RECOMMENDATIONS ON THE BEST PRACTICES FOR THE COLLECTION OF
SEXUAL ORIENTATION AND GENDER IDENTITY DATA ON FEDERAL STATISTICAL SURVEYS

Pursuant to the June 2022 Executive Order 14075 on Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex Individuals, the Office of the Chief Statistician of the United States developed this report to provide recommendations for Federal agencies on the current best practices for the collection of self-reported sexual orientation and gender identity (SOGI) data on Federal statistical surveys. These recommendations build on a long history of robust Federal effort to develop and refine SOGI measurement best practices.

The report highlights the importance of continual learning, offers considerations for including SOGI items on surveys, provides example approaches for collecting and reporting this information, offers guidance on how to safeguard SOGI data, and concludes with a summary of challenges that need further research. **It is not the intent of this report to limit the continued evolution and improvement of SOGI data collection methods.**

This report does not cover the collection of SOGI data about individuals as part of administrative transactions, for example, by way of forms required to apply for a job, benefits, or services. These administrative collections have different quality, privacy, legal, and ethical concerns that are outside the scope of this document.

**This report does not mandate any particular approach or create any new requirements for agencies.** In the future, Federal agencies may need to diverge from the recommendations in this report to reflect new, evidence-based best practices.

THE IMPORTANCE OF FEDERAL STATISTICS

Federal surveys play a vital role in generating the data that the public, businesses, and government agencies need to make informed decisions. Measuring sexual and gender minority (SGM) populations\(^1\) in Federal surveys improves understanding of SGM populations and supports evidence-based policymaking. By asking respondents about their sexual orientation and gender identity, Federal surveys have shown that, for example:

- the lesbian, gay, bisexual, or transgender (LGBT\(^2\)) community was hit harder by the economic impact of the COVID-19 pandemic [10],
- LGBT adults struggled more with mental health during the COVID-19 pandemic than non-LGBT adults [9],

---

\(^1\) Sexual and gender minority (SGM) populations include, but are not limited to, individuals who identify as lesbian, gay, bisexual, asexual, transgender, Two-Spirit, queer, and/or intersex. Individuals with same-sex or -gender attractions or behaviors and those with a difference in sex development are also included. These populations also encompass those who do not self-identify with one of these terms but whose sexual orientation, gender identity or expression, or reproductive development is characterized by non-binary constructs of sexual orientation, gender, and/or sex. ([NIH Sexual & Gender Minority Research Office](https://www.nih.gov/)).

\(^2\) This report uses the acronyms “LGBT”, or “LGBTQI+” depending on the source material being referenced.
college students who identify as gender minorities have had more difficulty finding safe and stable housing [33], and

- the rate of violent crime victimization of lesbian or gay persons has been more than two times the rate for straight persons [38].

**BEST PRACTICES EVOLVE THROUGH CONTINUOUS LEARNING**

SOGI measures need to be flexible and adapt over time to maintain usefulness. In addition to changes in terminology over time other changes could impact the ways SOGI data should be collected to meet the purposes of various surveys.

Many of the Federal agencies that currently collect SOGI data continue to conduct rigorous research and testing. Agencies undertake this work to ensure that measures are developed with the utmost care and consideration, reflecting the personal nature of these topics for survey respondents, and the importance of collecting accurate data. These research efforts assess the accuracy of responses, how difficult it is for an individual to respond to the questions, how comfortable respondents are answering the questions, and other relevant topics. Question development and implementation efforts are also informed by ongoing engagement with data providers and users.

Federal staff across agencies collaborate to learn from collective experience and promote development of questions that can be used across Federal data collections. In order to learn from interagency practices and expertise, The Office of Management and Budget first convened the Federal Interagency Working Group on Measuring Sexual Orientation and Gender Identity in 2015. This group was charged with addressing the lack of data on SGM populations, and documenting methodological considerations that support successful collection and accurate measurement of SOGI data. This working group concluded its tenure with publication in 2016 of three highly-cited papers. The papers documented SOGI measures in Federal surveys and the testing used to develop them, summarized results from evaluations of SOGI measures, and provided recommendations for a research agenda to continue improvement and adoption of SOGI measures in the Federal government.

