

Evaluations of Sexual Orientation and Gender Identity Survey Measures: What Have We Learned?

September 23, 2016

Contents

Executive Summary	3
I. Background and Purpose of the Working Paper	8
II. Relevant Concepts	8
a. Orientation, Identity, and Relationships	9
<i>Sexual Orientation</i>	9
<i>Sex and Gender</i>	9
<i>Household Relationships</i>	10
b. Measurement Error	10
<i>Sources of Error</i>	11
<i>Detecting Measurement Error</i>	11
III. Research Methods	12
IV. Results of Evaluation Studies	13
a. Sexual Orientation	13
<i>Sexual Identity</i>	13
<i>Sexual Attraction</i>	19
<i>Sexual Behavior</i>	22
<i>Congruency of Multiple Measures of Sexual Orientation</i>	23
b. Gender Identity	23
<i>Gender Expression</i>	28

c. Household Relationships	29
V. Discussion: Best Practices and Areas for Further Research.....	31
a. Sexual Orientation.....	31
<i>Sexual Identity</i>	31
<i>Sexual Attraction</i>	32
<i>Sexual Behavior</i>	32
b. Gender Identity	32
c. Household Relationships.....	33
d. Conclusion	34
References	35
Appendices	40
A. Surveys Measuring Sexual Identity in Cited Work.....	40
B. Sexual Identity Items Used in Cited Work.....	41
C. Sexual Attraction Items Used in Cited Work.....	49
D. Sexual Behavior Items Used in Cited Work.....	52
E. Gender Identity Items Used in Cited Work	53
F. Gender Identity Items in Federal Surveys	58
G. Scales Used to Measure Gender Expression.....	60
H. Measures of Household Relationships.....	61

Executive Summary

At a time when sexual and gender minority (SGM) populations are becoming more visible in social and political life, there remains a lack of data on the characteristics and well-being of these groups. In order to better understand the diverse needs of SGM populations, more representative and better quality data need to be collected. The U.S. Federal Government is taking several steps to more efficiently coordinate data collection efforts across its many departments. The Office of Management and Budget (OMB) convened the Federal Interagency Working Group on Measuring Sexual Orientation and Gender Identity (SOGI IWG) to begin addressing the deficiency of information for these populations and the concerns surrounding methodological issues in collecting such data. This document is the second in a series of working papers from the Federal Interagency Working Group that describe what is known about current measures on sexual orientation, gender identity, and household relationships and highlights the methodological questions that remain to be answered.

Although a few Federal agencies have collected information describing SGM populations for over a decade, some aspects of sexual orientation and gender identity (SOGI) have been more routinely measured than others. Federal agencies not currently collecting data on SOGI have expressed interest in doing so. However, before adopting a particular measure in any Federal data collection, it is important to evaluate the rigor of the measure, understand its measurement properties for the population of interest, and determine whether the measure is fit for its intended purpose.

This working paper identifies any current survey measures of SOGI where measurement error has been assessed and determines the nature and extent of such error. It also describes best practices based on these assessments to minimize SOGI measurement error in surveys.

For a meta-analysis of evaluations of SOGI measurement error in surveys, a comprehensive literature review was conducted. This included a review of Federal agency directed—or sponsored—reports, proceedings papers of the American Association for Public Opinion Research (AAPOR) conferences, and literature searches related to the topic of SOGI measurement. For literature searches, keywords related to “sexual orientation survey measurement” and “gender identity survey measurement” were used to search online academic research databases. These searches yielded reports, articles, and chapters from academic journals and books, including those in the fields of sociology, psychology, demography, public health, survey methodology, sexuality and gender research, and public opinion research. In order to yield the most relevant findings for the purposes of this paper, the scope of this literature review was limited to publications occurring between January 2011 and July 2016 (exceptions as noted). The review focused primarily on SOGI measurement research from surveys conducted in the United States, but some international studies were included.

Results of evaluation studies

The major findings from evaluation studies on the measurement of SOGI are organized around the concepts of *sexual orientation*, *natal sex*, *gender identity*, and *same-sex and opposite-sex households*.

Sexual orientation has three main dimensions: sexual identity, sexual attraction, and sexual behavior. Research on *sexual identity*, defined as the way someone identifies with a given sexual orientation, was found to be the most prevalent among the studies reviewed. Research in this area has focused on a diverse group of respondents, including sexual and gender minority populations, adult and adolescent populations, and English- and Spanish-speaking populations. The sexual identity studies reviewed used

both quantitative and qualitative methods to examine item performance and item nonresponse. In general, sexual identity items tested in these studies performed well, and most respondents appeared to have little difficulty answering the items. Some research did identify differences in responses/estimates, depending on how the item was worded and collected. Terminology differences and comprehension issues were particularly poignant for older adults (ages 70 and older) and less educated adults. For example, respondents experienced comprehension issues with “heterosexual,” suggesting “straight” is the more common term. Research on sexual identity measurement has also revealed issues related to language, particularly among Spanish-speaking respondents, as terms like “straight” have no translation; therefore, translations presenting the “not gay” (i.e., “No gay, o sea, heterosexual”) explanation first are recommended. While some research with adolescents has found sexual identity items to perform well, other research has found that sexual identity implies a degree of permanence that adolescents are not comfortable addressing. Additionally, some research identified social desirability biases when sexual identity questions were asked in less private modes. Finally, while item nonresponse rates to sexual identity items were relatively low across studies, they were found to vary by respondents’ demographic characteristics, including sex, age, race, and ethnicity.

Research on *sexual attraction*, which refers to the sex or gender to which someone feels attraction (e.g., whether an individual is attracted to males/men, females/women, or both), was more limited than the research on sexual identity. Half of the studies reviewed used cognitive interviewing techniques, while the other half involved quantitative evaluations of survey questions. The target populations differed across the studies, including adult women only, younger persons ages 12 to 21, and adults of all ages. One study found no significant differences in the reporting of same-sex attraction when a method that was proven to reduce social desirability was used. Another study found that respondents not only described sexual attraction as physical, but also included other factors such as affection, affiliation, and emotional preference. Consistent with this finding, interviews with younger respondents, ages 15 to 21, felt that the focus on “sexual” attraction was too restrictive, and reported other types of attraction. This same research also examined a combined question (identity and attraction), and found mixed results, including confusion between the two terms, which led to conflating the experiences (i.e., some respondents described themselves as gay or lesbian, but also felt some attraction to members of the opposite sex). In particular, among adolescent participants, research found that a sexual attraction question may be easier to answer since they may still be in the process of “figuring out” their sexual identity. Finally, quantitative results of multiple question formats for sexual attraction, including other items on sexual identity and behavior, found that item nonresponse varied by question wording, whether or not the concepts were combined, administration of the survey, and by respondents’ characteristics.

The studies that examined *sexual behavior*, which refers to the sex of a person’s sexual partners (e.g., individuals of the same sex, different sex, or both sexes), focused on how sexual behavior items performed, item nonresponse, and respondents’ reactions to capturing their sexual behavior. In general, researchers assess sexual behavior by asking respondents to report the sex of their sexual partners during a specified period of time in the past and/or during their lifetimes. Participants had mixed reactions to sexual behavior items, including feeling that items were too personal or being excited to see sexual behavior being measured in a health survey. Consistent with other research, nonresponse on sexual behavior items was found to vary by respondent age, education, and income. Research also found issues of misreporting, that is, when non-heterosexuals reported their behavior as exclusively heterosexual and when heterosexuals did not report same-sex sexual behavior. This also varied by sex, income, and age.

Additionally, a few studies examined the congruency of various measures of sexual orientation. In general, research found that a person's sexual identity is not necessarily congruent with their sexual attraction or behaviors. Some researchers have concluded that sexual orientation may be viewed on more of a continuum of sexuality. Given these findings, the appropriate measurement of sexual orientation will depend on the primary focus of the research.

There is a growing body of research on *gender identity*, or an individuals' self-identified sense of gender. Nonetheless, there is still no consensus on how to collect gender identity in surveys. The majority of research supports using a two-step method (capturing assigned sex at birth and current gender identity), per the recommendation of the Gender Identity in the U.S. Surveillance (GenIUSS) Group, World Professional Association for Transgender Health (WPATH), and Center of Excellence for Transgender Health (CoE). For example, research collecting SOGI at medical intake found that participants understood and were willing to answer the two-step items. Among other populations researched, including youth, the majority of respondents found questions on both natal sex and gender identity to be straightforward. Research has also tested this among Spanish-speaking participants who were able to answer the item related to natal sex, but the majority conflated sexual identity and gender identity. However, respondents still responded in a manner that was consistent with a gender item collected during the study screener. Even with this research, there are a number of open questions, including how to the order the items, whether to include reference to a birth certificate, and how the items perform in languages other than English and Spanish. Also unresolved is the appropriateness of using gender identity in realms that typically utilize binary measures of sex, such as gendered questions.

Research on *gender expression*, or an individual's external manifestation of gender, is also growing, but the number of evaluation studies is small. The studies examined on this topic asked respondents how they perceive their masculinity and femininity (first-order scales) and also how they think others perceive them (third-order scales), for a total of four scales. Research suggests that there is utility in asking all four scales, and that the masculine and feminine scales should be separated. The research also recommends asking the first-order scales if researchers have to choose one set of scales.

Household relationship data can be used to provide an indirect estimate of the SGM population through the measurement of same-sex couples. There were several studies identified in this working paper that attempted to quantify inconsistencies and over-reports in the measurement of same-sex couples using household data, and identify question design changes to improve the performance of sex and relationship questions for estimating the number of same-sex couples. These studies identified persistent inconsistencies in reports of sex and relationship to householder, leading to overestimates of same-sex households in the decennial Census, ACS, and other Federal surveys. The overestimates were largely attributed to accidental mismarking on sex by a very small proportion of opposite-sex married couples. While identifying and addressing the issues behind the documented inconsistencies is important for producing accurate estimates of same-sex couple households (and a number of other couple types), relying on sex and relationship information does not provide a direct measure of sexual orientation.

Areas for future research

This working paper reviewed recent studies evaluating the measurement of SOGI in surveys and has identified areas where future research is needed.

For *sexual identity*, the literature reveals a number of ways to ensure that reliable and accurate data can be collected. Question order studies are also an area of future inquiry, as placing sexual identity items near demographic survey questions rather than around sexual attraction and behavior questions may lead to less biased responses. Future research may be best focused on sexual identity measures among certain populations, including adolescents, older adults, and racial, ethnic, and language minorities, where research is more limited and the understanding of sexual identity terminology may differ. Finally, other areas for future research include survey operations and administration, pertaining to interviewer training and limitations to using proxy measurement.

Future research on *sexual attraction* should consider the definition of attraction as well as challenges with the understanding of question response options. There is research needed on question wording and response options, as well as mode of data collection. Question context and placement should also be considered. Nonresponse may also be a concern when measuring sexual attraction, and the conflation of sexual attraction and sexual identity common to many existing measures may contribute to item nonresponse. Consistent with sexual identity, it is recommended that more testing be conducted with adolescents to understand how they comprehend measures of sexual orientation broadly as well as sexual attraction specifically.

More research on *sexual behavior* that focuses on the structure and content of questions and that looks at how responses vary by demographic group would be beneficial. Also, both nonresponse and misreporting are areas that would benefit from future research, as existing findings suggest that nonresponse and misreporting vary by respondent demographics.

