operation. During CAPI, Census Bureau field representatives will conduct interviews by personal visit or phone. The CAPI instrument will be in English and Spanish.
- The internet instrument will be in English and Spanish. There will be a paper questionnaire in English.

The data collection protocols that will be different are:

- The first mailing for the August and September SOGI panels will be mailed a week later than the first mailings for the August and September ACS production data collection panels.²
- There will be no paper questionnaires in Spanish.³
- If respondents call TQA and opt to complete the survey over the phone, the interviewers will conduct the survey using the ACS production questionnaire.⁴ Since the TQA interviews will not include SOGI questions, they will be excluded from the analysis.
- Due to budget restraints, the CAPI operation, which normally occurs in the third month of data collection, will be delayed, and start in spring 2025.
- The 2024 ACS SOGI Test will not have the Telephone Failed-Edit Follow-Up (FEFU) operation. In ACS production, this operation follows up on households that provided incomplete information or reported more than five people on the roster of a paper questionnaire.⁵

² Due to the large sample size for this test and limits on processing capacity, mailout will be delayed to not interfere with ACS production mailout.

³ In 2022, 537 Spanish questionnaires were mailed back out of all mailable cases. Based upon this rate, we project that 74 Spanish questionnaires would be mailed back in the SOGI Test, which would not be cost-effective.

⁴ There is not a questionnaire separate from ACS production due to operational complexity and low TQA response.

⁵ The information obtained from the FEFU operation improves accuracy in a production environment but confounds the evaluation of respondent behavior in the test environment. For paper questionnaires, where the household size is six or more (up to 12), we will only collect name, age, relationship to Person 1, and date of birth of these additional persons but not detailed information as we do in the FEFU operation for

- Data collection will not include remote Alaska, Group Quarters, or Puerto Rico.
- The 2024 ACS SOGI Test will have a reinterview operation called the Content Follow-Up (CFU) operation. The reinterviews will be used to measure response reliability and test the accuracy of proxy reporting. We describe the CFU operation in more detail in Section 2.3.

Additional information about ACS data collection procedures can be found in the ACS and Puerto Rico Community Survey (PRCS) Design and Methodology Report (U.S. Census Bureau, 2022b).

2.3 Content Follow-Up

To measure response reliability, we will attempt a CFU reinterview with every household that completes an original interview and meets the eligibility requirements as described in Contard (2024).

CFU Timing

The CFU reinterview will be conducted approximately two to five weeks after the original interview. A case will be sent to the CFU operation one to two weeks after the original interview and must be completed within three weeks after being sent to CFU. This timing attempts to balance two competing needs: to minimize the possibility of real changes in answers due to a change in life circumstances between the two interviews and to minimize the possibility that the respondent repeats their previous answer based on their recollection of the original interview response rather than considering the most appropriate answer.

CFU Modes

Historically, CFU reinterviews have been conducted by phone. Low response rates in previous tests and the potential for a high workload in our phone centers for this test have caused us to modify the mode of collecting the CFU reinterview.

For the 2024 ACS SOGI Test, CFU reinterviews will be conducted by telephone for households that responded by mail in the original interview and will be conducted by internet for households that responded by internet in the original interview.

CFU Proxy Response

For some CFU cases, we will interview the same household member who provided information in the original interview, and for other CFU cases we will interview a different household member than the one who provided information in the original interview. The reason for this is to test the accuracy of proxy reporting for the SOGI questions. For all households with only

one adult, that person will be assigned as the CFU respondent.⁶ Households with more than one adult will be first stratified into two groups:

- Households with a member who was identified as a sexual or gender minority
- All other multiple-adult households

Within each stratum, households will be assigned evenly between two CFU respondent treatments. In one, the original respondent will be assigned as the CFU respondent. In the other CFU respondent treatment, a different adult in the household will be randomly selected and assigned as the CFU respondent.⁷ If there are any adults in the household who were identified as a sexual or gender minority and were not the original respondent, the CFU respondent will be randomly selected from those persons. Otherwise, the CFU respondent will be randomly selected from among all adults in the household who were not the original respondent.

CFU Content

The CFU instrument will not include all ACS questions, but for context and analysis purposes, it will include some ACS production questions in addition to the ones being tested for revision or addition. Questions will be asked in the reinterview regardless of whether the question was left blank in the original interview.

The questions that will be asked in the CFU reinterview include:

- Relationship
- Age
- Sex at Birth
- Current Gender
- Hispanic Origin
- Race
- Place of Birth
- Educational Attainment
- Health Insurance Coverage
- Disability
- Marital Status/History
- Sexual Orientation
- Fertility
- Wages

In the CFU reinterviews, we will ask the same version of the SOGI questions asked in the original interview for each treatment.

⁶ This includes households that have one adult and persons younger than 18.

⁷ Note that the age requirement of determining a household with a sexual or gender minority member is 15, however, this person cannot be the assigned CFU respondent unless they are at least 18.

For more information on the CFU operation and eligibility requirements, see Contard (2024).

2.4 Question Content for SOGI Field Test

2.4.1 Guidelines and Recommendations for Measuring SOGI

The SOGI questions in this test were developed using guidelines and recommendations for measuring SOGI in federal surveys. Those guidelines and recommendations are based on previous research on measuring SOGI, expert advice, and existing SOGI questions in federal surveys.

Federal agencies interested in collecting SOGI data independently researched how best to measure SOGI. To coordinate SOGI data collection efforts across the federal government, OMB convened the Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys (IWG) in 2015. The group was also convened to address the methodological issues surrounding the collection and measurement of SOGI data (IWG, 2016a).

The IWG published three working papers in 2016. The first paper, “Current Measures of SOGI in Federal Surveys” (IWG, 2016a), gave an overview of the current SOGI concepts and measures in federal household surveys. The second paper, “Evaluations of SOGI Measures” (IWG, 2016b), was a meta-analysis of assessments of SOGI measurement error in federal surveys. It also described how federal surveys defined SOGI, any measurement challenges, and areas of future research. The third paper, “Toward a Research Agenda” (IWG, 2016c), proposed research priorities and strategies for improving SOGI measurement in federal surveys. The top proposed priorities included 1) question terminology related to gender identity, 2) proxy reporting, and 3) translation of SOGI questions into other languages.

In 2020, the National Academies of Sciences, Engineering, and Medicine (NASEM) issued a report calling on the federal government to develop standards for collecting SOGI information. The National Institutes of Health then asked NASEM to have a committee of experts review the current SOGI measures in surveys and provide guidelines for collecting these data in various settings. The “Measuring Sex, Gender Identity, and Sexual Orientation” report (NASEM, 2022) was the outcome of this review. It provides recommended questions for measuring sex, gender identity, and sexual orientation, as well as future research areas.

Drawing from the research of these and other working groups, OMB issued the “Recommendations on the Best Practices for the Collection of Sexual Orientation and Gender Identity Data on Federal Statistical Surveys” report (OMB, 2023). It detailed considerations for including SOGI items on surveys, example modules for asking about SOGI, and guidelines for reporting and safeguarding SOGI data.

2.4.2 Question Content for Gender Identity and Sexual Orientation

The two versions we will test for the Gender Identity series of questions are provided in Figure 1. Both versions have the same wording for the Sex at Birth question and Current Gender question, but the Current Gender instructions for Version 1 instruct the respondent to mark only one box while the instructions for Version 2 say to mark one or more boxes.

Figure 1. Gender Identity Questions (Paper): Version 1 (Left) and Version 2 (Right)

4 What sex was Person 1 assigned at birth?
Mark (X) ONE box.

Male Female

5 What is Person 1's current gender?
Mark (X) ONE box. If this person is less than 15 years old, SKIP to the next question.

Male
 Female
 Transgender
 Nonbinary
 This person uses a different term – *Specify*

4 What sex was Person 1 assigned at birth?
Mark (X) ONE box.

Male Female

5 What is Person 1's current gender?
Mark (X) one or more boxes. If this person is less than 15 years old, SKIP to the next question.

Male
 Female
 Transgender
 Nonbinary
 This person uses a different term – *Specify*

The two-step approach to measuring gender identity uses a sex at birth question and a current gender question as a pair to provide estimates of cisgender men and women, transgender men and women, and those who identify using other terms. This approach has been used on several federal surveys, including the National Crime Victimization Survey (NCVS) and the Household Pulse Survey (HPS), and is supported in the NASEM (2022) report.⁸

Also as recommended in the NASEM (2022) report, in the test questionnaire, the Sex at Birth question replaces the Sex question from ACS production. The Sex at Birth question is included, along with Current Gender, in the basic demographic section, being asked after Age and before Hispanic Origin and Race.

The Sex at Birth question differs from the OMB (2023) recommendation by not referencing a person's original birth certificate, listing male before female, and allowing for proxy reporting. Versions of the question without the birth certificate language have appeared in other surveys and testing.⁹ It prevents confusion in non-English speakers that may not have a birth certificate available or parents of children with an X listed for sex. The Sex at Birth question also keeps the

⁸ For a detailed discussion on the two-step approach to measuring gender identity compared to a one-step approach, see NASEM (2022), Chapter 6.

⁹ The Census Bureau cognitively tested wording without the birth certificate language for the Current Population Survey (Holzberg et al., 2019). Versions of the question without the language have been fielded in the National Institutes of Health All of Us panel, the National HIV Behavioral Surveillance System, the National Adult Tobacco Survey, and the Canada and Scotland censuses.

original order of the male and female categories from the previous sex question on the ACS to prevent any accidental keying from interviewers who are used to the existing order.