In addition, the Federal Committee on Statistical Methodology’s (FCSM) Measuring Sexual Orientation and Gender Identity (SOGI) Research Group provides an ongoing venue for collaboration among Federal staff. This group focuses on exploring the measurement of SOGI, considering the differing dimensions of sex, gender, and sexuality. Any Federal agencies considering adding, expanding, or revising the collection of SOGI data are encouraged to participate in the FCSM SOGI Research Group.³

This report draws heavily on the work of the Federal agencies and interagency groups that have dedicated significant time and resources to researching, coordinating, and sharing best practices. As a result of these efforts, many Federal surveys are currently collecting valuable information on SGM populations. Best practices will continue to evolve with ongoing research on challenges such as proxy

---
³ Those interested in joining the FCSM SOGI Research Group should contact the group’s co-chairs for more information (contact information is available on the group’s website).
reporting, youth respondents, evolving terminology, translation into other languages, and privacy protection, which will certainly continue to improve the quality and availability of SOGI statistics.

**CONSIDERATIONS FOR INCLUDING SOGI ITEMS ON SURVEYS**

The following section presents a summary of practices and considerations in this area. As mentioned above, while we believe many of these practices and considerations will have ongoing relevance to SOGI data collection, we are also presenting these practices with acknowledgement of the dynamic and evolving nature of this subject.

**Starting with the planned uses:** As required by the Paperwork Reduction Act (PRA), Federal agencies must ensure that all survey questions provide useful data that meet the intended purposes without unduly burdening respondents.⁴ Consider how resulting data will be used when adding SOGI questions to a survey. For example, agencies may want to collect SOGI data to detect differences between groups, to estimate population sizes or features, and to meaningfully inform policies and programs intended to address inequality. As always, data should only be collected if there are plans to publish statistics, release data to the public or researchers, or to otherwise support the mission of the agency; and to minimize burden and privacy risk to respondents, only the minimum amount of information needed to meet the planned uses should be collected. The ability to protect confidentiality should also guide decisions about when and how to ask for this information. Keep in mind that sexual orientation and gender identity are two separate concepts. It may not always be necessary to ask about both.

**Strategies to ensure sufficient sample size:** There’s no best practice for the minimum survey sample size needed to collect SOGI data. In addition to the disclosure risk created by small samples, sample size also affects the ability of a survey to detect differences between groups. Underpowered studies have a high risk of finding that there are no statistically significant differences between groups in the sample data, even when real inequality exists in the population. During the survey planning stage, Federal agencies may find it helpful to use current estimates of SGM population prevalences⁵ to conduct a power analysis for key survey outcomes and to check estimated cell sizes against agency standards for publication.⁶ Surveys may need to combine data across years, reduce geographic detail, or collapse respondents into a single “sexual or gender minority” category in order to release results. Federal agencies may find that considering what information about SGM populations is most important to their analytical goals will be helpful when deciding what methods they use to address small population size.

**Considering sensitivity and burden for all respondents:** There may be some concern that adding SOGI questions to a survey will cause respondents to skip questions or abandon the survey altogether, harming the quality of the entire collection. But current surveys collecting SOGI data show that respondents are unlikely to skip SOGI questions (low item nonresponse), especially compared to other sensitive data items [1, 2, 4, 5, 6, 16, 15, 25, 33, 34]. Item nonresponse varies, though, by demographic group, with older people, women, non-Hispanic African Americans, Asians, and Hispanics, and those

---
⁴ For more information on agency responsibilities under the Paperwork Reduction Act, see https://pra.digital.gov/
⁵ For example, both Gallup and the Williams Institute release regular estimates of the LGBT population.
⁶ See, for example, Chapter 5 of the Statistical Standards for the National Center of Education Statistics.
with less education having higher rates [5, 11, 16, 17, 18, 31, 39]. The **addition of SOGI items typically does not cause significant survey breakoffs in household surveys** [5], although this also varies by demographic group [38].