There are a number of areas in which Federal research could add to the literature on how best to measure *gender identity*. While many organizations support a two-step measure, more research is needed on the assessment of gender identity using a one versus two-step approach. In addition, little is known about how question order may affect response for the two-step measure. Also benefiting from further research is the effect of referencing “original birth certificate” for the natal sex question. Evaluation research would also benefit from the inclusion of different racial, ethnic, cultural, and language minorities. For example, future evaluation research focused on increased recruitment or oversampling of transgender individuals, people of color, and people from cultures in which a third gender exists (e.g., “two-spirit” in American Indian tribes and “hijras” in South Asian cultures) would advance the literature. Recent variations of the two-step measure have included a verification item when natal sex and current gender identity do not agree. While items like these have been shown to reduce measurement error in the measurement of same-sex households, there are no studies examining their application to the measurement of gender identity. Finally, there are other methodological challenges related to the measurement of gender identity, including how to best to ask questions that have implications for public health, such as questions about mammograms, Pap tests, and birth control, once a respondent is identified as transgender.

Moreover, there appears to be limited research evaluating items measuring *gender expression*, or an individual’s external manifestation of gender. Additional research addressing this topic would prove beneficial and further develop knowledge on this concept central to SOGI research.

Studies measuring same-sex and opposite-sex households identified persistent inconsistencies in sex and relationship-to-householder reports, leading to overestimates of same-sex households in the decennial Census, ACS, and other Federal surveys. The overestimates are largely attributed to accidental mismarking on sex by a very small proportion of opposite-sex married couples. While many of the

testing efforts described recommend the use of a revised relationship to householder question, the evidence to date suggests that inconsistencies and overestimates of same-sex couple households are still likely. Additionally, a number of the reviewed studies mentioned the importance of automated edits in interviewer-administered surveys for reducing sex-relationship inconsistency rates. However, these studies failed to quantify the extent to which these edits led to changes in reporting of sex, relationships, or both. Furthermore, we know very little about the behavior of interviewers when confronted with these edits.

I. Background and Purpose of the Working Paper

At a time when sexual and gender minority (SGM)¹ populations are becoming more visible in social and political life, there remains a lack of data on the characteristics and well-being of these groups. In order to better understand the diverse needs of SGM populations, more representative and better quality data are needed. The U.S. Federal Government is taking several steps to more efficiently coordinate data collection efforts across its many departments. The Office of Management and Budget (OMB) convened the Federal Interagency Working Group on Measuring Sexual Orientation and Gender Identity (SOGI IWG) to begin addressing the dearth of data for these populations and the methodological issues in collecting such data.

Although a few Federal agencies have collected information describing SGM populations for over a decade, some aspects of sexual orientation and gender identity (SOGI) have been more routinely measured than others. Further, there are Federal agencies not currently collecting data on SOGI that have expressed interest in doing so. However, before adopting a particular measure in any Federal information collection, it is important to evaluate the rigor of the measure, understand its measurement properties for the population of interest, and determine its fit for its intended purpose.

This working paper identifies current survey measures of SOGI for which an assessment of measurement error has been conducted, determines the nature and extent of such error, and, based on those assessments, identifies best practices to avoid SOGI measurement error in surveys.² Specifically, this paper describes the ways in which SOGI has been measured, discusses quantitative and qualitative findings on the quality of existing measures, and identifies areas where future research is needed. Close attention has been paid to how SOGI is defined, how SOGI concepts are understood by respondents, and how measurement challenges (such as respondent misinterpretation, item nonresponse, and social desirability) lead to biased outcomes.

This document is the second in a series of working papers from the Federal Interagency Working Group. The first working paper, *Current Measures of SOGI in Federal Household Surveys*, provides an overview of concepts and current measurement in Federal household surveys and serves as the foundation for this document. A third working paper, *Toward a Research Agenda for SOGI in Federal Household Surveys*, will address the knowledge gaps identified in this working paper by proposing research priorities and strategies that will make the greatest impact on improving SOGI measurement in Federal surveys. Other working papers may be developed as research in this area matures.

II. Relevant Concepts

This section defines key terms and concepts central to this paper, consistent with definitions provided in the first working paper, *Current Measures of SOGI in Federal Household Surveys*.

¹ This working paper refers to the population of interest as SGM rather than as lesbian, gay, bisexual, and transgender (LGBT). We believe that SGM is more inclusive than LGBT because it allows for the inclusion of persons not specifically referenced by the acronym, such as genderqueer, two-spirit, etc.

² Evaluations of SOGI measures used in non-Federal surveys, such as the California Health Interview Survey (CHIS), have been reviewed for this paper because they can inform Federal measurement.

a. Orientation, Identity, and Relationships

Both the purpose of the survey and the specific dimension of SOGI being measured are important design and measurement considerations. Definitions for several key concepts are presented below. These include *sexual orientation, natal sex, and gender identity*. The concept of *same-sex and opposite-sex households* is also presented.

Sexual Orientation

Sexual orientation has three main dimensions: *sexual attraction, sexual behavior, and sexual identity*. *Sexual attraction* refers to the sex or gender to which someone feels attraction (e.g., whether an individual is attracted to males/men, females/women, or both). *Sexual behavior* refers to the sex of a person's sexual partners (e.g., individuals of the same sex, different sex, or both sexes). *Sexual identity* refers to the way a person self-identifies with a given sexual orientation (Sexual Minority Assessment Research Team, 2009).

The most commonly used terms to describe different sexual orientations are *lesbian, gay, bisexual, and heterosexual/straight*.³ In general, people who self-identify as *gay* or *lesbian* are primarily attracted to and/or have sex with people of the same sex. However, the concepts of sexual identity, attraction, and behavior do not always follow these patterns (e.g., individuals may not want to identify as *gay* or *lesbian* even if they are attracted to the same sex or only occasionally have different-sex relations). Therefore, although some surveys focus only on sexual identity, others measure all three components of sexual orientation.

When selecting among measures of sexual orientation, the purpose of the study should guide the dimension(s) measured (Sexual Minority Assessment Research Team, 2009). For example, surveys intended to describe sexual health may feature measures of sexual behavior, while surveys of early adolescents may better feature measures of sexual attraction or identity since early adolescents may be less likely to have sexual behavior experiences. Determining the “right” question for the purpose of the information collection can be challenging. Conceptual dimensions may be fluid over time for individuals and across age cohorts. Academically, these distinctions are clear, but the perceptions by the general public often are not.

Sex and Gender

Sex and gender are foundational concepts in research on SOGI. Generally speaking, the term *sex* refers to the biological characteristics that are used to categorize individuals as *male, female, or intersex*. *Sex* refers to “the genetic, hormonal, anatomical, and physiological characteristics on whose basis one is labeled at birth as either male or female” (Institute of Medicine, 2011, p. 25). The term, *gender*, on the other hand, refers to “the socially constructed characteristics of women and men—such as norms, roles, and relationships of and between groups of women and men” (World Health Organization, 2016).

³ The term *heterosexual* corresponds to those with *straight* identities and/or different-sex attraction and/or partners. The term *homosexual* is sometimes used to describe individuals who are attracted to those of the same gender, engage in sexual activity with those of the same gender, and/or self-identify as *lesbian* or *gay*. Because of its appearance as a mental disorder in the DSM-I and DSM-II, *homosexual* is sometimes seen as unfavorable. It was declassified as a mental disorder beginning with the publication of the DSM-III in 1973.

While *male* and *female* refer to sex, words like *masculine*, *feminine*, *man*, and *woman* all refer to gender.⁴ Gender is a multidimensional construct that has psychological, social, and behavioral dimensions that include gender identity and gender expression. It refers to the “cultural meanings of patterns of behavior, experience, and personality that are labeled masculine or feminine” (Institute of Medicine, 2011, p.25). Gender identity refers to a person’s internal sense of gender (e.g., being a man, a woman, or genderqueer) and potential affiliation with a gender community (e.g., women, trans women, or genderqueer).

The term *transgender* refers to a diverse population that departs significantly from gender norms (Institute of Medicine, 2011). Often, a person's gender identity is consistent with their sex assigned at birth. A person whose gender identity and sex assigned at birth are consistent can be referred to as *cisgender* (that is, a person who gender is “consistent in sex.”). The term *transgender* describes anyone who has a gender identity that differs from their sex assigned at birth (Spade, 2008). Some transgender individuals use hormones or elect for gender-affirming surgery, but not all transgender individuals do this.

There are several dimensions of gender that can be measured: *gender identity*, *gender expression*, and *gender dysphoria*. Some surveys might aim to measure *gender identity*, or an individual’s self-identified sense of gender. Others might be more interested in measuring *gender expression*, or an individual’s external manifestation of gender. A survey may also aim to identify all individuals who experience *gender dysphoria*, or the experience of a marked difference between self-identified gender and assigned gender for a period of at least six months associated with clinically significant distress caused by this incongruence (American Psychiatric Association, 2013). Someone may be diagnosed with gender dysphoria whether or not that person has taken any steps to align their gender expression with their (inner) gender identity.⁵

Household Relationships

Many Federal surveys collect household relationship data for all individuals living in the same housing unit. Household relationship data can be used to provide an indirect estimate of the SGM population through the measurement of same-sex couples. This approach, however, does not provide a direct measure of sexual orientation, and yields an incomplete estimate, since persons not living in the same household as their partner will not be identified (Interagency Working Group on Measuring Relationships in Federal Household Surveys, 2014).

b. Measurement Error

The goal of any survey question is to accurately measure the construct of interest, that is, achieve *construct validity*. However, even when the question or questions represent a valid measure of the

⁴ Some argue that the concepts of *sex*, *male*, and *female* are also socially constructed (Kelly, 2016).

⁵ Gender dysphoria is classified as a mental disorder by the DSM-5. However the DSM-5 states that gender nonconformity is not a mental disorder. Rather, it is the presence of clinically significant stress associated with the untreated condition. Nevertheless, there is some controversy around the categorization of the term, and especially its predecessor, *gender identity disorder*, as a mental illness (Lev, 2016). In fact, a recent study found additional support for classifying health-related categories related to transgender identity outside the classification of mental disorders in the ICD-11 (Robles et al., 2016).

underlying construct, measurement error may occur. For our purposes, measurement error represents the difference between the true value of the measurement or question and the provided response.

Sources of Error

Measurement error can occur for multiple reasons. Perhaps chief among these is clarity (or lack thereof) of the concept to be measured. The complexity and familiarity of the relevant concept can affect the ease with which a participant can respond accurately. Providing an answer involves comprehension of the question and survey task, retrieval of information from long-term memory, assessing the completeness and relevance of memories and a judgment of the level of precision that is called for by the question, and mapping an answer onto the appropriate scale or response option (Tourangeau et al., 2000).

For example, both sexual and gender identity can be conceptualized on a continuum, but they are often measured in nominal categories. Measurement error may occur when respondents feel that they are “between categories,” yet have to map their identities onto a fixed category. Vrangalova and Savin-Williams (2012) found that sexual identity categorized as five groups (i.e., heterosexual, mostly heterosexual, bisexual, mostly gay/lesbian, gay/lesbian) compared to three groups (i.e., heterosexual, bisexual, gay/lesbian) better reflected the nature of sexual orientation components.

Language, translation, and cultural differences can also affect question comprehension. As another example, although sex and gender are defined in research literature as two different concepts, common usage of these terms does not make the same distinction. Indeed, Federal surveys sometimes conflate these terms (Westbrook and Saperstein, 2015). In some cases, this may be done with the objective of collecting the data intended, rather than the data requested. For example, a question designed to measure natal sex may be posed as “What is your gender?” This disconnect often contributes to measurement and design issues. Question format, response option order, and placement within the body of a survey can also affect comprehension (DeMaio and Bates, 2012).