The Current Gender question is a mix of recommendations from NASEM (2022) and OMB (2023). The question stem aligns with NASEM (2022), except for changes to allow for proxy reporting. However, the response options align with the OMB (2023) recommendations, except for the addition of a nonbinary category. Nonbinary is a growing identity – approximately 1.2 million adults in the U.S. identify as nonbinary (Wilson and Meyer, 2021).

The mark one instruction in Version 1 of the Current Gender question follows the NASEM (2022) report. The mark-all-that-apply option in Version 2 of the question follows the OMB (2023) recommendations. However, few federal surveys use a mark-all-that-apply instruction for gender identity, and the California Health Interview Survey decided not to implement it due to concerns about erroneous increases in gender minority reporting (Hughes et al., 2023). Therefore, we are testing this option against the mark one option to further research in this area.

On the automated instruments (i.e., not the paper questionnaire), a confirmation question will be asked after the Current Gender question confirming the respondent's answers to both questions. Other surveys ask a confirmation question when Sex at Birth and Current Gender do not match to make sure the respondent did not accidentally click or say different answers (i.e., give a false positive). The effects of this misclassification error can be substantial because the percentage of gender minorities is small. For the 2024 ACS SOGI Test, we are asking the confirmation question for all responses to test if there is additional error from respondents accidentally clicking or saying the same answer to both questions (i.e., giving a false negative).

We will test one version of the Sexual Orientation question, provided in Figure 2. This version aligns with the OMB (2023) recommendations except for minor changes to work in the ACS instruments.¹⁰ It is also similar to versions used in other federal surveys (see Section 3.1) and to the recommendation from the NASEM (2022) report.

¹⁰ The question was reworded to allow for proxy reporting, and the “Don’t Know” response option was removed since respondents can skip questions in the survey.

Figure 2. Sexual Orientation Question (Paper)

26 Which of the following best represents how this person thinks of themselves?
Mark (X) ONE box.

Gay or lesbian

Straight, that is not gay or lesbian

Bisexual

This person uses a different term – *Specify* ↕

The wording of the Sexual Orientation question will be the same across all treatments. It will be asked after questions on marital status and history and before the question on fertility. This placement is part of the questions that apply only to persons age 15 or older.

2.4.3 Internet Write-In Box Display

The write-in box for both the Current Gender and Sexual Orientation questions will be displayed two different ways for internet respondents. Figure 3 and Figure 4 show the two different ways for the Sexual Orientation write-in box, but this will also apply to the Current Gender write-in boxes.

The first display option is shown in Figure 3. In this version, the write-in box is visible but greyed-out. The respondent must click on the box to write in an answer. This layout is the standard display option for write-in boxes on the ACS.

Figure 3. Internet Display Option for the Write-in Box: Version A

The screenshot displays the American Community Survey interface. At the top left is the United States Census Bureau logo. The title "American Community Survey" is centered at the top. A navigation bar includes "Instructions", "FAQs", and "Save and Log Out". On the right side, a sidebar titled "Where You Are" contains sections for "Basic Info", "Housing Questions", and "Person Info". The main content area features question 26: "Which of the following best represents how you think of yourself? (Help)". The options are: "Gay or lesbian", "Straight, that is not gay or lesbian", "Bisexual", and "I use a different term" with a right-pointing arrow and the text "Write in" next to it. Below the "I use a different term" option is a text input box. At the bottom of the question area are two buttons: "Previous" with a left arrow and "Next" with a right arrow. The footer contains "Contact Us" on the left and "Accessibility", "Privacy", and "Security" on the right.

The second display option for the write-in box is shown in Figure 4. In Version B, the write-in box pops up as a separate question only after a respondent checks the “I use a different term” response option and clicks the “next” button. That is, when the respondent first reaches this screen, they will not have a visual cue that a write-in response is available if they mark “I use a different term.” Fewer respondents may give “protest” write-ins (i.e., responses protesting the question content) if the existence of a write-in box is not immediately apparent.

Figure 4. Internet Display Option for the Write-in Box: Version B

The screenshot displays the American Community Survey web interface. At the top left is the United States Census Bureau logo. The title "American Community Survey" is centered at the top. Below the title is a navigation bar with "Instructions", "FAQs", and "Save and Log Out". On the right side, there is a vertical menu titled "Where You Are" with sub-items: "Basic Info", "Housing Questions", and "Person Info". The main content area shows question 26: "Which of the following best represents how you think of yourself? (Help)". The options are: "Gay or lesbian", "Straight, that is not gay or lesbian", "Bisexual", and "I use a different term" (which is selected). Below the options is a write-in box prompt: "➔ What is that term? (Help)" followed by an empty text input field. At the bottom of the main area are "Previous" and "Next" buttons. The footer contains "Contact Us", "Accessibility", "Privacy", and "Security" links.

Each treatment in the 2024 ACS SOGI Test will have a unique combination of question wording and write-in box display, giving us four treatments: Treatment 1A, Treatment 1B, Treatment 2A, and Treatment 2B. For more information about the treatments and the sample design of the test, see Section 4..

2.4.4 Additional Questionnaire and Instrument Changes

Some ACS questions include language related to gender as only male or female. For the 2024 ACS SOGI Test, we are including gender-neutral language (“they” instead of “he/she”) throughout the ACS questionnaire to accommodate identities outside this binary construct.

We are also updating the Relationship question for all modes and the instruction for the Fertility question on the paper questionnaire. Figure 5 shows the Relationship question for that is currently used in the ACS and the version that will be used in the 2024 ACS SOGI Test. The updated question no longer specifies opposite-sex and same-sex spouse or unmarried partner, as well as removes gendered descriptions from some of the other response categories.

Figure 5. Paper Questionnaire Wording for the Relationship Question: ACS Production Version (Left) and SOGI Version (Right)

<p>2 How is this person related to Person 1? Mark (X) ONE box.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Opposite-sex husband/wife/spouse <input type="checkbox"/> Opposite-sex unmarried partner <input type="checkbox"/> Same-sex husband/wife/spouse <input type="checkbox"/> Same-sex unmarried partner <input type="checkbox"/> Biological son or daughter <input type="checkbox"/> Adopted son or daughter <input type="checkbox"/> Stepson or stepdaughter <input type="checkbox"/> Brother or sister <input type="checkbox"/> Father or mother <input type="checkbox"/> Grandchild <input type="checkbox"/> Parent-in-law <input type="checkbox"/> Son-in-law or daughter-in-law <input type="checkbox"/> Other relative <input type="checkbox"/> Roommate or housemate <input type="checkbox"/> Foster child <input type="checkbox"/> Other nonrelative 	<p>2 How is this person related to Person 1? Mark (X) ONE box.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Spouse <input type="checkbox"/> Unmarried partner <input type="checkbox"/> Biological child <input type="checkbox"/> Adopted child <input type="checkbox"/> Stepchild <input type="checkbox"/> Sibling <input type="checkbox"/> Parent <input type="checkbox"/> Grandchild <input type="checkbox"/> Parent-in-law <input type="checkbox"/> Son-in-law or daughter-in-law <input type="checkbox"/> Other relative <input type="checkbox"/> Roommate or housemate <input type="checkbox"/> Foster child <input type="checkbox"/> Other nonrelative
---	--

Figure 6 shows the instruction for the Fertility question on the paper questionnaire for the current ACS and the version that will be used in the 2024 ACS SOGI Test. The updated instruction specifies that the question should be answered for those whose sex at birth was female.¹¹ This instruction is not needed in the other data collection modes where skip patterns are automated.

Figure 6. Paper Questionnaire Wording for the Instruction to the Fertility Question: ACS Production Version (Left) and SOGI Version (Right)

<p>J Answer question 25 if this person is female and 15 – 50 years old. Otherwise, SKIP to question 26a.</p>	<p>J Answer question 27 if this person's sex at birth was female, and they are 15-50 years old. Otherwise, SKIP to question 28a.</p>
---	---

The updates to the Relationship and Fertility questions are to align the language of current questions to the SOGI questions. We are not making comparisons amongst the treatments in the SOGI sample for these questions (as the same versions will be used in all four treatments). Instead, we will compare against ACS production.

¹¹ Using sex at birth for skip instructions was recommended in the NASEM (2022) report.

2.4.5 Qualitative Questions

The 2024 ACS SOGI Test will include some qualitative follow-up questions in the internet instruments. The questions will appear at the end of the survey, after the respondent answers questions for all household members. The purpose of these questions is to better understand respondents' behavior when interacting with the SOGI questions.

There will be four follow-up questions for Gender Identity and one follow-up question for Sexual Orientation. Respondents may not see all of the questions. The questions they see will depend on their responses to the SOGI questions and the treatment assignment.

Gender Identity Follow-Up Part 1 will be asked to respondents in Treatment 1 (mark only one). Responses from the question will help inform how Treatment 1 responses would vary if the respondent had a mark-all-that-apply option. The question is as follows (asked once of all household members):

Would your answer have changed if you could have selected more than one option (for anyone in your household)? If yes, what options would you have selected?

Gender Identity Follow-Up Part 2 will be asked for any person whose Current Gender answer is nonbinary or uses a different term. Responses from the question will help us better understand the behavior of respondents who do not choose the other three options: male, female, or transgender. The question is as follows (repeated for each eligible household member):

You indicated that (gender identity fill). Could you tell us more about how you came up with your answer? Are there any other answers you thought about choosing but did not?