**Using tested terminology:** The terminology used to describe sexual orientation and gender identity, and the way people identify with those concepts is fluid and evolving. It is **difficult to find response options that both comprehensively reflect LGBTQI+ identities and are universally understood** [24], particularly for youth or young adults [2, 3, 8, 19, 30, 35, 37]. Therefore, **it’s important to test terminology with a broad range of respondents, or rely on previously tested terminology.**

**Using tested translations:** Research to date has illustrated the **difficulty in translating SOGI terms.** Translation challenges include conveying the intended meaning in a culturally and linguistically appropriate way as well as a lack of terminology that conveys concepts in some languages [17, 20, 21, 22, 27, 29, 36].

**Using design elements that improve data quality:** **Allowing respondents to select multiple response options (mark all that apply)** prevents forcing respondents to choose between multiple applicable identities or orientations. When feasible, adding a **write-in response option can provide valuable information** [39]. However, evaluation of write-in responses show that not all responses indicate SGM status [27,41]. Write-in responses must be individually coded in order to use them as an indicator of SGM status, and data files should indicate when a response was coded. SOGI questions are typically **placed at the end of demographic questions,** with the gender identity series (as shown in the **Example SOGI Module** below) appearing together (i.e., not split up) for appropriate context. Because sex assigned at birth does not define gender [25], respondents should not be asked to provide their sex assigned at birth unless they are also given the opportunity to provide their current gender identity.

**Confirming gender minority status:** Every survey question suffers some degree of measurement error resulting from errors in data entry. Whenever an individual’s responses to the two-step gender identity items classify the respondent as a gender minority a **confirmation questions can be used** to verify. For example, a respondent who indicates their sex at birth was “Female” and selects “Male” as their current identity would be asked to confirm these responses (see **Example SOGI Module** below). Due to the relatively small size of the gender minority population, even small data entry error rates can result in a significant proportion of false positives within the set of respondents classified as gender minorities. This can greatly reduce a survey’s ability to detect real differences in outcome measures [22]. False negatives present a much lower risk of impacting analyses of differences between groups, although they may lead to underestimation of SGM population size. Giving a confirmation question to all respondents may be indicated if a goal of the survey is to estimate population size.

**HOW TO ASK ABOUT SEXUAL ORIENTATION AND GENDER IDENTITY**

There is no single, best practice set of questions for soliciting information about a person’s sexual orientation or gender identity. The context of the data collection, including planned uses for the data, alignment with other surveys or datasets, sample size, ability to code write-in responses, and ability to protect confidentiality should guide decisions about when and how to ask for this information. This
section provides some example approaches for asking about sexual orientation or gender identity. These examples are not an exhaustive presentation of appropriate ways to collect SOGI data.

More in-depth measures such as the *Example SOGI Module* below can support detailed and disaggregated reporting of sexual and/or gender minority individuals. This multi-question approach provides a structure capable of maximizing the amount of detail reported for both the sexual orientation and gender identity items, facilitating a more comprehensive identification. Response options may be modified as needed to meet specific data needs or as indicated by new evidence. Responses can also be aggregated to different levels (e.g., SGM) as needed for compliance with agency disclosure prevention protocols. This module, with slight variations, is typical for the large, general purpose statistical surveys that currently collect SOGI data.

A less detailed approach to collecting gender identity such as the *Example Gender Question* below can provide higher-level results that may meet data needs while minimizing burden, sensitivity, and privacy risk. This approach adds a third response option to traditional binary measures and collects only basic information about an individual’s current gender identity. This approach may suit surveys with smaller sample sizes or where privacy and confidentiality are of heightened concern, such as internal staff surveys. Surveys that currently collect gender using only binary response options (i.e., “male”/“female”) can use this approach to update to a more inclusive question.