Sensitive questions tend to elicit anxiety in respondents, leading them to alter their responses in attempts to protect their privacy, avoid embarrassment, or avoid providing responses that violate social norms. Asking them in face-to-face interviews, therefore, may lead to social desirability bias, the tendency for a respondent to present him/herself favorably by under-reporting less desirable qualities and over-reporting more desirable traits (Shoemaker et al., 2002; Tourangeau et al., 2000). In addition, social desirability may affect the objectivity and reliability of the manner in which SOGI questions are administered by interviewers. Further, perceived sensitivity of questions can affect the willingness of survey practitioners to include SOGI questions even when inclusion of these measures would support agency mission and data needs.

Detecting Measurement Error

Question evaluation methods such as expert content reviews, focus groups, cognitive interviewing, and usability testing in laboratory settings prior to the launch of a survey are particularly useful for detecting measurement error. The analysis of paralanguage and question probing through read-aloud, reason-aloud, and respond-aloud methods helps the survey methodologist better understand the participant’s understanding and difficulty with target questions.

Field pretests can be especially beneficial. They can feature a target test question and related questions to assess anticipated correlations, known as *concurrent validity*, and can enable statistical approaches such as behavior coding, item response theory, and latent class analysis. Field pretests can also feature split-ballot experiments in which the answers to alternative question wordings, obtained from comparable samples of respondents, are compared for robustness.

Two common indicators of measurement error that can be assessed using pretest as well as production survey data are item nonresponse and the analysis of survey breakoffs.

Item nonresponse is the failure of a survey respondent to answer a specific survey item. A soft refusal may occur when the respondent chooses “don’t know” for a response. A hard refusal occurs when the respondent chooses “refuse.” This is another way that social desirability may influence respondent behavior; respondents whose true answer is not socially desirable may be more likely to respond “don’t know” or refuse to answer the question at higher rates than respondents whose true answer is better accepted by society.⁶ However, item nonresponse may also be an important indicator of potential problems in survey wording or response options. Past research has shown that “don’t know” and “refused” responses may represent distinct forms of item nonresponse attributed to question sensitivity and cognitive effort (Shoemaker, et al. 2002). Questions requiring greater cognitive effort may lead to satisficing – the choosing of an answer, such as “don’t know,” that is adequate rather than optimal (Krosnick, et al. 1996).

Survey break-offs occur when survey respondents do not complete a survey. Break-offs may be intentional or unintentional and may be understood in context to respondent ability, motivation, and task difficulty. When surveys are administered in the field, interviewers may fail to administer specific questions for a number of reasons, resulting in nonresponse error.

Finally, item nonresponse and survey break-offs can contribute to nonresponse bias. Nonresponse bias can occur if the answers of respondents differ from the potential answers of those who did not answer (Little and Rubin, 1987; Singer, 2006). Generally speaking, the lower the response rate, the greater the likelihood of nonresponse bias. SOGI estimates could be susceptible to nonresponse bias if SGM populations are harder to reach or less likely to participate in the survey for other reasons. In addition, given the relatively low percentage of adults identifying as SGM, if nonresponse to SOGI questions is related to respondents’ sexual orientation or gender identity (a case of nonignorable nonresponse), the potential for bias in estimates of SGM may be considerable.

III. Research Methods

This section describes the methods used to assess possible measurement errors associated with measurement of SOGI.

To conduct the proposed meta-analysis of evaluations of SOGI measurement error in surveys, the authors conducted a comprehensive literature review. This included a review of Federal agency directed- or sponsored- reports relevant to the topic. The authors also reviewed proceedings papers of the American Association for Public Opinion Research (AAPOR) conferences that related to the topic of

⁶ It is important to note that for some questions, “don’t know” responses may be valid reflections of the true value. This could especially be true with SOGI questions where a respondent’s identity may yet to be formed, is fluid, or is in flux.

SOGI measurement. In addition, the writing group conducted literature searches using keywords related to “sexual orientation survey measurement” and “gender identity survey measurement” using online academic research databases. These searches yielded reports, articles, and chapters from academic journals and books, including those in the fields of sociology, psychology, demography, public health, survey methodology, sexuality and gender research, and public opinion research. Papers identified during review of the aforementioned materials were also examined.

To manage the breadth of this review in a way that would yield the most relevant findings for the purposes of this paper, the literature review was limited to publications occurring between January 2011 and July 2016 (exceptions as noted). In addition, the review centrally focused on SOGI measurement research on surveys conducted in the United States, but some international studies were also reviewed. In total, the search yielded approximately 100 references.

Each eligible publication was coded according to dimension of SOGI, question format, testing for measurement error, and findings and recommendations regarding measurement.

IV. Results of Evaluation Studies

This section presents the results of evaluation studies on the measurement of SOGI. First, findings from research on sexual orientation questions, including questions on sexual identity, sexual attraction, and sexual behavior, are discussed. Second, findings from research on gender identity questions, including questions on gender identity and gender expression, are discussed. Lastly, results from studies on same-sex relationship questions are presented. For each set of questions, the questions used in the studies, the sample characteristics of the population studied, the methods of analysis and evaluation in the studies, and the major findings are examined.

a. Sexual Orientation

As noted above, sexual orientation includes the concepts of sexual identity, sexual behavior, and sexual attraction. Researchers often use the term sexual orientation as a synonym for sexual identity (Coffman et al., 2013; Gruskin et al., 2001; VanKim et al., 2010). For example, Vrangalova and Savin-Williams (2012) used the term *sexual orientation identity* in their study although they asked respondents to identify their sexual orientation. Fisher and colleagues (2001) used the term *sexual orientation*, which included sexual identity and sexual behavior. Other researchers simply use the term sexual identity when asking respondents to self-identify their sexual orientation (Gates, 2011; Miller and Ryan, 2011).

Sexual Identity

As noted earlier in this memo, sexual identity is defined as the way someone self-identifies with a given sexual orientation. We found more than 20 quantitative and qualitative studies that attempted to assess the measurement of various sexual identity items (see Appendix A). These studies examined the performance of sexual identity items, how various item responses compared to each other, item nonresponse, and respondents’ reactions to capturing their sexual identity (see Appendix B for sexual identity item wording).

Population Surveyed

Studies measuring sexual identity have been conducted with respondents of widely varying populations. Some researchers select a population of interest based on respondents' *sexual identity* (Austin et al., 2007; Clark et al., 2005; Fisher et al., 2001; NCES). For example, Fisher et al. (2001) focused on lesbian, bisexual, and transgender women in the Washington, DC, area. In cognitive testing for the High School Longitudinal Study of 2009 (HSL:09), the National Center for Education Statistics had 10 of 30 respondents identify as LGBTQ and the remaining respondents as straight or heterosexual. Of the 40 women who participated in Clark and colleagues' (2005) cognitive interviews, 19 were women who partner with women (WPW) and 21 were women who partner with men (WPM).

Some research studies have been conducted with particular *age groups*, such as the adult and adolescent noninstitutionalized U.S. populations (e.g., Dahlhamer et al., 2014; Jans et al., 2015; Martinez et al., 2016; Miller and Ryan, 2001; Poston and Chang, 2015; Saewyc et al., 2004; Cahill et al., 2014; Clark et al., 2005; Coffman et al., 2013; Gruskin et al., 2001; Joloza et al., 2010; Martinez et al., 2016; Ridolfo et al., 2012; Stern et al., 2016; VanKim et al., 2010). For the research reviewed in this paper, the minimum age of respondents was typically 12, 16, or 18 years and the maximum was 65 and older. Research by Grant and Jans (n.d.) discusses differences in responses of respondents who were 70 and older.

Most studies included respondents across *race/ethnic groups*. In general, respondents included persons who were Hispanic or Latino; White; Black or African American; Asian or Asian Pacific Islander; and Native American (Coffman et al., 2013; Fisher et al., 2001; Martinez, et al., 2016; Miller and Ryan, 2011; NCES; Vrangalova and Savin-Williams, 2012). Some studies included both English- and Spanish-speaking respondents (Miller and Ryan, 2001; Stern et al., 2016; VanKim et al., 2010).

Other studies have identified the population of interest based on *occupation or program participation*, such as students (McCabe et al., 2012), Medicare beneficiaries (Stern et al., 2016), and members of health programs (Gruskin et al., 2001). Some recent research has also begun to examine sexual identity in the adult incarcerated population (Glaze, 2015).

Although not the intended focus of the research, populations included in these studies also varied by *educational attainment* of the respondent. For example, Fisher and colleagues' (2001) respondents were highly educated in that more than half had graduate or professional degrees. In the HSL:09 cognitive testing, about 70% were currently or previously enrolled in a postsecondary program (National Center for Education Statistics, forthcoming).

Quantitative Analyses: Concurrent/Congruent Validity

The sexual identity studies reviewed in this paper used a number of quantitative methods to examine measure validity. For item performance, some studies compared concordance on responses to several versions of the question or to other survey findings (Case et al., 2006; Dahlhamer et al., 2014; Saewyc et al., 2004). To look at item response, one study used test-retest reliability and factor analysis to determine if the items used to measure sexual identity conformed to expectations (Mohr and Kendra, 2011). Other studies completed both univariate and multivariate analyses to examine the functioning of items (Coffman et al., 2013; Fisher et al., 2001; Joloza et al., 2010). Overall, based on these quantitative findings, the majority of respondents appear to have no difficulty answering sexual identity items. In

particular, LGBT respondents feel that the questions presented on surveys allow them to accurately describe their sexual orientation (Cahill et al., 2014).

The National Health Interview Survey (NHIS) examined responses to follow-up questions regarding what respondents meant when indicating “something else” or “I don’t know the answer” (Dahlhamer, et al., 2014). The most common responses to the “something else” follow-up question(s) were “You do not use labels to identify yourself” and “You do not think of yourself as having sexuality.” The majority (90%) of adult respondents who initially answered “I don’t know the answer” said that they had not figured out or were in the process of figuring out their sexuality, they did not understand the words, they did not know the answer, or simply refused. When comparing sexual identity estimates before and after back-coding using the follow-up question responses, there were no changes found in the sexual identity distribution. Therefore, the NHIS stopped featuring these follow-up questions in the 2015 survey. Cognitive research by Clark et al. (2005) found that respondents who answered “think of yourself in different terms” used terms including asexual, questioning, and crossdresser to describe themselves (Clark et al. 2005).

When comparing the findings in the NHIS to the National Health and Nutrition Examination Survey (NHANES) and the National Survey of Family Growth (NSFG), differences were found in the measurement of sexual identity: a higher percentage of adults identified as gay or lesbian in the NHIS compared to the NHANES, and a higher percentage of women identified as gay or lesbian in the NHIS compared to the NSFG (Dahlhamer et al., 2014). Additionally, a substantially lower percentage of NHIS adults identified as bisexual compared to adults in the NSFG and NHANES. The various survey features, including question wording, mode of administration, survey content, and placement of the questions, were suggested as possible reasons for these differences (Dahlhamer et al., 2014).