Gender Identity Follow-Up Part 3 will be asked for any person under 18 whose Sex at Birth and Current Gender answers are different. For the 2024 ACS SOGI Test, the universe for the Current Gender question is persons age 15 or older. Age minimums for gender identity questions are not consistent across surveys, with some asking questions to youth younger than 15 years old (NASEM, 2022). Responses to the qualitative follow-up question will help inform potential future research related to the Current Gender age universe. Due to the late addition during test design, the question is only asked on the CFU instruments, with the question as follows (repeated for each eligible household member):

How old (were you/was (Name)) when they started to feel that their sex assigned at birth and current gender were different?

Gender Identity Follow-Up Part 4 will be asked to respondents in Treatment 2 (mark-all-that-apply) who chose more than one option for Current Gender for any household member. It is a counterpart question to Gender Identity Follow-Up Part 1, and responses from the question will help inform how Treatment 2 responses would vary if the respondent had a mark only one

option. Due to the late addition during test design, the question is only asked on the CFU instruments, with the question as follows (repeated for each eligible household member):

You indicated that (gender identity fill). If you could have selected only one option, what option would you have selected?

Sexual Orientation Follow-Up will be asked for any person whose Sexual Orientation answer is uses a different term. Responses from the question will help us better understand the behavior of respondents who choose the write-in option. The question is as follows (repeated for each eligible household member):

You indicated that (you use/(Name) uses) a different term. Could you tell us more about how you came up with that answer? Are there any other answers you thought about choosing but did not?

Responses from the qualitative questions are for informational purposes only – to better inform how respondents interact with the SOGI questions and future research. The responses from them do not factor into the decision criteria in Section 5..

3. LITERATURE REVIEW

Questions about SOGI on the ACS could help federal agencies improve their understanding of sexual and gender minority populations, as well as help inform civil rights and equal employment enforcement. While there are questions about SOGI on other federal surveys (see Table 1 and Table 2 in Section 3.1), this information is not at the detailed level of geography as in the ACS.

The current ACS (2024) instrument has a sex question and a relationship question. The sex question has two response options: male or female. It does not ask specifically about sex at birth, nor is there a question asking about gender identity.

The relationship question has response options "Opposite-sex husband/wife/spouse", "Opposite-sex unmarried partner", "Same-sex husband/wife/spouse", and "Same-sex unmarried partner", among others.¹² However, the relationship question is only asked of household members other than person 1 and only captures living arrangement information, so no information related to sexual orientation is collected for single-person households, nor is it collected for those not in a romantic relationship with person 1 for multiple-person households.

Neither of these existing questions is sufficient for collecting information about the sexual and gender minority population.

¹² The Relationship question response options will be different than these for the SOGI Test. See Section 2.4 for screenshots of the changes.

3.1 SOGI Questions in Federal Surveys

SOGI questions have been implemented already in some federal surveys (both surveys administered and not administered by the Census Bureau). Some of the Census Bureau-administered surveys that collect SOGI information include the National Health Interview Survey (NHIS), the National Crime Victimization Survey (NCVS), and the Household Pulse Survey (HPS).

National Health Interview Survey (NHIS)

In 2013, NHIS began collecting data regarding sexual orientation. This survey tracks the intersection of sexual orientation, health status, and access to healthcare. Gender identity questions were later added in 2022.

The NHIS is an interviewer-administered survey. While everyone in the household is rostered, only one adult (age 18 years or older) is asked about SOGI. During this roster, age, sex, race, and ethnicity is asked of everyone in the household. Next, one adult (age 18 years or older) and one child (age 17 years or younger) are chosen randomly to complete the rest of the survey. The SOGI questions are asked of the sample adult. Sexual orientation is asked prior to current gender. The sexual orientation question is asked separately for males and females based on the response to the sex question during rostering. The wording of the question is the same, however the response options differ slightly. Both males and females are asked, “Which of the following best represents how you think of yourself?”. The response options are:

- Gay (for males)/Lesbian or gay(females)
- Straight, that is not gay
- Bisexual
- Something else
- Don’t know
- Refused

Early research on the NHIS sexual orientation question showed some evidence of question skipping for older adults, less-educated adults, and adults answering in Spanish. The NHIS is an interviewer-administered survey, so it is unknown whether this was the result of interviewers’ sensitivity toward asking this question, respondents’ refusal to answer, or an alternative reason (Dahlhamer et al., 2014).

The current gender question wording and response options for the NHIS are as follows:

“For this next question you may select more than one answer. Do you currently describe yourself as male, female, transgender, nonbinary, or another gender?”

- Male

- Female
- Transgender
- Nonbinary
- Another Gender

If the response given is “Another Gender”, a write-in response is prompted.

For the 2022 and 2023 NHIS, after the current gender question, respondents were then asked their sex assigned at birth.¹³ A confirmation question was asked if sex assigned at birth did not match their current identity. Cognitive testing of the gender identity series of questions showed the confirmation question was very important for data quality in the two-step series, with several respondents changing their response and preventing false positives (Miller, 2022). However, cognitive testing of the two-step series also indicated some people who are not gender minorities may confuse the sex assigned at birth question with a question on sexuality, thus choosing the sex of their partner (Miller et al., 2021). The 2024 NHIS does not include the sex assigned at birth or confirmation question; a question about sex is still included in the roster section of the survey.

National Crime Victimization Survey (NCVS)

In 2016 the NCVS added questions about SOGI. The results of this survey can show the intersection of being a crime victim with sexual orientation or gender identity.

The NCVS is an interviewer-administered survey. The SOGI questions are asked of those 16 and older and are asked in the 1st, 3rd, 5th, and 7th interviews (out of a total of 7 interviews) (Truman and Morgan, 2022).¹⁴ The NCVS data collection process rosters an entire household during the first visit. This roster includes the name, age, sex, race and ethnicity, marital status, and education level for each household member. After rostering, interviewers screen respondents for potential victimizations. Respondents are asked some socio-demographic questions in addition to the crime screener questions, including questions on SOGI.¹⁵

The NCVS basic screen questionnaire asks about sexual orientation before current gender. The sexual orientation question is the same as the one used in the NHIS.

Research showed the “something else” response was typically selected by English-speaking respondents when they did not want to disclose their identity whereas Spanish-speaking

¹³ The 2022 NHIS also included a test of the gender identity questions with sex at birth asked before gender identity.

¹⁴ Upon implementation in 2016, the SOGI questions were asked to all respondents age 16 or older. From July 2019 through December 2021, these questions were restricted to victims age 16 or older. The administration of the SOGI questions was reverted back to all respondents age 16 or older beginning January 2022.

¹⁵ For more information on NCVS data collection, see the NCVS technical documentation (Bureau of Justice Statistics, 2017).

respondents tended to select “something else” due to confusion about the terminology (Truman et al., 2019).

There are two questions on the NCVS for gender identity: first a question asking sex assigned at birth and then a question on how they currently describe their gender. A confirmation question is asked if the answers to these questions do not match.

The wording and response options for the sex assigned at birth question and the current gender question are as follows:

“What sex were you assigned at birth, on your original birth certificate?”

- Male
- Female
- Refused
- Don’t know

“Do you currently describe yourself as male, female or transgender?”

- Male
- Female
- Transgender
- None of these

Item nonresponse to the SOGI questions have been low, compared to other NCVS questions (Truman and Morgan, 2022).

Household Pulse Survey (HPS)

The HPS is an experimental data product designed to analyze the COVID-19 pandemic’s impact on U.S. households, both socially and economically. Questions regarding SOGI were added in 2021. Some results from this survey show the pandemic may have had a greater impact on the sexual and gender minority populations (Anderson et al., 2021).

The HPS is collected by internet only. Respondents to the HPS must be at least 18 years of age. There is no rostering of the household and no proxy reporting. In the HPS, respondents are asked their sex assigned at birth, followed by a question on how they define their gender, and then they are asked about their sexual orientation. This differs from the previous two surveys which ask about sexual orientation before current gender.

The wording and response options of the sex at birth question and current gender question are below:

“What sex were you assigned at birth, on your original birth certificate? “

- Male
- Female

“Do you currently describe yourself as male, female or transgender?”

- Male
- Female
- Transgender
- None of these

If the answers to these questions are not the same, a confirmation question is asked.

The wording and response options of the sexual orientation question are:

“Which of the following best represents how you think of yourself?”

- Gay or lesbian
- Straight, that is not gay or lesbian
- Bisexual
- Something else, please specify _____
- I don't know

It is important to note that the estimates from HPS are to measure the COVID-19 pandemic's impact on these populations and should not be used for prevalence estimates of the lesbian, gay, bisexual, and transgender (LGBT) population (ITWG, 2021).

Phase 3.10 of the HPS, which collected data from August 23 to October 30, 2023, included a split-sample test on an alternative version of the current gender question. The alternative version had a mark-all-that-apply format, added a nonbinary option, and changed the “none of these” option to “I use a different term” with a write-in field. The results from this test may help inform the analysis of the 2024 ACS SOGI Test.

Other Federal Surveys that Currently Collect SOGI

Some of the federal surveys not administered by the Census Bureau that collect SOGI information include the Behavioral Risk Factor Surveillance System (BFRSS), National Health and Nutrition Examination Survey (NHANES), National Intimate Partner and Sexual Violence Survey (NISVS), National Survey on Drug Use and Health (NSDUH), National Survey of Family Growth (NSFG), Population Assessment of Tobacco and Health (PATH) survey, and Youth Risk Behavior Survey (YRBS).