A single question about LGBT status such as the *Example LGBT Status Question* below can provide the necessary information to determine a respondent’s SGM status by asking about aspects of sexual orientation and gender identity at the same time. This type of question may be useful as a screener question or as a tool for basic equity analysis.

Other tested and verified measures can be found in the *Additional Resources for Asking about SOGI* section below.

### EXAMPLE SOGI MODULE

#### SEXUAL ORIENTATION

Q1. Which of the following best represents how you think of yourself?

- Gay or lesbian
- Straight, that is not gay or lesbian
- Bisexual
- I use a different term [free-text]
- I don’t know

#### GENDER IDENTITY

---

7 This module is based on the module in the Census Bureau’s *Household Pulse Survey*, which was in turn based on modules in the Bureau of Justice Statistic’s National Crime Victimization Survey and the National Center of Health Statistic’s National Health Interview Survey.
Q1. What sex were you assigned at birth, on your original birth certificate?

- Female
- Male

Q2. How do you currently describe yourself (mark all that apply)?

- Female
- Male
- Transgender
- I use a different term [free-text]

Q3. Just to confirm, you were assigned (FILL) at birth and now you describe yourself as (FILL). Is that correct?8

- Yes
- No <skip back to Q1 and/or Q2 to correct>

EXAMPLE GENDER QUESTION9

Q1. Are you:

Mark all that apply.

- Female
- Male
- Transgender, non-binary, or another gender

EXAMPLE LGBT STATUS QUESTION10

Q1. Which of the following do you consider yourself to be? You can select as many as apply.

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

8 To minimize false positives, Q3 should be asked of individuals whose response to Q1 (sex at birth) differ from their response to Q2 (current gender identity).
9 This question is based on ongoing research by the National Center for Health Statistics.
10 This question is based on the Gallup Poll.
REPORTING SOGI INFORMATION

Federal agencies may consider strategies such as pooling data from multiple time points (e.g., months or years of survey data), aggregating detailed groups where needed, and providing measures of uncertainty to generate the most useful results. More information on these strategies can be found in the FCSM’s Framework for Data Quality.

SEXUAL ORIENTATION REPORTING CATEGORIES

Depending on sample size, the recommendation is to report out five categories for “sexual orientation.” Evaluation of National Health Interview Survey (NHIS) data found that people who use “Don’t know” are qualitatively different than those that use “something else” for sexual orientation [6, 7]. Those who report using “a different term” may also be qualitatively different from those who respond “don’t know.” Therefore, reporting these categories out separately can provide valuable information. If sample size allows, estimates may be further disaggregated (for example collecting and reporting “gay” and “lesbian” separately).

1. Gay or lesbian
2. Straight
3. Bisexual
4. A different term
5. Don’t know

In the event the respondent did not answer the question, data can be reported as ‘missing.’ Imputed data on sexual orientation may be unreliable. Because sexual minority groups are relatively small, a small change resulting from imputation procedures can result in relatively large changes in statistical results. Categorizing the response as missing avoids risks associated with imputed data.

GENDER REPORTING CATEGORIES

When using a two-step approach to measuring gender, as depicted in the Example SOGI Module, the recommendation is to report out four categories for gender based on the following logic when the analytical goal is to illuminate any potential differences in outcome measures for gender minorities. If sample size allows, estimates may be further disaggregated (for example reporting trans men and trans women, or frequent write-in responses, separately).

1. Cisgender Male – would be determined by Q1 “male” AND Q2 “male” OR “I use another term”, accompanied by a write-in response that codes as male.
2. Cisgender Female – would be determined by Q1 “female” AND Q2 “female” OR “I use another term”, accompanied by a write-in response that codes as female.

---

11 Many surveys, including the NHIS, use “something else” as a response option for sexual orientation. This report instead recommends using the response option “I use a different term” followed by a free-text field, based on the findings of a recent NASEM panel [25].