One study used an expanded measure of sexual identity, adding “mostly heterosexual” and “mostly gay/lesbian” as response categories (Vrangalova and Savin-Williams, 2012). This research found that “mostly heterosexual” was the most frequently chosen non-heterosexual label for men and women. Sex differences in the number of women compared to men choosing the “mostly heterosexual” label was highly significant. For the “mostly gay/lesbian” responses, more participants identified as bisexual or gay/lesbian than as mostly gay/lesbian. Overall, this research supported the addition of the two “mostly” categories (Vrangalova and Savin-Williams, 2012).

Some research has examined social desirability in responses to sexual identity. Coffman et al. (2013) argued that the standard methods of eliciting respondents’ sexual orientation underestimates the true fraction of individuals who identify as something other than heterosexual. The authors concluded that stigma is still felt by many in this population, often leading to social desirability bias in responses to sexual identity questions.

Research generally has found congruence between several measures of sexual identity (McCabe et al., 2012). In addition, concordance between reported sexual identity on a pilot study and the cohort study was high (Case et al., 2006). Mohr and Kendra (2011) generally found support for the validity and reliability of Lesbian, Gay, and Bisexual Identity Scale (LGBIS) subscale scores among university students that identified as LGB. This suggests that the LGBIS may be an appropriate means of assessing the different dimensions of LGB persons in this particular population.

Quantitative Analyses: Item Nonresponse

To examine potential item nonresponse bias, several studies used regression analyses by including other demographic characteristics as covariates (Kim and Fredriksen-Goldsen, 2013; Gruskin et al., 2001). Other studies looked at descriptive analyses of item nonresponse (Grant and Jans, n.d.; Ridolfo et al., 2012; VanKim et al., 2010). One study had the capability to compare nonresponse over time (Jans et al., 2015). Overall, nonresponse appears to be relatively low across all studies that have examined sexual identity, varying from less than 1% to just over 6% (Case et al., 2006; Dahlhamer et al., 2014; Grant and Jans, n.d.; Ridolfo et al., 2012; VanKim et al., 2010).

In general, studies that looked at nonresponse counted “not sure,” “don’t know,” and refusals as item nonresponse (Dahlhamer et al., 2014; Kim and Fredriksen-Goldsen, 2013; VanKim et al., 2010). Some researchers also treated “something else” responses as missing for select analyses (Dahlhamer et al., 2014). In some cases, researchers excluded respondents that did not answer all of the related sexual identity items (Mohr and Kendra, 2011).

Although low overall, item nonresponse was found to vary by respondent demographic characteristics. For example, research has found that nonresponse to sexual identity questions tends to increase with age (Gruskin et al., 2001). Focusing on the time spent on the sexual identity question, the NHIS found that there were significantly more adults ages 65 and older in the shortest and longest time groups (based on quintiles), suggesting possible comprehension problems often associated with item nonresponse (Dahlhamer et al., 2014).

Nonresponse also varied by sex of the respondent, although the direction of this finding seemed to vary by age. Item nonresponse has been found to be higher for women compared to men (Grant and Jans, n.d.; Gruskin et al., 2001). In some cases, it is particularly high for women who speak a language other than English (Grant and Jans, n.d.) In contrast, Saewyc et al., (2004) found that boys had higher nonresponse than girls in a survey of students ages 12 to 20.

The majority of studies found that rates of item nonresponse varied by race and ethnicity. In general, non-Hispanic African Americans, Asians, and Hispanics had higher nonresponse rates than non-Hispanic whites (Grant and Jans, n.d.; Gruskin et al., 2001; Jans et al., 2015; Kim and Fredriksen-Goldsen, 2013; Saewyc et al., 2004). This finding also held when other covariates correlated with item nonresponse, such as age, income, education, were examined (Kim and Fredriksen-Goldsen, 2013). Specifically, non-Hispanic African Americans, Asians, and Hispanics were about 2.6, 12.5, and 6.4 times more likely, respectively, than non-Hispanic whites to respond “not sure or don’t know” (Kim and Fredriksen-Goldsen, 2013). Item nonresponse has also been found to be higher among persons who speak languages other than English (Dahlhamer et al., 2014; Grant and Jans, n.d.; Jans et al., 2015; Saewyc et al., 2004). NHIS time analyses have also shown evidence of question shortcutting for adults answering in Spanish—that is, there were more respondents answering in Spanish who were in the shortest-time group to complete the question than respondents answering in English (Dahlhamer et al., 2014).

Nonresponse on sexual identity items also varies by education. Persons with less education (less than a high school degree or no college education) were less likely than those with at least some college education to respond to sexual identity questions (Dahlhamer et al., 2014; Gruskin et al., 2001). NHIS time analyses have also shown evidence of question shortcutting for adults with less education (Dahlhamer et al., 2014).

Item nonresponse may also vary by the question wording. For example, the NSFG modified its sexual identity question to include more commonly used identity categories and dropped the definition that had been included in previous versions. These changes reduced the level of “other” responses from about 4.2% in 2002 to 0.4% in 2006 (Ridolfo et al., 2012).

Context of administration and scope of items included in the survey also appear to affect item nonresponse. Research completed by Saewyc et al., (2004) with students ages 12 to 20 found that for surveys without the sexual behavior questions, the combined self-labeling/attraction questions had the lowest nonresponse, followed by the combined attraction/intentions questions. The self-labeling/attraction item conflates the two concepts, which is likely a contributing factor to the item nonresponse. Lower nonresponse was observed in surveys administered by public health nurses compared to teachers. Items with the shortest stems, fewest response options, and fewest words had the lowest nonresponse rates. Identity and attraction items that followed sexual behavior questions had higher item nonresponse than similar questions that appeared before the behavior items.

Item nonresponse may also vary over time. In the California Health Interview Survey (CHIS), sexual minority self-identification increased over time as item nonresponse declined (Jans et al., 2015). CHIS respondents whose interviews were completed in non-English languages showed the greatest declines over time in item nonresponse for sexual identity (Jans et al., 2015).

Qualitative Analyses: Cognitive Testing

The reviewed studies also assessed items using cognitive interviewing to observe participants’ reactions and responses to measures of sexual identity. These studies completed cognitive interviews with participants of various demographic groups, including adolescents and adults, and LGB and non-LGB persons. (Clark et al., 2005; Ingraham et al., 2015; Martinez et al., 2016; National Center for Education Statistics, forthcoming; Ridolfo et al., 2012; Stern, et al., 2016; National Opinion Research Center (NORC), 2016).

The terminology used in sexual identity questions has been one area of focus in available cognitive interviewing research. Some have found that participants have comprehension difficulty with the term “heterosexual,” and respondents recognize “straight” as the more common term (Miller and Ryan, 2011; National Center for Education Statistics, forthcoming). However, some straight participants felt that the terms “heterosexual” and “homosexual” were more proper and less offensive than “straight,” “gay,” and “lesbian” (National Center for Education Statistics, forthcoming). Similar comprehension issues were found in other cognitive testing. In particular, some participants, especially low-income participants, did not know how to define “heterosexual” (Ridolfo et al., 2012). Further, Hispanic cognitive interview participants with less education (particularly those who completed the survey in Spanish) were less likely to be familiar with the sexual identity terms (Ridolfo et al., 2012). Research has also found that certain participants, particularly those of lower socioeconomic status and Spanish-speakers, had difficulty with the term “bisexual,” confusing it with heterosexual or conceptualizing it as behavior (Miller and Ryan, 2011). This research, however, recommended keeping the term as it is the appropriate term used by bisexual respondents (Miller and Ryan, 2011).

Participant reactions to sexual identity items varied by demographic characteristics, as well. Heterosexual participants tended to define themselves as “not gay,” that is, distancing themselves from the stigmatized gay identity (Ridolfo et al., 2012). Among heterosexual participants, there was a mix of

preference for the terms “straight” and “not gay” (Miller and Ryan, 2011; Ridolfo et al., 2012). Transgender participants who associate with the queer community may have problems associating themselves with the term “heterosexual” even though this may be the appropriate description; for instance, transgender male-to-female may have relationships with men (Ridolfo et al., 2012).

Testing from the CHIS showed a high rate of misclassification of adults ages 70 and older into the bisexual category. Because of these findings, CHIS limited the sample universe to persons age 70 and younger (Grant and Jans, n.d.). Additional testing of sexual identity items among an older population of adults revealed that the items tested well for English-speaking adults (Stern et al., 2016). Sexual identity items tested in the context of a victimization survey among persons ages 16 and older performed well for both adolescents and adults (Martinez et al., 2016). Other testing among adolescents found that many respondents felt that sexual identity was difficult to answer, in some cases because it implied a degree of permanence with which they were not comfortable (Austin et al., 2007; Saewyc, 2011).

Ridolfo et al. (2012) identified a complex relationship between gender and sexual identity specifically for transgender persons. That is, sexual identity may be difficult to answer for persons who are in the process of coming out or transitioning genders. These respondents most often chose “other” or “don’t know” responses in lieu of the other options provided. In other cases, transgender respondents chose “heterosexual” because they had fully re-conceptualized their gender identity and were in a relationship with partners of their previous genders. These cognitive interviews illustrated that transgender respondents rejecting traditional terms used to describe sexual identities were more likely to experience difficulty in responding.

Research on sexual identity measurement also revealed issues related to language. The sexual identity measure for the NSFG was cognitively tested and showed using terms such as “straight” and “gay” produced better comprehension in English administration; however, there is no translation for “straight” in the Spanish language (Ridolfo et al., 2012). Some have concluded that the presence of the word “heterosexual” on the Spanish language question helps respondents make sense of other response categories (Miller and Ryan, 2011). And in the same manner, the absence of the word “heterosexual” on the English-language question is helpful to reduce response difficulty (Miller and Ryan, 2011). Cognitive interviewing of an older adult population revealed some comprehension problems among Spanish-speaking respondents and recommended changing the order of the sexual identity response category from “Heterosexual, o sea, no gay” to “No gay, o sea, heterosexual” (Stern et al., 2016). The reasoning for this change was to improve Spanish-speaking respondents’ ability to properly report their sexual identity by presenting the “not gay” explanation first (Stern et al., 2016). This was further tested with Spanish-speaking adults, and the recommended change tested well (NORC, 2016).

Glaze (2015) examined the functioning of a sexual identity item used in the Survey of Prison Inmates (SPI). Overall, interviewers did not report negative or unusual reactions to the question. Consistent with other research, some respondents had trouble with the term “heterosexual,” particularly Spanish-speaking inmates, but after the question was reread they were able to answer appropriately (L. Glaze, BJS and T. Smith, RTI International, personal communication, February 22, 2016). In doing so, some respondents strongly asserted that they were straight. There were no differences observed between Spanish-speaking and English-speaking respondents’ reactions (L. Glaze, BJS and T. Smith, RTI International, personal communication, February 22, 2016).

Future Research

Researchers evaluating sexual identity measures often note directions for future research to improve sexual identity measurement. For example, with regard to measuring health outcomes related to sexual identity, Gruskin and colleagues (2001) note that their study may not have completely captured respondents' experiences because they only included one question on sexual orientation, compared to asking multiple questions on sexual orientation and sexual behavior. Saewyc and colleagues (2004) suggest that future research focus on additional measures of behavior and either self-labeling or sexual attraction when trying to effectively measure health outcomes related to sexual identity.