The current federal surveys that collect data on sexual orientation and gender identity are in Table 1 and Table 2, respectively. Federal surveys given to specific universes (e.g., inmates) or specific cohorts (e.g., the High School Longitudinal Study of 2009 given to 2009 ninth-graders or

the Baccalaureate and Beyond Longitudinal Study: 2016/2020 given to those who earned a bachelor’s degree in 2016) are omitted.

Table 1. Federal Surveys that Collect Sexual Orientation Data

Survey	Year Began Collecting	Data Collection Mode	Age Range
BRFSS [†]	2014	CATI (phone)	18+
HPS	July 2021	Internet	18+
NCVS	July 2016	CAPI (in-person and phone)	16+
NHANES	2001 (question revised 2015)	ACASI [^]	18-59
NHIS	2013 (question revised 2015)	CAPI (in-person and phone)	18+
NISVS	2010	CATI (phone)	18+
NSDUH	2015	ACASI [^]	18+
NSFG	2008	CAPI (in-person), ACASI [^]	15-49
PATH	2013 (question revised 2014)	ACASI [^]	14+
YRBS	2015 (question revised 2021)	Paper	13-18

[†] As part of the optional SOGI Module.

[^] Audio Computer-Assisted Self-Interviewing

Table 2. Federal Surveys that Collect Gender Identity Data

Survey	Year Began Collecting	Data Collection Mode	Age Range
BRFSS [†]	2014	CATI (phone)	18+
HPS	July 2021	Internet	18+
NCVS	July 2016	CAPI (in-person and phone)	16+
NHIS	2022	CAPI (in-person and phone)	18+
PATH	2013	ACASI [^]	14+
YRBS	2023 [*]	Paper	13-18

[†] As part of the optional SOGI Module.

^{*} A gender identity question was on the standard high school questionnaire in 2023, but was available for addition by states to their questionnaires before then.

[^] Audio Computer-Assisted Self-Interviewing

The NHANES, NSDUH, NSFG, and PATH surveys use Audio Computer-Assisted Self-Interviewing (ACASI) for collecting information on SOGI. ACASI allows respondents to enter answers directly into a computer after reading a question directly from a laptop or hearing a question through headphones. It is used to administer questions that may be considered sensitive or personal. See “Current Measures of Sexual Orientation and Gender Identity in Federal Surveys” (IWG, 2016a) for more information on ACASI compared to CAPI for administering SOGI questions.

3.2 Additional SOGI Testing

Joint Program in Survey Methodology (JPSM) Survey Practicum

In 2016, the JPSM Survey Practicum set out to identify ways to optimize response and reduce measurement error to SOGI questions for an online survey.

Results from the analysis showed there were significantly higher rates of item nonresponse in proxy responses compared to self-responses to the question on sexual orientation, but not for the sex at birth and current gender questions. Despite this, the item nonresponse for all SOGI questions was low (under three percent), and they found that proxy respondents were generally able and willing to answer SOGI questions (Ortman et al., 2017).

Cognitive Testing of SOGI for Current Population Survey (CPS)

The Census Bureau conducted cognitive testing of SOGI for the CPS, which collects data regarding the U.S. labor force. The cognitive testing research assessed responding difficulties, sensitivity issues, and the effect of proxy reporting on data quality.

Generally, the participants did not find responding to the SOGI questions difficult for both self-reporting and proxy-reporting, and many judged other CPS questions more difficult to answer (e.g., income, employment, and disability). When there were difficulties indicated, it tended to be due to uncertainty about their own identity or not knowing the sexual orientation or gender identity of another individual in the household (for proxy-reporting).

Participants in this research deemed sexual orientation and disability to be the most sensitive items on the CPS, with LGBT respondents more often finding one or both SOGI questions to be sensitive. Some found their identity personal to talk about, and others were not satisfied with the question wording and response options. While question wording was not a focus of the cognitive testing, respondents provided some valuable feedback on the wording tested.

Despite some indications of sensitivity, all respondents were willing to answer the SOGI questions for themselves and other people in their household (Holzberg et al., 2019). However, participants of this testing were paid volunteers who were willing to interact with the federal government.

4. RESEARCH QUESTIONS AND METHODOLOGY

4.1 Sample Design

The 2024 ACS SOGI Test will consist of a national sample of roughly 480,000 housing unit addresses, excluding Puerto Rico and remote Alaska. This sample will be independent of the ACS production sample. The sample design will largely be 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 PRCS Design and Methodology report (U.S. Census Bureau, 2022b).

There will be three sampling strata, defined by 2020 Census same-sex coupled household rates and 2020 Census urban and rural areas definitions at the tract level.¹⁶ Tracts will be sorted into the three strata according to the following rules:

- **Stratum 1:** The percentage of households featuring same-sex partners in a tract is greater than or equal to one percent.
- **Stratum 2:** The percentage of households featuring same-sex partners in a tract is less than one percent, but the percentage of households in the tract considered rural is greater than or equal to 50 percent.
- **Stratum 3:** The percentage of households featuring same-sex partners in a tract is lower than one percent and the percentage of households in the tract considered rural is less than 50 percent.

The strata were defined to increase the number of sampled households with gender and sexual minorities while ensuring adequate representation of rural households.¹⁷ Addresses within defined sampling strata will be geographically sorted and selected for the sample by systematic sampling.

Addresses in sample for the test will be systematically split evenly across four treatments, Treatment 1A, Treatment 1B, Treatment 2A, and Treatment 2B. For more information regarding the 2024 ACS SOGI Test sample design, see Risley (2023).

4.2 Experimental Design

There will be four treatments for the test: Treatment 1A, Treatment 1B, Treatment 2A, and Treatment 2B. The treatments are defined by their unique combination of question wording for the Current Gender question and write-in box display for the Current Gender and Sexual Orientation questions. As described in Section 2.4, there are two versions of the Current Gender question instructions – Version 1 that specifies “Mark only one” and Version 2 that specifies “Mark one or more”. There are two versions of the write-in display box for both the Current Gender and Sexual Orientation questions (for internet mode only). Version A will display the write-in box when the question is asked. Version B will only show the write-in box if a respondent selects “I use a different term” and clicks the “next” button. The paper questionnaires for Treatments 1A and 1B will be identical, and the paper questionnaires for Treatments 2A and 2B will be identical. Refer to Figure 1 through Figure 4 in Section 2.4 for the different versions.

Table 3 provides the treatment breakdown in terms of question wording and write-in box display.

¹⁶ The data is from 2020 Census table PCT15 and table H2, respectively.

¹⁷ Defining strata only by 2020 Census same-sex coupled household rates would overrepresent urban areas. Since people in rural areas may react differently to the SOGI questions compared to those in urban areas, it is important to have adequate responses from households in rural areas.

Table 3. Breakdown of SOGI Test Treatments

Topic	Treatment 1A	Treatment 1B	Treatment 2A	Treatment 2B
Sex at Birth Question Wording	Version 1	Version 1	Version 1	Version 1
Sex at Birth Write-in Display	NA	NA	NA	NA
Current Gender Question Wording	Version 1	Version 1	Version 2	Version 2
Current Gender Write-in Display	Version A	Version B	Version A	Version B
Sexual Orientation Question Wording	Version 1	Version 1	Version 1	Version 1
Sexual Orientation Write-in Display	Version A	Version B	Version A	Version B

When evaluating differences between the two versions of the Current Gender question, Treatments 1A and 1B will be analyzed together as Treatment 1, and Treatments 2A and 2B will be analyzed together as Treatment 2. Likewise, when evaluating differences between the two write-in box styles in the internet mode, Treatments 1A and 2A will be analyzed together as Treatment A, and Treatments 1B and 2B will be analyzed together as Treatment B.

4.3 Research Questions

Gender Identity Questions

*Benchmarks*¹⁸

- How do the gender identity estimates from each treatment compare to those from:
 - National Health Interview Survey (most recent estimates available)
 - National Crime Victimization Survey (most recent estimates available)
 - Household Pulse Survey (most recent phase available)
 - Estimates from the Williams Institute (Herman et al., 2022)¹⁹
 - 2022 Gallup telephone surveys
 - Statistics Canada 2021 Census of Population
 - The United Kingdom's Census 2021 (England and Wales)
 - New Zealand 2023 Census
 - ACS Production (comparing Sex and Sex at Birth)

¹⁸ See Section 4.4.2 for more information on benchmark sources.

¹⁹ This report uses state-level data from the Centers for Disease Control and Prevention (CDC)'s Behavior Risk Factor Surveillance System (2017-2020) and Youth Risk Behavior Survey (2017 and 2019) to calculate a national estimate of the number of adults (ages 18 and older) and youth (ages 13 to 17) who identify as transgender.

Estimates will be compared for the population as a whole and broken down by age, for benchmark sources where gender identity by age is available.

2. How does the item missing data rate for Sex at Birth nominally compare to the Sex question from ACS production? How does the item missing data rate for the entire Gender Identity series nominally compare to the Sex question from ACS production?

Item Missing Data Rates

For Research Questions 3-5, item missing data rates will be calculated for both missing responses alone and missing responses combined with “protest” and uncodable write-ins.