12 Write-in responses may be coded and included in other reporting categories where appropriate. Responses that don’t fit into other categories or that don’t have written clarification may be reported as “a different term.”
3. Gender minority - would be any of these combinations:
   a. Q1 female and Q2 male
   b. Q1 male and Q2 female
   c. Q2 transgender
   d. Q2 “I use another term”, accompanied by a write-in response that codes the response as gender minority

4. Another gender identity – would be determined by Q2 “I use another term”, unless accompanied by a write-in response that codes the response into a different category

There is not enough evidence at this time to support classifying those who report “another gender identity” into the “gender minority” category, unless write-in responses are collected and coded to one of the first three reporting categories.

In the event the respondent did not answer one or both of the questions, data can be reported as ‘missing’ to prevent obscuring differences between groups. As in Sexual Orientation, gender identity based on imputed data may be unreliable.

SAFEGUARDING SOGI DATA

Survey data are typically linked or linkable to a specific individual. Even slight modifications to existing binary measures of sex or gender on surveys can create significant privacy risks for respondents. Removal of direct identifiers (e.g., name or telephone number) only mitigates some of the privacy risk; it may be possible to determine an individual’s SGM status if re-identified through other information on the survey. Unless properly safeguarded against re-identification, breaches, and mishandling, SGM status revealed to unauthorized parties or for unauthorized purposes could be used to target individuals, deny them access to programs or services, or cause other harms. Steps to minimize privacy risk not only reduce the likelihood of such harms, but also promote the trust needed for respondents to feel comfortable providing data in a way that maintains the usefulness of the data.

In implementing these safeguards, Federal agencies should engage relevant agency experts on data governance, including their Statistical Official, agency general counsel, and the Senior Agency Official for Privacy (SAOP), who is responsible for ensuring compliance with applicable privacy requirements and managing privacy risks consistent with the agency’s mission. Agencies are responsible for complying with the Privacy Act of 1974, if applicable, and any other applicable laws, regulations, and policies (e.g., OMB Circular A-130). The FCSM Data Protection Toolkit contains helpful tools for maximizing the usefulness of data while protecting privacy.

As with any other data, and in accordance with applicable laws, regulations, and policies, agencies should:

---

13 OMB Memorandum M-17-12, Preparing for and Responding to a Breach of Personally Identifiable Information, § III(C) (Jan. 3, 2017).
14 OMB Memorandum M-16-24, Role and Designation of Senior Agency Officials for Privacy (Sept. 15, 2016).
• Evaluate the privacy risks associated with collecting, maintaining, and using such data throughout its life cycle\textsuperscript{15} - including when involving intermediaries (e.g., state, local, tribal, and territorial governments; schools; grantees; contractors; or other entities) - and balance the need for new information against any privacy risks;\textsuperscript{16}
• Have a plan for data governance across the data life cycle, including disposition of data;
• Assess available statutory authority to protect the confidentiality of survey data, and make sure that the agency provides respondents with the strongest confidentiality protections available;
• Carefully control access to microdata, especially if it is linked or linkable to a specific individual;
• Develop a plan to handle Freedom of Information Act requests that could disclose sensitive data

\begin{mdnote}
\textbf{ONGOING RESEARCH}
\end{mdnote}

The Federal statistical system is actively conducting further research, but currently does not have the ability to provide best practices or recommendations for several areas. Additional research is needed for: bridging and linking to binary gender measures; collecting SOGI data by proxy, from youth respondents, in employment settings, on administrative forms, or in languages other than English; and collecting data about the intersex status of respondents [15]. Additional information is detailed below.

For non-response bias analysis, longitudinal analysis, weighting responses, or generating blended estimates from multiple data sources, gender identity data may need to be linked with other datasets or bridged to previous survey rounds that collected gender with only binary (male/female) response options. Any approach that an agency takes to transform more inclusive gender identity data to a binary set of response options will by necessity discard or disregard information provided by respondents about how they identify.