Several researchers recommend further research on the effect of context in which sexual identity questions are administered. Specifically, an important component of survey methodology and interviewing is interviewer training. Research suggests that data quality can be improved by focusing interviewer training on the implications of invalid data and asking questions as worded, especially questions on sensitive topics (Grant and Jans, n.d.; Ridolfo et al., 2012). Furthermore, sensitive questions, like questions on sexual identity, may need to be asked in private areas so the respondent is comfortable. Research related to privacy and discreteness when administering sexual identity questions is needed (Grant and Jans, n.d.; Ridolfo et al., 2012). Other areas of future research related to survey administration of sexual identity questions include the consistent administration of questions and issues related to administering these questions via different survey modes (Grant and Jans, n.d.). Further, Ridolfo and colleagues (2012) suggest putting the sexual identity items near demographic questions rather than around sexual attraction and behavior questions; not doing so may lead to biased responses (Ridolfo et al., 2012).

Other researchers have urged exploring innovative ways to measure sexual orientation that would address alternative sexual identity terms that may be more relevant to other racial and ethnic groups (e.g., same-gender loving, down-low) (Sexual Minority Assessment Research Team, 2009). This research is important to fully measure sexual identity in the national population (Kim and Fredriksen-Goldsen, 2013). Saewyc et al. (2004) also suggest a closer examination of adolescents' understanding of sexual identity questions.

Sexual Attraction

As defined earlier, *sexual attraction* refers to the sex or gender to which someone feels attraction (e.g., whether an individual is attracted to males/men, females/women, or both). Four studies were identified that evaluated questions on sexual attraction. Two of these studies used cognitive interviewing techniques, while the other two involved quantitative evaluations of survey questions. The target populations differed considerably across the four studies, including women ages 40 and older, adolescents between the ages of 12 and 21, and adults of all ages. Appendix C lists the questions tested in the cited studies.

Although they found differences for sexual identity and behavior, Coffman et al. (2013) found no significant difference in the reporting of same-sex attraction when using a method that was proven to reduce social desirability bias.⁷ This may indicate that current best practices that ensure privacy and

⁷ Coffman and colleagues utilized a variation of "item count" technique. The authors randomized respondents into two groups. The control group was given a baseline set of four innocuous questions, told to indicate the total number of "yes" responses, and then immediately asked a sensitive question. The other group was told to indicate

confidentiality, in this case the use of a self-administered online survey, may be sufficient for collecting information about sexual attraction. However, the authors note that incorrect interpretation of the question resulted in increased reports of both non-heterosexual identity and same-sex sexual behavior, and may be related to the experimental method used. Further research is needed.

Based on cognitive interviews of women ages 40 and older, Clark et al. (2005) evaluated the following question on sexual attraction: “People differ in their physical attraction to other people. Which statement best describes your feelings? Would you say ... only attracted to women, mostly attracted to women, equally attracted to women and men, mostly attracted to men, only attracted to men, not sure, none of the above statements describe your feelings.” The primary response problem to emerge from an initial round of interviewing was the request for clarification of the type of attraction being asked about in the question. A revised question was tested that added the term “physically” to the first five response options (see Appendix C). Using the revised question, when participants were asked to define physical attraction, answers included sexual, chemical, and erotic examples. Although 80% of the participants provided a definition of attraction that included sexual intimacy, other factors were mentioned including affection, affiliation, and emotional preference. These constitute important alternative interpretations of physical attraction among middle-aged and older women.

Austin and colleagues (2007) performed cognitive interviews with 30 adolescents between the ages of 15 and 21. In addition to questions on sexual identity and behavior, the study authors evaluated a combined identity-attraction question and an attraction-only question (see Appendix C). The results were mixed for the combined question, with some of the participants noting the intermediate response options between “completely heterosexual” and “bisexual” and between “bisexual” and “completely homosexual” as positive features of the question. However, several sexual minority participants found the identity terms and parenthetical statements of attraction to be confusing, conflating two different experiences. For example, some youth noted the dilemma presented by the category “completely homosexual” with its parenthetical attraction statement of “gay/lesbian, attracted to persons of the same sex.” Some of these participants stated that they describe themselves as gay or lesbian, but also feel some attraction to members of the opposite sex. The combined question also presented problems for transgender youth who weren’t sure if they should answer based on their biological sex or their gender identity. Some transgender youth answered “not sure” or “bisexual” as a way to avoid choosing between the two.

The sexual attraction-only question (second item listed in Appendix C) appeared to work well with the participants, with all but two youth describing the item as specific, clear, and easy to answer. Consistent with the Clark et al. (2005) study, the two dissenting youth thought the focus on “sexual” attraction to be too restrictive, and reported other types of attraction including romantic-emotional attraction. For most of the youth, however, this narrow definition of attraction was regarded as a positive aspect of the item. Other positive aspects included self-explanatory response options and the less-threatening nature of the question, relative to the sexual identity item. Some of the participants thought that young adolescents in particular would find this question easier to answer since they are still in the process of “figuring out” their sexual identity.

the total number of the “yes” responses to a set of five questions that included the original four questions given to the control group *plus* the sensitive question. The true value of the sensitive question was deduced by comparing the responses between groups.

In a quantitative study of survey responses, Saewyc and colleagues (2004) analyzed five questions combining elements of sexual identity and sexual attraction and one question combining elements of sexual attraction and sexual intentions. As shown in Appendix C, three of the five sexual identity/attraction questions shared the same response options but differed in their question stems. The other two identity/attraction questions also shared response options but used different question stems. One or more of these questions was asked in each of eight adolescent health surveys with varying sampling designs and demographic compositions. All eight surveys were administered in schools. Overall, the surveys covered students in grades 7-12, with the ages of participants ranging from 12 to 20.

The authors compared item nonresponse rates to the questions across surveys as the primary performance measure. Overall, the identity/attraction items produced item nonresponse rates in the range of 0.6% to 16.2% for females and 0.7% to 18.2% for males. These large variations in item nonresponse rates were also observed across years and surveys for identically-worded questions, suggesting causal factors other than question wording. For example, for surveys without sexual behavior questions, the combined identity/attraction questions produced the lowest nonresponse, followed by the combined attraction/intention question. However, when the identity and attraction items followed sexual behavior questions they tended to produce higher item nonresponse than similar questions that appeared before the behavior items, highlighting the importance of question context and location as factors affecting performance. While all surveys were administered in classrooms via paper and pencil, some were administered by teachers, some by research staff, and some by public health nurses. Interestingly, lower item nonresponse was observed in surveys administered by the public health nurses. The authors also noted that the conflation of attraction and identity in these questions likely contributed to item nonresponse, although no questions that tapped identity alone or attraction alone were available for comparison.

Only two of the eight surveys included an attraction/intention question, with each survey using a slightly different question stem from the other (but same response options). The item nonresponse rates were vastly different between the two surveys, with one producing rates of 25.5% (females) and 23.2% (males), and the other producing rates of 6.0% (females) and 6.8% (males). The authors noted that these questions were confusing and likely produced comprehension problems since they ask about both current feelings and future behaviors.

Across all seven identity/attraction and attraction/intention questions, higher item nonresponse was observed for boys, younger students, students with a learning disability, students held back a grade, students who reported speaking a language other than English at home, and African-American and Southeast Asian students. Finally, students who were "not sure" of their orientation were more likely to pick "not sure" to other questions in the survey.

Saewyc and colleagues strongly suggested that more testing is needed to understand how adolescents comprehend measures of sexual orientation. As for recommendations, they suggested that "[t]o effectively measure the highest priority health concerns, the best choice would include behavior plus either self-labeling or attraction" (345.e12).

Sexual Behavior

As mentioned above, sexual behavior refers to the sex of a person's sexual partners (e.g., individuals of the same sex, different sex, or both sexes). We identified four quantitative and qualitative studies that attempted to assess the measurement of sexual behavior. These studies looked at how sexual behavior items performed, item nonresponse, and respondents' reactions to capturing their sexual behavior. In general, researchers assess sexual behavior by asking respondents to report the sex of their sexual partners, during a specified period of time in the past and/or during their lifetime. Notably, only one study focused solely on sexual behavior (Berg and Lien, 2006). The other studies simultaneously measured two or three dimensions of sexual orientation (behavior and identity, or behavior, identity, and attraction) (Case et al., 2006; Poston and Chang, 2015; Vrangalova and Savin-Williams, 2012). A discussion of measures of sexual behavior in conjunction with identity and attraction can be found in the section entitled *Congruency of Multiple Measures of Sexual Orientation*. The current section discusses findings related to sexual behavior. Appendix D lists the questions tested in the cited studies.

Populations Surveyed

The populations included in studies that evaluate sexual behavior questions vary by demographic characteristics of respondents. These demographic characteristics include sexual identity, sexual attraction, sex, income, race, and marital status. Demographic characteristics also varied by educational attainment, age, city size, and language.

Qualitative Analyses of Sexual Behavior Items

Case et al. (2006) piloted a sexual identity question (referred to as "identity or orientation") and a sexual behavior question on a sample drawn of never married, previously married, and currently married female participants of the 1993 Nurses' Health Study II (NHSII). Pilot participants were asked to write in their reactions to the sexual identity and behavior questions. Regarding the sexual behavior item (see Appendix D), some participants focused on the question wording, correcting grammar (suggesting "with whom you have had sex" vs. "who you have had sex with") and indicating confusion with the use of the phrases "man/men" and "woman/women." Several participants also noted that the term "adult lifetime" did not include teenage years. Some participants felt the item was too personal, but others were excited to see sexual behavior being included as an aspect of women's health.

Quantitative Analyses: Item Nonresponse and Misreporting

Berg and Lien (2006) developed a probability model that simultaneously dealt with misreporting and item nonresponse using a self-reported question on sexual behavior in the General Social Survey (GSS). The goal of the model was to estimate the frequency of same-sex sexual behavior in the presence of non-ignorable item nonresponse and misreporting. The researchers included income, race/ethnicity, marital status, parental status, education, age, and city size as covariates in the model, and they performed separate analyses by sex. Among women, Berg and Lien found that item nonresponse decreases as income increases, increases as age increases, and increases as education level increases. Among men, they found item nonresponse decreases as city size increases.

Misreporting, a type of measurement error, occurred in the study when non-heterosexuals reported their behavior as exclusively heterosexual and when heterosexuals did not report same-sex sexual behavior. The authors found that among men, misreporting increased as income increased and

misreporting decreased as age increased. They also found that white men had a higher probability of misreporting compared to men of all other races. Among men and women, misreporting decreased as city size increased.

Congruency of Multiple Measures of Sexual Orientation

Some studies that were reviewed examined multiple dimensions of sexual orientation (sexual identity, sexual attraction, and sexual behavior), and whether respondents answered consistently among these dimensions. In general, research finds that a person's sexual identity is not necessarily congruent with their sexual attraction or behaviors (Fisher et al., 2001; Poston and Chang, 2015; Ridolfo et al., 2012; Saewyc, 2011; Saewyc et al., 2004; Vrangalova and Savin-Williams, 2012). Some researchers have concluded that sexual orientation may be viewed on more of a continuum of sexuality (Poston and Chang, 2015; Vrangalova and Savin-Williams, 2012).

Research among adolescents and adults has found that persons who reported being heterosexual or gay or lesbian, did not necessarily also report consistent responses to sexual attraction or behavior items (Fisher et al., 2001; Poston and Chang, 2015; Saewyc et al., 2004; Ridolfo et al., 2012; Vrangalova and Savin-Williams, 2012). That is, some persons who identify as gay or lesbian also report opposite-sex attraction or behaviors; and some persons who identify as heterosexual report same-sex attraction or behaviors. Using an expanded categorical measure of sexual identity that included "mostly gay/lesbian" and "mostly heterosexual," Vrangalova and Savin-Williams (2012) found that respondents who reported these non-exclusive responses also reported non-exclusive responses to other sexual orientation components, particularly in terms of attraction. This same research found that males typically reported more consistent responses to their sexual orientation than females.