3. Are the item missing data rates for the Gender Identity series different between Treatment 1 and Treatment 2 (for each question and the series as a whole)?
4. Are the item missing data rates for the Gender Identity series different between Treatment 1 and Treatment 2 within mode (mail and internet)?
5. For internet responses, are the item missing data rates for Current Gender different between the two internet write-in box styles (Treatment A vs Treatment B)? Is there a difference in the rate of responses where the “use a different term” option is selected but a write-in is not provided?
6. Are the item missing data rates for the Gender Identity series different between Treatment 1 and Treatment 2 by demographic characteristics (household respondent status, relationship type, age category, race, Hispanic origin, urban/rural status, and English proficiency)?
7. For each treatment, how do the item missing data rates for the Gender Identity series compare to those for other items in the Basic Person section, sensitive items, or items with proxy-reporting differences, overall and by household respondent status?
 - a) Basic Person items: Relationship, Age, Hispanic Origin, Race
 - b) Sensitive items: Place of Birth, Citizenship, Disability, Total Income
 - c) Items with proxy-reporting differences: Educational Attainment, Health Insurance Coverage, Time to Work

Response Distributions

8. Are the Gender Identity distributions different between Treatment 1 and Treatment 2 (for each question and the series as a whole)?

- a) We will look at distributions from each part of the two-step question (Sex at Birth and Current Gender) as well as both steps taken together (Gender Identity).
 - b) We will compare distributions not accounting for and accounting for write-in responses.
9. Are the Gender Identity distributions different between Treatment 1 and Treatment 2 within mode (mail and internet)?
 10. Are the Gender Identity distributions different between Treatment 1 and Treatment 2 by demographic characteristics (household respondent status, relationship type, age category, race, Hispanic origin, urban/rural status, and English proficiency)?
 11. Within each treatment, are the Gender Identity distributions different among demographic characteristic categories (e.g., different among race groups)?
 12. How do the proportions of write-ins for Gender Identity compare between Treatment 1 and Treatment 2, overall and by mode?
 13. Within each treatment, what are the proportions of write-ins for Gender Identity by demographic characteristics (age category, race, Hispanic origin, urban/rural status)?
 14. How do the proportions of “protest” write-ins for Gender Identity compare between Treatment 1 and Treatment 2, overall and by mode?
 15. How do the distributions of coded categories of write-ins compare between Treatment 1 and Treatment 2? How do they compare to those from previous studies (e.g., Stats Canada list and cognitive testing)?
 16. For internet responses, how do the proportions of write-ins for Current Gender compare between the two internet write-in box styles (Treatment A vs Treatment B)? How do the proportions of “protest” write-ins or uncodable write-ins compare? How does the content of write-ins qualitatively compare?
 17. For Treatment 2, in all modes, what are the proportions of multiple marks, overall and by mode? What combinations of multiple marks were selected? For Treatment 1 in the mail mode, what are the proportions of multiple marks? What combinations of multiple marks were selected?

18. For mail responses and by treatment, how often is the Current Gender question answered for people under age 15? How does this nominally compare to the rate at which question 21 was answered for people under age 15?²⁰

*Response Reliability*²¹

19. Are the measures of response reliability for the Gender Identity series different between Treatment 1 and Treatment 2, overall and by mode of original interview (mail and internet)?

20. What are the response reliability measures for the Gender Identity series when the same household respondent responds to both interviews? How do those measures compare to those when a different respondent responds in the CFU reinterview? If sample allows, how do the response reliability measures compare by household composition (i.e., simple vs. complex households)?

21. For households in which the same household member completed both interviews, how do the response reliability measures compare between self-responses and proxy responses?

22. For each treatment, how do the measures of response reliability for Gender Identity nominally compare to those for other items in the Basic Person section, sensitive items, or items with proxy-reporting differences? How do the measures compare by proxy-reporting status (i.e., compare across items for self-responses and compare for proxy responses)?

a) Basic Person items on CFU reinterview: Relationship, Age, Hispanic Origin, Race

b) Detailed Person items on CFU reinterview: Place of Birth, Educational Attainment, Health Insurance Coverage, Disability, Marital Status/History, Fertility, Wages

23. *For households in which a different household member responded in CFU:* How do the measures of response reliability compare between persons who had a self-response in one interview and proxy response in the other, and those who had two different proxy responses? For persons who had a self-response and a proxy response, what proportion of proxy responses did not match the self-response?

²⁰ Question 21 asks about difficulty doing errands alone. It is supposed to be answered only for persons age 15 or over, and it is the first question out of the remaining detailed person questions that exclude persons under age 15.

²¹ See Section 4.4.5 for definitions of the response reliability measures.

Additional Metrics

24. Using internet paradata, how do the following behaviors for the Gender Identity series compare between treatments? How do the rates of these behaviors compare to those for other sensitive or complex questions from Research Question 9?
 - a) Breakoff rates?
 - b) Help screen access rates?
 - c) Backing (going back to different screens)?
 - d) Answer changes?
 - e) Time on screen?
25. What proportion of respondents would have received the Gender Identity confirmation question if only asked for respondents who provided different answers to Sex at Birth and Current Gender? How does this rate compare between treatments?
26. For respondents whose responses to the two Gender Identity questions matched, what proportion of them changed their response in reaction to the confirmation question (i.e., the “false negative” rate)? How does this compare to the proportion for those whose initial answers to the two Gender Identity questions did not match (i.e., the “false positive” rate)?
27. *When coded data from the qualitative questions are available:* For households in Treatment 1 (mark only one), what proportion said their answer would have changed for someone in their household if they could have selected more than one option? Which response options did they select, and which other options would they have selected if possible?
28. *When coded data from the qualitative questions are available:* For households in Treatment 2 (mark all that apply) who selected more than one response option for Current Gender, which response option would respondents have selected if they could only select one?
29. *When coded data from the qualitative questions are available:* In responses to the qualitative follow-up question, what factors did respondents tend to mention when describing how they selected “nonbinary” or “use a different term” as their response? Which of the other response options did they mention considering most often? How did these answers differ between self- and proxy responses? What factors were mentioned for youth ages 15 to 17?
30. *When coded data from the qualitative questions are available:* In responses to the qualitative follow-up question on how old a person was when they started to

feel/express that their sex assigned at birth and current gender were different, what was the distribution of ages given by respondents?

31. *Potential analysis using modeling*: How does the type of household relationship (e.g., close relative vs. extended relative vs. roommate) affect the quality of proxy reporting? How does it affect item missingness?

Sexual Orientation Question

*Benchmarks*²²

1. How do the sexual orientation estimates compare to those from:
 - National Health Interview Survey (most recent estimates available)
 - National Crime Victimization Survey (most recent estimates available)
 - Household Pulse Survey (most recent phase available)
 - 2022 Gallup telephone surveys
 - The United Kingdom’s Census 2021 (England and Wales)
 - New Zealand 2023 Census

Estimates will be compared for the population as a whole and broken down by age, for benchmark sources where sexual orientation by age is available.

Item Missing Data Rates

For Research Questions 2-3, item missing data rates will be calculated for both missing responses alone and missing responses combined with “protest” and uncodable write-ins.

2. What are the item missing data rates for Sexual Orientation, overall and by mode?
 - a. Are the rates significantly different between Treatment 1 and Treatment 2? If so, the remaining metrics will account for treatment differences.
3. For internet responses, are the item missing data rates for Sexual Orientation different between the two internet write-in box styles (Treatment A vs Treatment B)? Is there a difference in the rate of responses where the “different term” option is selected but a write-in is not provided?
4. What are the item missing data rates for Sexual Orientation by demographic characteristics (household respondent status, relationship type, age category, race, Hispanic origin, urban/rural status, and English proficiency)?

²² See Section 4.4.2 for more information on benchmark sources.

5. How do the item missing data rates for Sexual Orientation compare to those for sensitive items or items with proxy-reporting differences?
 - a) Sensitive items: Place of Birth, Citizenship, Disability, Total Income
 - b) Items with proxy reporting differences: Educational Attainment, Health Insurance Coverage, Time to Work
 - c) Gender Identity items: Sex at Birth, Current Gender, Gender Identity series

Response Distributions

6. What are the response distributions for Sexual Orientation, overall and by mode?
 - a. We will check if the distributions are significantly different between Treatment 1 and Treatment 2. If so, the remaining metrics will account for treatment differences.
7. What are the response distributions for Sexual Orientation by demographic characteristics (household respondent status, relationship type, age category, race, Hispanic origin, urban/rural status, and English proficiency)? Are the distributions different within characteristic categories (e.g., different between race groups)?
8. For internet responses, are the Sexual Orientation distributions different between the two internet write-in box styles (Treatment A vs Treatment B), overall and by demographic characteristics (household respondent status, age category, race, Hispanic origin, urban/rural status, and English proficiency)?
9. What are the proportions of write-ins for Sexual Orientation, overall and by mode?
10. What are the proportions of write-ins for Sexual Orientation by demographic characteristics (age category, race, Hispanic origin, urban/rural status)?
11. What are the proportions of “protest” write-ins for Sexual Orientation, overall and by mode?
12. How does the content of the write-ins for Sexual Orientation qualitatively compare to those from previous studies (e.g., cognitive testing)?
13. For internet responses, how do the proportions of write-ins for Sexual Orientation compare between the two internet write-in box styles (Treatment A vs Treatment B)? How do the proportions of “protest” write-ins or uncodable write-ins compare? How does the content of write-ins qualitatively compare?