Little is known about the ability and willingness of respondents to answer SOGI questions on behalf of other household members (i.e., proxy response). It is common practice to ask one person to provide information for other members of the household on Federal surveys. There is some concern that proxy reporting to SOGI questions may generate inaccurate responses because respondents may not be comfortable sharing information about others or may not know the information. However, the limited research carried out to date suggests respondents are generally willing and able to answer for other members of the household [13, 26]. Further research is needed to evaluate proxy versus self-response, as well as item nonresponse to these questions to ensure they produce high quality measures on Federal household surveys.

There are special considerations when determining what age groups receive these questions and how the information is solicited, particularly if asking SOGI questions of younger populations (i.e., the population under 18 years of age). Question wording and response options that are appropriate for asking these questions of adults are not necessarily sufficient for soliciting information from youth respondents. There are unique issues associated with asking SOGI questions among youths, such as parental consent and assent laws, as well as designing data collection methods that protect respondent

\textsuperscript{15} OMB Circular A-130, Managing Information as a Strategic Resource, Appendix II \S\ 4 (Jul. 28, 2016).
\textsuperscript{16} OMB Circular A-130, Managing Information as a Strategic Resource, \S\ 4(i) (Jul. 28, 2016).
privacy. In addition, there are challenges with greater salience among younger populations, such as outness, fluidity over time, dynamic and evolving terminology, and uncertainty/questioning status [34].

Further research is needed on the sensitivity and potential impacts on response rates, breakoffs, and response accuracy when adding SOGI items to non-household surveys, such as business surveys and employee surveys. For example, while studies of household surveys have not shown significant breakoffs, a survey of farm operators did find that adding SOGI items resulted in reduced response rates [42]. While technically beyond the scope of this report, more research is also needed on how to collect and protect data collected as part of administrative transactions, for example, by way of forms required to apply for a job, benefits, or services.

Additional research is needed on best practices for collecting SOGI data in languages other than English using translations that are linguistically and culturally appropriate.

Finally, additional research is needed on how best to ascertain information about intersex status (i.e., people who are born with, or naturally develop over time, sex traits that do not correspond to a single sex), particularly in the field of survey research [25].

OMB encourages continued testing and research to address these issues with the goal of improving the quality, availability, and consistency of sexual orientation and gender identity data.
ADDITIONAL RESOURCES FOR ASKING ABOUT SOGI

FEDERAL RESOURCES

Federal Committee on Statistical Methodology’s (FCSM) Measuring Sexual Orientation and Gender Identity (SOGI) Research Group


National Institutes of Health Sexual & Gender Minority Research Office

Centers for Medicare & Medicaid Services Sexual and Gender Minority Clearinghouse

National Academies Measuring Sex, Gender Identity, and Sexual Orientation (2022)

National Academies Understanding the Well-Being of LGBTQI+ Populations Consensus Study Report (2020)


NON-FEDERAL RESOURCES

How Do You Measure the LGBT Population in the U.S.? (gallup.com)

Survey Measures - Williams Institute (ucla.edu)

Best Practices for Asking Questions about Sexual Orientation on Surveys (SMART) - Williams Institute (ucla.edu)

Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys (GenIUSS) - Williams Institute (ucla.edu)

Summary and Final Recommendations – California Health Interview Survey Sexual Orientation and Gender Identity Working Group

Measuring Aspects of Sexuality and Gender: A Sexual Human Rights Challenge for Science and Official Statistics | CHANCE (amstat.org)

REFERENCES


Materials posted on this page are released to inform interested parties of research and to encourage discussion. Opinions expressed are those of the authors and do not reflect the official position of the Office of the Chief Statistician of the United States or of the Office of Management and Budget.


[34] Spock, Strategies to Improve Measurement of Sexual Orientation and Gender Identity Among Youth, 2022, Retrieved at: https://www.jahonline.org/article/S1054-139X(22)00646-2/fulltext