Given incongruent responses to questions tapping the three dimensions of sexual orientation (i.e., identity, behavior, and attraction), the appropriate measure depends on the primary objective of the research (Saewyc, 2011; Vrangalova and Savin-Williams, 2012). For example, some have suggested that health researchers or educators may want to focus more on sexual behavior, while still including measures of identity and attraction, in order to gain appropriate information from respondents (Fisher et al., 2001; Saewyc et al., 2004). It is also important to note that prevalence rates of homosexuality, bisexuality, and heterosexuality may vary by definition and are found to be highest when a more inclusive definition is used (Poston and Chang, 2015).

b. Gender Identity

In its 2011 report, *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*, the Institute of Medicine (IOM) recommended the expansion of research methods yielding data to inform decisions about LGBT health. Although advances have been made in the measurement of sexual identity since the report's release, much of the literature on transgender individuals has relied on data from convenience samples in limited geographic areas and may not be generalizable to the overall transgender population. Initial research has shown that transgender individuals are at greater risk of discrimination in housing, employment, and health (Center for American Progress, 2015). Other research points to a higher likelihood of binge drinking (Scheim, Bauer, and Shokoohi, 2016), higher likelihood of suicide attempt (Grant et al., 2010) and higher levels of sexual victimization in prisons and jails (Beck et al., 2014).

Yet, at the time of this working paper's release, few Federal surveys have included measures of gender identity. In part, this is because measures of gender identity have not been tested by Federal agencies until just recently. While it is clear that the measurement of gender identity is important, researchers face a choice between different measures. Without an evaluation of measurement functioning, Federal agencies may be reluctant to collect these data for fear that measurement problems would affect the quality of the statistics produced (and used to inform policy making). Another challenge is that, where gender identity measures are included, different measures are used in different surveys. As with other measures (e.g., functional disability), the use of different measures can make it difficult to compare estimates across surveys. The purpose of this section is to summarize recent evaluations of gender identity measures in the U.S.

Gender Identity Measures

At their most basic level, gender identity measures can be broken into two types: single-item and two-step measures.

Single-item measures, as the name implies, use a single item to capture gender identity. In their 2014 report, *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys*, the Gender Identity in the U.S. Surveillance (GenIUSS) Group recommended using single-item measures "when valid, self-report measures of assigned sex at birth and current gender identity are not on a survey and cannot be added (or replace existing measures)" (GenIUSS Group, 2014).

A number of Federal surveys utilize a single-item measure. The Behavioral Risk Factor Surveillance System (BRFSS) introduced an optional single gender identity item in 2013 (see Appendix F). Baker and Hughes (2016) reported that states utilizing the BRFSS sexual and gender identity (SOGI) module reported that it is well-received by respondents and does not negatively affect the quality of data collected. The authors also said that the Missouri state BRFSS coordinator reported no survey break-offs and minimal item nonresponse.

The National Inmate Survey (NIS) also uses a single-item approach to gender identity. While groundbreaking in its ability to estimate the transgender-identified inmate population in U.S. prisons and jails, the NIS "choose one response" approach compels inmates who identify as both "male" or "female" and "transgender" to decide between both identities. Another challenge with the NIS measures, which also affects the BRFSS measure, is related to respondents who do not identify as transgender but whose gender identities are incongruent with their natal sex. If these respondents choose "male" or "female," the BRFSS and NIS may undercount the transgender population.

The Population Assessment of Tobacco and Health (PATH) uses a variation of the single-item approach. After a definition of transgender is read to respondents, those who respond that they are transgender are asked a follow-up item to further clarify their gender identity. This item, however, may suffer from the same analytic challenges as the BRFSS and NIS items described above.

Two-item measures generally capture assigned sex at birth (i.e., natal sex) and current gender identity. The GenIUSS Group recommends utilizing two step measures, "when two demographic items can be added to an adult survey (or, in most instances, a standing measure of sex replaced and a measure of current gender identity added)." Two-step measures are also recommended for use by the Center of

Excellence for Transgender Health (CoE)⁸ and the World Professional Association for Transgender Health (WPATH; Deutsch et al., 2013). The Federal surveys utilizing two-step measures are the National Crime Victimization Survey (NCVS), Survey of Prison Inmates (SPI), and the National Adult Tobacco Survey (see Appendix F).

Although there has been limited testing on the single-item method, there have been a number of tests of the two-part method. The challenge with these methods is that the language of these items varies across tests (see Appendix E for the language used in the studies cited in this section), making the comparison of results difficult. Regardless, each study advances the understanding of how respondents understand and react to questions about both sex and gender.

Lombardi and Banik (2016) recruited a convenience sample of transgender and cisgender people in Ohio (n=50) to examine how they understood and answered a two-step gender identity item. Both cisgender and transgender participants were able to answer the question about natal sex, referring to their physiology (e.g., genitalia) or birth certificate.⁹ Cisgender participants were split about whether sex was inherently biological/physiological or assigned by someone like a physician, while approximately two-thirds of transgender participants described sex as being assigned by a health professional or someone else. Cisgender participants were able to answer the “sex or gender” question without hesitation. About half felt that sex and gender measured the same concepts, and all of them responded in the same manner as they responded to natal sex. All but two of the transgender participants chose a gender that differed from their natal sex, and all but one saw sex and gender as measuring different concepts. There was, however, a preference among transgender participants to drop “sex” from “sex or gender.” Based on these results, the authors felt the two-step measure could be used to distinguish between the transgender and cisgender populations.

In a study examining the collection of SOGI at medical intake, Cahill et al. (2014) tested a version of the two-step gender identity items on a diverse group of 301 patients in four community health centers. The researchers found that study participants understood and were willing to answer the tested two-step items. In fact, 84% of participants strongly agreed or somewhat agreed that they would answer the question on natal sex, and 86% of participants strongly agreed or somewhat agreed that they would answer the gender identity question. More than three-quarters of participants strongly agreed that they understood the gender identity item response options. Compared to younger participants, participants ages 65 and older were less likely to understand what the questions were asking or the response options. But their mean responses were in the neutral range of the scale indicating an average level of understanding.

Reisner et al. (2014) studied the validity of natal sex and gender identity in youth. Their study found that, compared to cisgender participants, gender minority participants scored higher on both recalled childhood gender nonconformity measures (criterion-related validity) and current socially assigned gender nonconformity measures (construct validity). Both gender minority participants and cisgender participants indicated that the natal sex question was clear and easy to answer. Both groups mentioned that “physical anatomy” was used to assign sex at birth. All participants felt the response options of “male” and “female” were appropriate, including a single intersex participant who acknowledged that

⁸ See <http://transhealth.ucsf.edu/trans?page=lib-data-collection>.

⁹ It is worth noting that the tested item did not refer to birth certificate.

these are the only options available on birth certificates in the U.S. More than three-quarters of participants understood gender to be an internal concept that could differ from one's natal sex. As with natal sex, all participants felt that the gender identity question was straightforward, but gender minority participants did express difficulty in choosing the most appropriate response option. Some expressed switching their responses given that they were only allowed to select one gender. The majority of cisgender participants were not sure what was meant by the "do not identify" response option. That being said, they were still able to respond accurately, and two participants who identified as gender non-conforming or gender variant selected the "do not identify" response option. More than three-quarters of participants liked the response options offered in the gender identity questions, but one who identified as genderqueer suggested a write-in response option.

In a cognitive test of English-speaking, cisgender individuals (ages 16 and older), Martinez et al. (2016), examined how direct the natal sex item should be. The authors found that the majority of study participants preferred a direct natal sex question (i.e., "On your original birth certificate, was your sex assigned as male or female?") over a less direct, but more common, question (i.e., "What sex were you assigned at birth, on your original birth certificate?") Regardless of preference, however, almost all respondents (89%) said they understood both questions equally well. The researchers also examined whether participants preferred gender labels (i.e., "Do you currently describe yourself as a man, woman, or transgender person?") over sex labels (i.e., "Do you currently describe yourself as male, female, or transgender?") The majority of study participants preferred sex labels (i.e., male, female) to gender (i.e., man, woman) labels, but the majority of respondents said that both questions were understood equally well. That said, it was pointed out that cisgender youth may not consider themselves to be old enough to be "man" or "woman."

Glaze (2015) tested the two-part gender identity item while testing the functionality of the SPI instrument with a group of cisgender prison inmates. Unlike prior studies, the two-step measure was implemented without specific follow-up, but the reactions of the inmates were observed. The research found that participants did not react negatively to the items and were able to provide their responses.

The studies discussed so far have only looked at comprehension of the two-step gender identity method in English. Stern et al. (2016) examined comprehension of these items in the English- and Spanish-speaking Medicare disabled (ages 18-64) and aged (ages 65 and older) populations. The authors found that English speakers had no problems comprehending the item related to natal sex. When asked about the gender identity item, the majority of English-speaking cisgender participants associated the term "transgender" with the physical component of transitioning. Some English-speaking respondents conflated sexual identity (i.e., being straight, bisexual, lesbian, or gay) with gender identity, but all of them responded to the gender identity item in a manner that was consistent with a gender item collected during the study screener.

Spanish-speaking study participants were able to answer the item related to natal sex, though some responded with their gender expression (e.g., "masculine" or "feminine") rather than their sex (Stern et al., 2016). The majority of Spanish-speaking participants conflated sexual identity and gender identity, and some expressed discomfort in being asked about their gender identity. In addition, a number of Spanish-speaking cisgender participants expressed that their gender identity was "normal" or "non-deviant." Even with the issues the researchers identified in the Spanish-speaking population, all of them

responded to the gender identity item in a manner that was consistent with a gender item collected during the study screener.

Future Federal Research

There are a number of areas in which Federal research could add to the literature on how best to measure gender identity. First, little is known about how question order in the two-step measure may affect response. Saperstein and Westbrook (2014) used a randomized design in a large Web panel and found no evidence of an order effect. However, Lombardi and Banik (2016) found that most transgender participants in their study preferred current gender identity be asked prior to natal sex. The GenIUSS Group recommends asking natal sex first, while WPATH and CoE recommend asking current gender identity first. More research is needed.

Another unresolved issue in the literature is whether to include a reference to one's original birth certificate in the natal sex item. While there have been no studies focusing on this specific issue, there is anecdotal evidence that transgender individuals preferred its exclusion. The GenIUSS group recommended that the two-step approach includes a birth certificate reference, while CoE's most recent recommendation excludes it. Again, more research is needed.

The literature reviewed above points to a widespread ability to respond to the two-step measure in varied populations, including elders, prison inmates, and Spanish speakers. However, current research would also benefit from the inclusion of racial, ethnic, and cultural minorities, as well as language minorities other than Spanish.¹⁰ Current gender identity research has focused mostly on comprehension in the English- and, more recently, Spanish-speaking populations. Future research, specifically Federal research, may benefit from increased recruitment of transgender individuals, people of color, and people from cultures in which a third gender exists (e.g., "two-spirit" in American Indian tribes and "hijras" in South Asian cultures).

Finally, recent variations of the two-step measure have included a verification item when natal sex and current gender identity do not agree. These verification measures are utilized in the NCVS and the CHIS as questions asked directly of respondents and in the SPI as an interviewer instruction. They were also used in the measures tested by Lombardi and Banik (2016). While items like these have been shown to reduce measurement error in the measurement of same-sex households (O'Connell and Feliz, 2011), there are no studies examining their application to the measurement of gender identity.