14. For mail responses, how often is the Sexual Orientation question answered for people under age 15? How does this compare to the rate at which question 21 was answered for people under age 15?²³

*Response Reliability*²⁴

15. What are the measures of response reliability for Sexual Orientation, overall and by mode of original interview (mail and internet)?
16. What are the response reliability measures for Sexual Orientation when the same household respondent responds to both interviews? How do those measures compare to those when a different respondent responds in the CFU reinterview?
17. For households in which the same household member completed both interviews, how do the response reliability measures compare between self-responses and proxy responses?
18. How do the measures of response reliability for Sexual Orientation compare to those for sensitive items or items with proxy-reporting differences? How do the measures compare by proxy-reporting status (i.e., compare across items for self-responses and compare for proxy responses)?
 - a. Detailed Person items on CFU reinterview: Place of Birth, Educational Attainment, Health Insurance Coverage, Disability, Marital Status/History, Fertility, Wages
19. *For households in which a different household member responded in CFU:* How do the measures of response reliability compare between persons who had a self-response in one interview and proxy response in the other, and those who had two different proxy responses? For persons who had a self-response and a proxy response, what proportion of proxy responses did not match the self-response?

Additional Metrics

20. Using internet paradata, how do the following behaviors for the Sexual Orientation question compare between the two internet write-in box styles? How do these rates compare to those for other complex or sensitive questions in Research Question 5?
 - a) Breakoff rates?
 - b) Help screen access rates?

²³ Questions 21 asks about difficulty doing errands alone. It is supposed to be answered only for persons age 15 or over, and it is the first question out of the remaining detailed person questions that exclude persons under age 15.

²⁴ See Section 4.4.5 for definitions of the response reliability measures.

- c) Backing (going back to different screens)?
- d) Answer changes?
- e) Time on screen?

21. *When coded data from the qualitative question are available:* In responses to the qualitative follow-up question, what factors did respondents tend to mention when describing how they selected “use a different term” as their response? Which of the other response options did they mention considering most often? How did these answers differ between self- and proxy responses?
22. *Potential analysis using modeling:* How does the type of household relationship (e.g., close relative vs. extended relative vs. roommate) affect the quality of proxy reporting? How does it affect item missingness?

Other ACS Questions

Item Missing Data Rates

1. What are the item missing data rates for the Relationship question, overall and by mode (internet and mail)? How do they compare to historical ACS missingness rates?
2. For mail responses, what are the item missing data rates for the Fertility and Grandparents as Caregivers questions? How do these compare to historical ACS missingness rates?
3. What are the item missing data rates for the Fertility question, overall and within treatment? What are the rates by mode (internet and mail)? How do they compare to historical ACS missingness?
4. In the mail mode, what proportion of people ages 15 to 50 whose sex assigned at birth was female did not answer the Fertility question? What proportion of people who answered the Fertility question are not age 15 to 50 and assigned female at birth?

Response Distributions

5. What are the response distributions for spouse and unmarried partners, overall and by mode (internet and mail)? What is the overall response distribution for Relationship? How do they compare to historical ACS estimates? How do they compare between treatments?
6. What are the response distributions of different-gender couple, same-gender couple, transgender couple, and nonbinary couple? How do those distributions compare between treatments?

7. What are the response distributions for the Fertility question, overall and within treatment? What are the distributions by mode (internet and mail)? How do they compare to historical ACS estimates?

Additional Metrics

8. For the internet responses, what is the average total time spent in the instrument? How does this compare between treatments? How does this compare to ACS production?

4.4 Analysis Metrics

For the 2024 ACS SOGI Test, we will not make typical ACS production edits because the primary concern of the test is examining the unaltered responses to the SOGI questions provided directly by respondents. For this reason, responses will not be imputed either. A few edits may be made to non-SOGI data, such as calculating a person’s age based on their date of birth, but such edits will be minimal.

Comparisons between treatments will be conducted using first-order adjustment Rao-Scott chi-square tests (Rao & Scott, 1987) or two-tailed t-tests at the $\alpha=0.1$ level of significance, as specified in the following sections. In statistical tests involving multiple comparisons, we will control for the overall Type I error rate by adjusting the resulting p-values using the Hochberg method (Hochberg, 1988).

4.4.1 Unit-Level Analysis

As part of our analysis for the 2024 ACS SOGI Test, we will test for unit-level differences between treatments for the original interview and the CFU reinterview.

We will calculate and compare the unit response rate overall and by mode for each treatment. The unit response rate is generally defined as the weighted proportion of sample addresses eligible to respond that provided a complete or sufficient partial response.²⁵

We will calculate how the unit response rates were distributed by mode as follows for each treatment:

$$\text{Self-Response Portion of the Final Response Rate} = \frac{\text{Number of sample addresses that provided a response by mail or internet}}{\text{Total number of sample addresses that were eligible to respond to the survey}} \times 100$$

²⁵ In general, a sufficient partial response is one that has at least minimal information, which indicates an attempt to respond. The specific definition of a sufficient partial response is sensitive and for Census Bureau internal use only.

$$\text{Mail Portion of the Final Response Rate} = \frac{\text{Number of sample addresses that provided a response by mail}}{\text{Total number of sample addresses that were eligible to respond to the survey}} \times 100$$

$$\text{Internet Portion of the Final Response Rate} = \frac{\text{Number of sample addresses that provided a response by internet}}{\text{Total number of sample addresses that were eligible to respond to the survey}} \times 100$$

The universe for the unit response rates consists of all addresses in the initial sample that are eligible to respond to the survey. Any nonresponding addresses that were sampled out of CAPI will not be included in any of the response rate calculations. Later, when the CAPI operation is completed, we will calculate the CAPI portion of the final response rate as well as the final response rate with CAPI response included.

We will also calculate response rates for the CFU reinterview as the weighted proportion of sample addresses eligible to respond to the CFU reinterview that provided a response. We will calculate CFU response rates overall and by mode of original interview for each treatment. Additionally, we will calculate CFU response rates by CFU strata and CFU treatment.

We will test for differences in response rates using two-tailed t-tests. Response rate comparisons will correspond to what comparisons are being made in the research questions (e.g., Treatment 1 vs. Treatment 2 and Treatment A vs. Treatment B).

Additionally, we will calculate distributions of respondents' socioeconomic and demographic characteristics. We will calculate distributions by treatment as the proportion of valid responses in a category to all valid responses. We will compare the distributions between treatments of the following demographic and household-level characteristics:

- Hispanic origin
- Race
- Age
- Household respondent status
- Limited English proficiency status
- Urban and rural status
- Household size

We will calculate distributions for the original interview by mode and for the CFU reinterview by mode of original interview. We will use Rao-Scott chi-square tests of independence (Rao & Scott, 1987) to test for differences in the distributions of each demographic characteristic.

4.4.2 Benchmarks

To roughly gauge the accuracy of the responses to the SOGI questions, we will nominally compare select estimates from the test to similar estimates from external reliable sources (i.e., a benchmark).

Differences between the benchmark data sources and the 2024 ACS SOGI Test data, such as the sample design (sampling technique and universe), question wording, and type of survey, will be noted in each analysis. We will compare the latest available NCHS, NCVS, and HPS estimates to the SOGI Test data as a comparison of this test to other federal surveys.²⁶ We will also look at benchmarks from other U.S. sources as well as censuses from other countries. Those benchmark sources are briefly described below.

Gallup telephone surveys

Gallup produces an annual national estimate of sexual and gender minority status in adults based on aggregated polling data from their telephone surveys. Respondents are asked the following:

“Which of the following do you consider yourself to be? You can select as many as apply.”

- Straight or heterosexual
- Lesbian
- Gay
- Bisexual
- Transgender
- Something else

In 2022, respondents who selected “something else” were asked to give their preferred identity; most said either pansexual, asexual, or queer (Jones, 2023).

Williams Institute report

The Williams Institute, as part of the University of California Los Angeles School of Law, conducts independent research on SOGI law and public policy. Herman et al. (2022) estimated the number of adults (ages 18 and older) and youth (ages 13 to 17) who identify as transgender using state-level data from the Centers for Disease Control and Prevention (CDC)'s Behavior Risk Factor Surveillance System (BRFSS) (2017-2020) and Youth Risk Behavior Survey (YRBS) (2017 and 2019). Both surveys have optional gender identity questions that can be included on the state-specific questionnaires. The report was able to use BRFSS information from 41 states and YRBS information from 15 states, calculating estimates for the remaining states through multilevel models.

²⁶ Although the HPS is not intended to provide SOGI prevalence estimates, it will still be a worthwhile comparison for showing the possible range of estimates.

The Herman et al. (2022) report provides national estimates of transgender women adults, transgender men adults, transgender gender-nonconforming adults, and transgender youth (ages 13 to 17). These estimates are also given by age category, race/ethnicity, and state.

Statistics Canada 2021 Census of Population

Canada's 2021 Census of Population included a question on gender and distinguished it from sex, for the first time compared to previous censuses. Sex at birth, followed by gender, were asked about for all household members, including children. The questions were the following:

“What was this person’s sex at birth? Sex refers to sex assigned at birth.”

- Male
- Female

“What is this person’s gender? Refers to current gender, which may be different from sex assigned at birth and may be different from what is indicated on legal documents.”

- Male
- Female
- Or please specify this person’s gender (with write-in box)

2021 Census data releases include estimates of cisgender men and women, transgender men and women, and nonbinary persons. The data releases also include a two-category gender variable, men+ and women+, which distributes nonbinary persons to the other two categories.²⁷

United Kingdom’s Census 2021 (England and Wales)

The census of England and Wales included for the first time, in 2021, questions about gender identity and sexual orientation. The questions were voluntary and asked of people age 16 years and older. The gender identity question was the following:

“Is the gender you identify with the same as your sex registered at birth?”