Additional Challenges Related to Measuring Gender Identity

In addition to how best to *ask* gender identity, there are other methodological challenges. For example, little is known on how best to ask gendered questions, such as questions about mammograms and birth control, once a respondent is identified as transgender. Deutsch et al., (2013) recommends collecting an organ inventory in a medical setting to ensure proper gendering of these type of questions. Instead of an organ inventory, it may be possible to use introductory language (e.g., "The next question was designed for women who are not transgender, so let me know if it does not apply to you"), but research

¹⁰ Ryan (2013) reports that in 2011, 20.8% of the U.S. population spoke a language other than English at home, of whom 22.4% spoke English either "not well" or "not at all." Spanish speakers comprise 62.0% of those who spoke a language other than English at home (25.9% of whom spoke English "not well" or "not at all"), followed by Chinese speakers (4.8%, 29.6% of whom spoke English "not well" or "not at all").

is needed on the most effective language. Neither of these strategies has been applied in a Federal survey setting.

Deutsch and colleagues (2013) also recommend the collection of name and pronoun preferences in a medical setting, but this may be just as important in a survey setting. This information can be used to properly tailor mailing materials to transgender participant preferences (e.g., using the proper salutation in a survey mailing) (Ingraham et al., 2015). This information can also be used to properly gender transgender respondents in a proxy survey.

Summary

Since the 2011 IOM report, research on transgender measurement has increased, but there is still no consensus on how to collect gender identity in surveys. The majority of research supports using a two-step method, per the recommendation of the GenIUSS Group, WPATH, and CoE, but there are a number of open questions, including how to order the items, whether to include reference to a birth certificate, and how the items perform in languages other than English and Spanish. Also unresolved is the appropriateness of using gender identity in realms that typically utilize binary measures of sex, such as gendered questions. With a growing number of Federal agencies interested in measuring gender identity and engaging in evaluation of these measures, the Federal government can play an important role in advancing the literature.

Gender Expression

Gender expression is defined as the way one sees themselves or their outward display of gender. Research on gender expression is limited. We identified only two studies that examined how respondents perceive their masculinity and femininity and also how they think others perceive them (Correll et al., 2014; Magliozzi et al., 2016). Both studies were based on surveys administered using Amazon Mechanical Turk (MTurk). Respondents were required to be age 18 or older and U.S. residents. The final number of respondents in both surveys was approximately 1,520. Gender expression was measured through the use of first-order and third-order scales. Appendix E presents the wording and response options for the scales included in Correll et al. (2014) and Magliozzi et al. (2016).

For the first-order and third-order scales, respondents were presented with two separate scales – one for masculinity and one for femininity, for a total of four scales (Correll et al., 2014; Magliozzi et al., 2016). Seven response categories were given for each scale, ranging from “not at all” to “very.” Correll and colleagues (2014) assigned numbers to the poles, categories 1 (“not at all”) and 7 (“very”). Categories 2 through 6 were only given numerical values and were not assigned nominal values in order to avoid any negative connotations if respondents didn’t categorize themselves on either pole of the scale. Response categories were the same for the first-order and third-order scales.

Magliozzi and colleagues (2016) presented response options slightly different than Correll and colleagues (2014). The nominal categories given to pole responses were the same in both studies, “not at all” and “very.” However, the pole responses were not given a numerical value to go along with the nominal response. Numerical values 1 through 5 were given to response options between the poles. A value of 1 was closest to the response “not at all” and a value of 5 was the category closest to “very” (Magliozzi et al., 2016). Response categories were the same for the first-order and third-order scales.

Correll and colleagues (2014) recommend that future researchers keep the masculine and feminine scales separate and administer the four separate scales to respondents. Magliozzi and colleagues (2016) echo these recommendations, noting that “femininity and masculinity scales will allow researchers to account for variations within gender categories...” (p. 7). Therefore, based on the current research, there appears to be value and utility in providing separate scales to respondents. Correll and colleagues (2014) also recommend asking the first-order scales if researchers have to choose one set of scales. The authors mention that the first-order scales offer “the clearest counterpoint to the interviewer’s classification of the respondent’s sex/gender that occurs at the beginning of the GSS” (Correll et al., 2014).

Given the paucity of studies evaluating gender expression items, additional research addressing this topic would prove beneficial and further develop knowledge on this concept central to SOGI research.

c. Household Relationships

As noted in the previous working paper, *Current Measures of SOGI in Federal Household Surveys*, several Federal surveys collect household relationship data for all individuals living in the same housing unit. This information, in conjunction with sex, can be used to produce indirect estimates of the SGM population through the measurement of same-sex couples. For example, in a two-person household if person two is said to be the husband, wife, or spouse of person one and both persons were reported to be of the same sex, these individuals would be coded as a same-sex married couple. Relying on entries to separate questions to derive this status, however, increases the potential for erroneous categorization due to keying errors in computer-assisted surveys or ambiguous markings on mail surveys to one or both of these questions. Not surprisingly, past research has identified sex-relationship inconsistencies that have led to the overestimation of the number of same-sex couples (O’Connell and Feliz, 2011; DeMaio et al., 2013; Kreider and Lofquist, 2014; Lewis et al., 2015). This is usually attributed to accidental mismarking on sex by a very small proportion of opposite-sex married couples. As opposite-sex couples are substantially larger in number than same-sex couples, this often results in a large over-count of same-sex couples (Lewis et al., 2015). We identified six (6) quantitative and qualitative studies that attempted to quantify these inconsistencies and over-reports, and/or identify question design changes to improve the performance of sex and relationship questions for estimating the number of same-sex couples.

Using a statistical names directory in which first names are assigned a probability of “maleness,” DeMaio et al. (2013) estimated that 28.3% of reported same-sex couple households in the 2010 Census were likely opposite-sex. In addition, the majority of sex-relationship inconsistencies occurred in the enumerator-administered nonresponse follow-up (NRFU) form. The matrix format used with the NRFU, as opposed to the sequential format used with the standard Census form, appeared to be the primary cause of data-reporting errors with opposite-sex couples. Data from the 2010 American Community Survey (ACS), using the same sex and relationship questions, revealed fewer reports of same-sex couples. In this case, the presence of automated edits in the telephone and personal-visit data collection phases of the ACS, where inconsistencies in sex-relationship reporting prompted interviewers to verify entered responses (see Appendix H, Figure 1 for an example), appeared to greatly reduce same-sex couple over-reports (O’Connell and Feliz, 2011).

Kreider and Lofquist (2014) corroborated the findings of DeMaio et al. (2013) via research in which couples’ sex as reported on the 2010 Census or 2010 ACS was matched to sex on file with the Social

Security Administration (SSA). Over 57% of couples reported as same-sex married in the ACS and nearly 73% of these couples in the 2010 Census were found to be opposite-sex in SSA data. The concomitant figures for same-sex unmarried couples were around 7% for both the Census and ACS. In contrast, less than 1.0% of reported opposite-sex married couples in both the ACS and decennial Census, and 1.6% of opposite-sex unmarried couples in the ACS and 0.8% of these couples in the 2010 Census appeared to be same-sex in the SSA data. Again, the lower rate of inconsistencies in the ACS data were attributed to the automated edits in the telephone and personal-visit instruments.

Cognitive interviewing and focus group testing of the relationship question used in the 2010 Census and the 2010 ACS (see Appendix H, Figure 2) showed that both same-sex and opposite-sex couples interpreted the relationship question to be asking about legal status (Bates et al., 2010). Same-sex couples who were legally married classified themselves as married couples, while unmarried, same-sex couples classified themselves accordingly. However, based on these studies, a set of changes were recommended to the existing relationship question, including 1) the addition of “spouse” to the “husband/wife” category; 2) moving the “unmarried partner” category up from nearly last to second on the list of options; and 3) adding separate categories for “opposite-sex husband/wife/spouse,” “opposite-sex unmarried partner,” “same-sex husband/wife/spouse,” and “same-sex unmarried partner.” Cognitive testing of these revisions uncovered few response problems (DeMaio and Bates, 2012; Smirnova and Scanlon, 2013).

The revised relationship question that emerged from these testing efforts (see Appendix H, Figure 3) was added to the 2013 American Housing Survey (AHS) as part of an interviewer-administered, split-panel test. Using the experimental AHS data, Lewis et al. (2015) explored inconsistencies in sex and relationship reporting. Like the DeMaio et al. (2013) study, the research used probable sex based on the Census Bureau’s names directory to assess where misreporting occurred. While the new relationship question, along with built-in edits, appeared to reduce the level of inconsistencies and, therefore, misreporting, measurement error remained. For example, of 82 couples reported to be same-sex married at the relationship question, yet opposite sex at the sex question, all 82 were found to be opposite-sex couples. Hence, relationship-sex inconsistency appeared to be driven by errors at the relationship question. Results were largely similar for couples reported to be same-sex unmarried at the relationship question. Of the 21 couples with information on probable sex, 23.8% were likely same-sex, with the remaining 76.2% judged to be opposite-sex. Shifting to couples reported to be opposite-sex unmarried at the relationship question (52 with probable sex information), 72.2% appeared to be same-sex couples, with the remaining 27.8% judged to be opposite-sex. As observed with reported same-sex couples, the bulk of misreporting took place at the relationship item. Finally, of the couples reported to be opposite-sex married at the relationship question, 14.2% were judged to be same-sex couples, with the remaining 85.8% of couples deemed to be opposite-sex. In contrast to the other couple types, misreports at the sex question were the far more prevalent problem for these couples. In sum, for three of the four couple types, misreporting at the relationship question was more prevalent. However, errors at the sex question were more prevalent for reported opposite-sex married couples, and since these couples are the most prevalent couple type, errors at the sex question explained relationship-sex inconsistency for a substantial portion of couples overall (Lewis et al., 2015).

The six reviewed studies identified persistent sex-relationship to householder inconsistencies, leading to overestimates of same-sex households in the decennial Census, ACS, and other Federal surveys. The overestimates are largely attributed to accidental mismarking on sex by a very small proportion of opposite-sex married couples. While many of the testing efforts described above recommend the use of a revised relationship to householder question (see Figure 5), the evidence to date suggests that inconsistencies and overestimates of same-sex couple households are still likely. While the revised design shows promise, the current ACS, for example, continues to use the relationship question fielded with the 2010 ACS.

A number of the reviewed studies mentioned the automated edits in the interviewer-administered phases of the ACS as an explanation for the lower sex-relationship inconsistency rates compared to the 2010 Census. However, these studies failed to quantify the extent to which these edits led to changes in reporting of sex, relationships, or both. We know very little about the behavior of interviewers when confronted with these edits. Do they verify the entered information with respondents? How often do they suppress the edit to keep the interview moving? Are the error messages intuitive and easy to navigate?

Finally, while identifying and addressing the issues behind the documented inconsistencies is important for producing accurate estimates of same-sex couple households (and a number of other couple types), relying on sex and relationship information does not provide a direct measure of sexual orientation. Furthermore, it yields an incomplete estimate, since persons not in a relationship or not living in the same household as their partner will not be identified (MRFHS 2014).

V. Discussion: Best Practices and Areas for Further Research

This working paper reviewed recent studies evaluating the measurement of SOGI in surveys. With the goal of identifying best practices for collecting SOGI data, this paper described the ways in which SOGI has been measured, discussed quantitative and qualitative findings on the quality of existing measures, and identified areas where future research is needed. The current literature has paid close attention to how SOGI is defined, how SOGI concepts are understood by respondents, and how measurement challenges, such as respondent misinterpretation, item nonresponse, and social desirability, lead to biased outcomes.