- Yes
- No (with write-in box)

Using responses from a sex question, estimates of transgender man, transgender woman, nonbinary, and other gender identity are available from Census 2021. The sexual orientation question was the following:

²⁷ For more information on Statistics Canada 2021 Census of Population sex at birth and gender questions, see the [Age, Sex at Birth and Gender Reference Guide, Census of Population, 2021](#).

“Which of the following best describes your sexual orientation?”

- Straight/heterosexual
- Gay or lesbian
- Bisexual
- Other sexual orientation (with write-in)

Estimates are available from Census 2021 for the first three categories along with pansexual, asexual, queer, and all other sexual orientations using the write-in responses.²⁸

New Zealand 2023 Census

The New Zealand Census of Populations and Dwellings included for the first time, in 2023, questions about gender, sexual identity, and variation of sex characteristics (i.e., intersex). The gender question series was the following:

“What is your gender?”

- Male
- Female
- Another gender (with write-in box)

“What was your sex at birth?”

- Male
- Female

The sexual orientation question was, “Which of the following best describes how you think of yourself?” It had the following response options:

- Heterosexual/straight
- Gay or lesbian
- Bisexual
- Another identity (with write-in box)
- Prefer not to say

Results from the 2023 Census will be available starting late May 2024.²⁹

²⁸ For more information on the SOGI questions from Census 2021 of England and Wales, see [Sexual orientation and gender identity variables Census 2021](#).

²⁹ For more information on the SOGI questions on the New Zealand 2023 Census, see [2023 Census fact sheet: Rainbow communities](#).

4.4.3 Item Missing Data Rates

Respondents leave items blank for a variety of reasons including not understanding the question (clarity), unwillingness to answer a question as presented (sensitivity), and lack of knowledge of the data needed to answer the question. The item missing data rate (for a given item) is the proportion of eligible persons for which a required response is missing.

For each item, it is important to carefully define both the universe of eligible units and the criteria that determine whether a response to that item is missing or not missing. Usually the definition of “missing” will include paper and internet questionnaires where no answer was provided. For items with write-in boxes, we will calculate two versions of the item missing data rates: one version counts “protest” and uncodable write-in responses as missing, and the other version counts them not as missing.

We will compare the item missing data rates via two-tailed t-tests.

4.4.4 Response Distributions

Comparing the response distributions between question versions will allow us to assess whether the question change affects the resulting estimates. We will calculate response distributions as the proportion of valid responses in a category to all valid responses. Most of the response distributions will account for the write-in responses. When comparing response distributions between treatments or demographic categories, we may combine certain categories of responses that very few respondents select.

We will make comparisons using a Rao-Scott chi-square test (first-order adjustment) that checks for a significant difference between two sample distributions (Rao & Scott, 1987).

4.4.5 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 2024 ACS SOGI Test, we will measure response error using response reliability.

A survey question has good response reliability if respondents tend to answer the question consistently. We will measure 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. In the SOGI Test, we are also interested in assessing the reliability of proxy responses. We will do this by calculating simple response variance when the questions are re-asked of a different household member. In both cases, we will use the following measures:

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

The first two measures, GDR and IOI, will be calculated for individual response categories. The L-fold index of inconsistency will be calculated for questions that have 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.

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 will be weighted to make them more representative of the population.

We will calculate 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 will look at the differences in the GDR and IOI for each response category. We will test the significance of these differences, using two-tailed t-tests.

So far, we have only discussed response reliability with respect to single response categories. If a question has three or more response categories (or “comparison categories” in cases where it is necessary to collapse some response categories for comparison), we can also measure the overall response reliability of a question using the L-fold index of inconsistency, IOI_L . It is possible to look at the difference in IOI_L between treatments and test for significance as with the single category measures.

Suppose a question has L response categories. Let X_{ij} be the weighted count of sample units (households or persons) for which we have CFU responses in category i and original interview responses in category j . Here, both i and j range from 1 to L. Table 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 will check 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 will use the following adjustment of the IOI, developed by Flanagan (2001):

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

4.5 Weighting

All estimates from the 2024 ACS SOGI Test will be weighted. The weights will be a function of the initial probability of selection (base weight). A CFU nonresponse adjustment will be used to create weights for the CFU analysis.³⁰

We will estimate the variances of the point estimates and differences using the Successive Differences Replication (SDR) method with replicate weights – the standard method used in the ACS (see U.S. Census Bureau, 2022b). We will calculate the variance for each rate and for the difference between rates using the formula below:

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

where:

$X_r = \hat{c}$ the estimate calculated using the r^{th} replicate

$X_0 = \hat{c}$ the estimate calculated using the full sample

The standard error of the estimate (X_0) is the square root of the variance.

³⁰ The weight creation process for the 2024 ACS SOGI Test 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 SOGI Test weighting procedure, see Risley (2023).

5. DECISION CRITERIA

When evaluating general data quality from the Gender Identity series, including the quality of proxy responses, the most important results of this analysis are, in order of priority:

Table 6. Decision Criteria for Evaluating Gender Identity

Priority	Research Question	Decision Criteria
1	1	The percent male and percent female for Sex at Birth being nominally close to the percentages from non-edited ACS production is preferable.
	2	The item missing data rate for Sex at Birth being lower or comparable to the Sex question from ACS production is preferable.
2	3, 4, 6, 7	Low item missingness for the entire Gender Identity series is preferable. Specifically, the item missingness being lower or comparable to ACS questions that are sensitive or have proxy-reporting differences is preferable.
3	19, 22	High response reliability is preferable. Specifically, the response reliability nominally being higher or comparable to ACS questions that are sensitive or have proxy-reporting differences is preferable.
4	1	The estimates (sex at birth and gender) being nominally close to benchmarks is preferable.
5	20-23	The difference in response reliability between proxy responses and self-responses being nominally smaller or comparable to ACS questions that are sensitive or have proxy-reporting differences is preferable.
6	24	The behavior of respondents on the internet instrument answering the Gender Identity questions (e.g., breakoff rates and help screen access rates) being comparable to how they answer other sensitive or complex questions is preferable. Specifically, the metrics for Gender Identity being smaller or comparable to those for other questions is preferable.

When evaluating general data quality from the Sexual Orientation question, including the quality of proxy responses, the most important results of this analysis are, in order of priority:

Table 7. Decision Criteria for Evaluating Sexual Orientation

Priority	Research Question	Decision Criteria
1	2, 4, 5	Low item missingness is preferable. Specifically, the item missingness being lower or comparable to ACS

		questions that are sensitive or have proxy-reporting differences is preferable.
2	15, 18	High response reliability is preferable. Specifically, the response reliability being nominally higher or comparable to ACS questions that are sensitive or have proxy-reporting differences is preferable.
3	16-19	The difference in response reliability between proxy responses and self-responses being nominally smaller or comparable to ACS questions that are sensitive or have proxy-reporting differences is preferable.
4	1	The estimates being nominally close to benchmarks is preferable.
5	20	The behavior of respondents on the internet instrument answering the Sexual Orientation question (e.g., breakoff rates and help screen access rates) being comparable to how they answer other sensitive or complex questions is preferable. Specifically, the metrics for Sexual Orientation being smaller or comparable to those for other questions is preferable.

When evaluating mark all vs. mark one in the Current Gender question, the most important results of this analysis are, in order of priority:

Table 8. Decision Criteria for Mark All vs Mark One for Gender Identity

Priority	Research Question	Decision Criteria
1	3, 4, 6	The treatment with the lowest item missingness is preferable.
2	14-15	The treatment with the lowest rate of “protest” or uncodable write-ins is preferable.
3	1	The treatment with estimates nominally closest to benchmarks is preferable.
4	17	A high rate of multiple marks, relative to the rate of gender minorities, may indicate a need for a mark all format.
5	19	The treatment with the highest response reliability is preferable. However, we expect Treatment 2 to have lower response reliability due to the more complex mark all format.
6	24	The treatment with the lowest respondent burden in the internet instrument (e.g., breakoff rates and help screen access rates) is preferable.

When evaluating the confirmation question in the Gender Identity series, the most important results of this analysis are, in order of priority:

Table 9. Decision Criteria for Confirmation Question for Gender Identity

Priorty	Research Question	Decision Criteria
1	26	Respondents changing their responses to the Gender Identity questions would indicate a need for a confirmation question. Specifically, a substantial false negative rate (the proportion of respondents whose initial answers do not identify them as a gender minority, but whose answers after the confirmation question do) would indicate a need for a confirmation question for everyone, regardless of initial answer choice.
2	25	Potential improvements in data quality from priority 1 will be weighed against potential increases in burden for all respondents receiving a confirmation question compared to only household members whose Sex at Birth and Current Gender do not match.

When evaluating the internet write-in box styles in the SOGI questions, the most important results of this analysis are, in order of priority:

Table 10. Decision Criteria for Internet Write-in Box Style

Priorty	Research Question	Decision Criteria
1	GI 5, SO 3	The write-in box style with the lowest item missingness is preferable.
2	GI 16, SO 13	The write-in box style with the lowest rate of “protest” or uncodable write-ins is preferable.

When evaluating changes to existing ACS questions (e.g., Relationship), the most important results of this analysis are, in order of priority:

Table 11. Decision Criteria for Changes to Existing ACS Questions

Priorty	Research Question	Decision Criteria
1	1-4	The item missing data rates being comparable to historical ACS missingness rates is preferable.
2	5-7	The response distributions being comparable to historical ACS estimates is preferable. However, some differences are expected due to sample differences between the SOGI Test and ACS production.

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

6. ASSUMPTIONS AND LIMITATIONS

6.1 Assumptions

1. The 2024 ACS SOGI Test sample is representative of the entire ACS sample frame, with respect to response rates.
2. Each treatment sample (1/4 of the full SOGI Test sample) is representative of the entire ACS sample frame.
3. The one-week delay in sending the first mailing for the August and September SOGI panels, compared to the first mailings for the August and September ACS production data collection panels, will not affect response rates, overall and between treatments. Additionally, the delay will not result in a change in respondent characteristics by mode.
4. We assume that the frequency of real changes in answers due to a change in life circumstances between the original interview and CFU will be similar between treatments.

6.2 Limitations

1. Group Quarters are not included in the sample for the test.
2. Sample housing unit addresses from remote Alaska and Puerto Rico are not included in the sample for the test.
3. Interviews will be conducted in English and Spanish only. Respondents who need language assistance in another language will not be able to participate in the 2024 ACS SOGI Test. TQA responses are not included in the analysis because survey responses completed via the TQA operation were only conducted using the ACS production data collection instrument.
4. The CFU reinterview will be conducted by phone for cases that responded by mail in the original interview (different data collection modes). As a result, the data quality measures derived from the reinterview may include some bias for the mail mode due to the differences in mode of data collection. Cases that responded by internet in the original interview will respond via internet again in the CFU reinterview, and so there will not be a potential bias due to differences in mode of data collection. However, there will be mode differences between CFU respondents (with some having internet and some having phone CFU reinterviews) that could impact analysis of the CFU metrics.

5. Due to variations on the internet and phone operations, there will be some differences in the length of time between the original interviews and CFU reinterviews between the internet and CATI CFU modes. These differences could affect respondent's answers to survey questions due to changes in life circumstances or recollection of the original interview answers. However, the timing differences between modes is so minor that we expect any impacts to be minimal.
6. This timing attempts to balance two competing needs: to minimize the possibility of real changes in answers due to a change in life circumstances between the two interviews and to minimize the possibility that the respondent repeats their previous answer based on their recollection of the original interview response rather than considering the most appropriate answer.
7. We have not had an internet CFU operation before. Historically, the CFU operation has only been conducted by phone. Though we think the internet mode will gain more response than having only phone, there is a risk of our assumption not being true. Therefore, the quality of data available to answer research questions related to the CFU reinterview may be lower if there is not sufficient response to that operation.
8. Some research questions relate to which household member was the household respondent in the original and CFU interviews. We assume that the first person listed on the household roster was the respondent; however, in self-response modes, this assumption may not always be true. Therefore, there is a limitation in the accuracy of our analysis based on household respondent status.

7. POTENTIAL CHANGES TO ACS

This test could result in the addition of SOGI questions to the ACS. We expect adding the SOGI questions to the ACS would help us capture populations we previously missed and allow us to calculate estimates for these populations. Incorporating new questions into the ACS requires redesigning survey instruments, revising our data processing specifications, and updating interviewer training materials.

8. REFERENCES

- Anderson, L., File, T., Marshall, J., McElrath, K., & Scherer, Z. (2021). "New Household Pulse Survey Data Reveals Differences between LGBT and Non-LGBT Respondents During COVID-19 Pandemic." U.S. Census Bureau.
<https://www.census.gov/library/stories/2021/11/census-bureau-survey-explores-sexual-orientation-and-gender-identity.html>
- Bureau of Justice Statistics (2017). *National Crime Victimization Survey, 2016: Technical documentation*. NCJ 251442. Retrieved December 21, 2023, from <https://bjs.ojp.gov/redirect-legacy/content/pub/pdf/ncvstd16.pdf>
- Contard, L. (2023). *Requirements for the Content Follow-Up Reinterview in the 2024 American Community Survey Sexual Orientation and Gender Identity Test*. Forthcoming.
- Dahlhamer, J.M., Galinsky, A.M., Joestl, S.S., & Ward, B.W. (2014). Sexual orientation in the 2013 National Health Interview Survey: A quality assessment. *Vital and Health Statistics*, 2(169), 1-24. https://www.cdc.gov/nchs/data/series/sr_02/sr02_169.pdf
- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys. (IWG, 2016a). *Current Measures of Sexual Orientation and Gender Identity in Federal Surveys*. Retrieved November 7, 2023, from https://nces.ed.gov/FCSM/pdf/current_measures_20160812.pdf
- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys. (IWG, 2016b). *Evaluations of Sexual Orientation and Gender Identity Survey Measures: What Have We Learned?* Retrieved November 7, 2023, from https://nces.ed.gov/FCSM/pdf/Evaluations_of_SOGI_Questions_20160923.pdf
- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys. (IWG, 2016c). *Toward a Research Agenda for Measuring Sexual Orientation and Gender Identity in Federal Surveys: Findings, Recommendations, and Next Steps*. Retrieved November 7, 2023, from https://nces.ed.gov/FCSM/pdf/SOGI_Research_Agenda_Final_Report_20161020.pdf
- Herman, J.L., Flores, A.R., O'Neill, K.K. (2022). *How Many Adults and Youth Identify as Transgender in the United States?* The Williams Institute, UCLA School of Law.
<https://williamsinstitute.law.ucla.edu/wp-content/uploads/Trans-Pop-Update-Jun-2022.pdf>
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4), 800-802. <https://doi.org/10.2307/2336325>

Holzberg, J., Ellis, R., Kaplan, R., Virgile, M., & Edgar, J. (2019). Can they and will they? Exploring proxy response of sexual orientation and gender identity in the Current Population Survey. *Journal of Official Statistics*, 35 (4), 885-911. <https://doi.org/10.2478/jos-2019-0037>

Hughes, T., Lordi, N., Park, R., Shrivastava, P., Fu, J., Steinmertz, V., Aguilar, D., & Wan, W. (2023). *CHIS sexual orientation and gender identity 2022 cognitive pretest report: Findings from Waves I and II*. UCLA Center for Health Policy Research. Retrieved January 5, 2024, from <https://healthpolicy.ucla.edu/sites/default/files/2023-05/chis-2022-sogi-pretest-report-waves-i-and-ii.pdf>

Interagency Technical Working Group on Sexual Orientation and Gender Identity Items in the Household Pulse Survey. (ITWG, 2021). *Report and recommendations*. Office of Management and Budget.

Miller, K., Willson, S., & Ryan, V. (2021). *An initial cognitive evaluation of a 2-step gender identity measure*. National Center for Health Statistics - CCQDER. <https://wwwn.cdc.gov/QBank/Report.aspx?1219>

Miller, K. (2022, August 2). *SOGI Question Design and Evaluation Collaborating Center for Question Design and Evaluation Research (CCQDER) [Webinar]*. Centers for Disease Control and Prevention. <https://www.cdc.gov/nchs/video/dae-sexual-orientation-gender-identity/DAE-Sexual-Orientation-and-Gender-Identity-2022-Low-Res.mp4>

National Academies of Sciences, Engineering, and Medicine. (2022). *Measuring sex, gender identity, and sexual orientation*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26424>

Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Retrieved December 21, 2023, from https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Office of Management and Budget. (2023). *Recommendations on the best practices for the collection of sexual orientation and gender identity data on federal statistical surveys*. Retrieved January 4, 2024, from <https://www.whitehouse.gov/wp-content/uploads/2023/01/SOGI-Best-Practices.pdf>

Ortman, J.M., Bates, N.A., Brown, A., & Sawyer, R.C. (2017). *Optimizing Self and Proxy Response to Survey Questions on Sexual Orientation and Gender Identity*. Presented at the Annual Meeting of the Population Association of America. Retrieved November 7, 2023 from <https://paa.confex.com/paa/2017/mediafile/ExtendedAbstract/Paper13962/Ortman%20Bates%20Brown%20Sawyer%20PAA%202017.pdf>

- Rao, J., & Scott, A. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics*, 15(1), 385–397. <https://doi.org/10.1214/aos/1176350273>
- Risley, M. (2023). *High-level sampling and weighting requirements for the 2024 American Community Survey Sexual Orientation and Gender Identity Test*. Forthcoming.
- Truman, J.L., Morgan, R.E., Gilbert, T., & Vaghela, P. (2019). Measuring sexual orientation and gender identity in the National Crime Victimization Survey. *Journal of Official Statistics*, 35(4), 835-858. <https://doi.org/10.2478/jos-2019-0035>
- Truman, J.L., & Morgan, R.E. (2022). *Violent victimization by sexual orientation and gender identity, 2017-2020*. Bureau of Justice Statics. NCJ 304277. Retrieved December 21, 2023, from <https://bjs.ojp.gov/library/publications/violent-victimization-sexual-orientation-and-gender-identity-2017-2020>
- U.S. Census Bureau. (2022a). *U.S. Census Bureau statistical quality standards*. Retrieved November 7, 2023, from <https://www2.census.gov/about/policies/quality/qualitystandards.pdf>
- U.S. Census Bureau. (2022b). *American Community Survey and Puerto Rico Community Survey Design and Methodology, Version 3.0*. Retrieved November 7, 2023, from <https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>
- Wilson, B. D. M., & Meyer, I. H. (2021). *Nonbinary LGBTQ adults in the United States*. Williams Institute UCLA School of Law. Retrieved January 5, 2024, from <https://williamsinstitute.law.ucla.edu/publications/nonbinary-lgbtq-adults-us/>