The paper began with a review of the key concepts of sexual orientation (including sexual identity, sexual attraction, and sexual behavior), gender identity, and household relationships, and then reviewed evaluation studies on each topic.

a. Sexual Orientation

Sexual Identity

The literature on sexual identity measures reveals a number of ways to improve their performance and enhance their validity and reliability. Research on question order effects could improve data quality, as placing sexual identity items near demographic survey questions rather than around sexual attraction and behavior questions may lead to less biased responses. In addition, the literature on sexual identity may benefit from more research on how best to reduce racial and ethnic bias, including the use of terminology that may be more relevant to other racial, ethnic, and language minorities. Future research on adolescents' understanding of sexual identity questions would be beneficial, as research on

adolescents is limited and existing research has identified age differences in comprehension and interpretation. Similarly, potential differences in the responses provided by older adults would benefit from further study.

Other areas of future research include survey operations and administration, specifically interviewer training and whether proxy measurement should be allowed. Research suggests that data quality can be improved by focusing interviewer training on the implications of invalid data and asking questions as worded, especially questions on sensitive topics. Research related to privacy and discretion when administering sexual identity questions is also needed. Other areas of future research include the consistent administration of questions across different survey modes (e.g., telephone, face-to-face, internet survey) and the use of proxy respondents.

Sexual Attraction

Sexual attraction refers to the sex or gender to which someone feels attraction. Future research is needed on question wording, response options, as well as mode of data collection. A better understanding of how adolescents and adults interpret and define “attraction” is also warranted.

We encourage a focus on question context and placement in future research. For example, in one study, when the identity and attraction items followed sexual behavior questions they tended to produce higher item nonresponse than when identity and attraction items appeared before the behavior items. This highlights the importance of question context and location as factors affecting performance.

Nonresponse is also a concern when measuring sexual attraction. The conflation of sexual attraction and sexual identity may contribute to item nonresponse. It is recommended that more testing be conducted with adolescents in particular to understand how they comprehend measures of sexual orientation broadly as well as sexual attraction specifically.

Sexual Behavior

Sexual behavior, operationalized for our purposes as sex of a person’s sexual partners, is one of three dimensions of sexual orientation. Current studies investigate how sexual behavior items perform, item nonresponse, and respondent’s reactions to capturing their sexual behavior. The reviewed evaluations drew upon groups that varied by demographic characteristics (e.g., income, race, marital status, age, language). More research on the structure and content of questions and how responses vary by demographic group would help advance valid measurement across subgroups.

Nonresponse is a fertile area for future research. Findings suggest that item nonresponse decreases as income increases, increases as age increases, and increases with higher levels of education.

Misreporting, a type of measurement error, is another area for future investigation. One study found that when non-heterosexuals reported their behavior as exclusively heterosexual and when heterosexuals did not report same-sex sexual behavior, misreporting became apparent. These authors found that among men, misreporting increased as income increased, white men had a higher probability of misreporting compared to men of all other races, and misreporting decreased as age increased. Among men and women, misreporting decreased as city size increases.

b. Gender Identity

There are a number of areas in which Federal research could add to the literature on how best to measure gender identity. While many organizations support a two-step measure, more research is needed on the assessment of gender identity using a one versus two-step measure. In addition, little is known about how question order in the two-step measure may affect response; some literature found that most transgender participants preferred current gender identity be asked prior to natal sex. The GenIUSS Group recommends asking natal sex first, while WPATH and CoE recommend asking current gender identity first. More research is needed on question order as well as survey design.

Another unresolved issue is whether to include a reference to one's original birth certificate in the natal sex item. While there have been no studies focusing on this specific issue, there is anecdotal evidence that transgender individuals preferred its exclusion. The GenIUSS group recommended two-step items include a birth certificate reference, while CoE's most recent recommendation excludes it. Again, more research is needed.

Current research would also benefit from the inclusion of racial, ethnic, and cultural minorities, as well as language minorities other than Spanish. The literature to date has focused mostly on comprehension in the English and, more recently, Spanish-speaking populations. Future research, specifically Federal research, may benefit from increased recruitment of transgender individuals, people of color, and people from cultures in which a third gender exists (e.g., "two-spirit" in American Indian tribes and "hijras" in South Asian cultures).

Recent variations of the two-step measure have included a verification item when natal sex and current gender identity do not agree. These verification measures are utilized in the NCVS and the CHIS as questions asked directly of respondents, and in SPI as an interview instruction. While items like these have been shown to reduce measurement error in the measurement of same-sex households, there are no studies examining their application to the measurement of gender identity.

In addition to how best to *ask* gender identity, there are other methodological challenges related to the measurement of gender identity. For example, little is known about how best to ask gendered questions, such as questions about mammograms and birth control, once a respondent has identified as transgender. Researchers have recommended the collection of name and pronoun preferences in a medical setting, but this may be just as important in a survey setting. This information can be used to properly tailor mailing materials to transgender participant preferences (e.g., using the proper salutation in a survey mailing). This information can also be used to properly refer to the gender of transgender respondents in a proxy survey.

Moreover, there appears to be limited research evaluating items measuring gender expression, or an individual's external manifestation of gender. Additional research addressing this topic would prove beneficial and further develop knowledge on this concept central to SOGI research.

c. Household Relationships

The studies evaluated identified persistent sex-relationship to householder inconsistencies, leading to overestimates of same-sex households in the decennial Census, ACS, and other Federal surveys. The overestimates are largely attributed to accidental mismarking on sex by a very small proportion of opposite-sex married couples. While many of the testing efforts described recommend the use of a revised relationship to householder question, the evidence to date suggests that inconsistencies and

overestimates of same-sex couple households are still likely. While the revised design shows promise, the current ACS, for example, continues to use the relationship question fielded with the 2010 ACS.

A number of the reviewed studies mentioned the automated edits in the interviewer-administered phases of the ACS as an explanation for the lower sex-relationship inconsistency rates compared to the 2010 Census. However, these studies failed to quantify the extent to which these edits led to changes in reporting of sex, relationships, or both. We know very little about the behavior of interviewers when confronted with these edits.

Finally, while identifying and addressing the issues behind the documented inconsistencies is important for producing accurate estimates of same-sex couple households (and a number of other couple types), relying on sex and relationship information does not provide a direct measure of sexual orientation.

d. Conclusion

Federal agencies have expressed interest and enthusiasm for improving the measurement SOGI. The Federal Government is interested in better understanding the SGM community in order to better serve its needs. This cannot be accomplished without sound data on the population. This working paper described what is known about current measures on sexual orientation, gender identity, and household relationships and highlighted the methodological questions that remain to be answered.

References

- American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington, DC: American Psychiatric Association.
- Austin, B, Conron, K, Patel, A, and Freedner, N. 2007. "Making Sense of Sexual Orientation Measures: Findings from a Cognitive Processing Study with Adolescents on Health Survey Questions." *Journal of LGBT Health Research*, 3:1, 55-65, DOI: 10.1300/J463v03n01_07
- Baker, K, and Hughes, M. 2016. "Sexual Orientation and Gender Identity Data Collection in the Behavioral Risk Factor Surveillance System." Center for American Progress.
<https://cdn.americanprogress.org/wp-content/uploads/2016/03/29090401/BRFSSdatacollect-brief-03.31.16.pdf>
- Bates, N, DeMaio, TJ, Robins, C, and Hicks, W. 2010. "Classifying Relationship and Marital Status among Same-Sex Couples." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Chicago, IL, May 13-16.
- Beck, A, Berzofsky, M, Caspar, R, and Krebs, C. 2014. "Sexual Victimization in Prisons and Jails Reported by Inmates, 2011-12." *Bureau of Justice Statistics*.
<http://www.bjs.gov/index.cfm?ty=pbdetail&iid=4654>
- Bureau of Justice Statistics. 2015 "Attachment B_2016 SPI [Survey of Prison Inmates] QuestionnaireTrack Changes_11-18-15 (V2)" Department of Justice.
- Case, P, Austin, B, Hunter D, Willett, W, Malspeis, S, Manson, J, and Spiegelman, D. 2006. "Disclosure of Sexual Orientation and Behavior in the Nurses' Health Study II." *Journal of Homosexuality* 51(1): 13-31. doi:10.1300/j082v51n01_02.
- Cahill, S, Singal, R, Grasso, C, et al. 2014. "Do Ask, Do Tell: High Levels of Acceptability by Patients of Routine Collection of Sexual Orientation and Gender Identity in Four Diverse American Community Health Centers." *PLoS One* 9: e107104.
- Clark, M, Armstrong, G, Bonacore, L. 2005. "Measuring Sexual Orientation and Gender Expression among Middle-Aged and Older Women in a Cancer Screening Study." *Journal of Cancer Education* 20(2): 108-112.
- Center for American Progress and Movement Advancement Project. 2015. "Unfair Price: The Financial Penalty for Being Transgender in America." Movement Advancement Project.
<http://www.lgbtmap.org/unfair-price-transgender#sthash.zMVOLI8f.pdf>
- Coffman, K, Coffman, L and Ericson, K. 2013. "The Size of the LGBT Population and the Magnitude of Anti-gay Sentiment are Substantially Underestimated." No. w19508. National Bureau of Economic Research.
- Correll, S, Ridgeway, C, Saperstein, A, and Westbrook, L. (2014). Gender Identity and Diversity: Revision and Updates. (unpublished report)

- Interagency Working Group on Sexual Orientation and Gender Identity. . 2016. "Current Measures of Sexual Orientation and Gender Identity in Federal Surveys." Accessed July 2016 at: https://fcsm.sites.usa.gov/files/2014/04/WorkingGroupPaper1_CurrentMeasures_08-16.pdf
- Dahlhamer, J, Galinsky, A, Joestl, S, and Ward, B. "Sexual Orientation in the 2013 National Health Interview Survey: A Quality Assessment." *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research* 169 (2014): 1-32.
- DeMaio, TJ and Bates, N. 2012. "New Relationship and Marital Status Questions: A Reflection of Changes to the Social and Legal Recognition of Same-Sex Couples in the U.S." Survey Methodology Paper 2012-02. Suitland, MD: U. S. Census Bureau.
- DeMaio, TJ, Bates, N, and O'Connell M. 2013. "Exploring Measurement Error Issues in Reporting of Same-Sex Couples." *Public Opinion Quarterly*, 77:145-158.
- Deutsch, M, Green, J, Keatley, J, et al. 2013. "Electronic Medical Records and the Transgender Patient: Recommendations from the World Professional Association for Transgender Health EMR Working Group." *Journal of the American Medical Informatics Association* 20 (4): 700-703.
- Fisher, S, Michaels, M, and Kahn, E. 2001. "A Needs Assessment Study of LBT Women: The Relationship between Sexual Orientation/Gender Identity, Health-Seeking Behaviors, and Perceived Quality of Care Issues." *Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.*
- Gates, G. 2011. "How Many People are Lesbian, Gay, Bisexual, and Transgender?" The Williams Institute. Accessed July 2016. <https://escholarship.org/uc/item/09h684x2>
- Glaze, L. 2015. "Nonsubstantive Changes to OMB #1121-0152, Survey of Prison Inmates, 2016" Department of Justice.
- Grant, D, and Jans, M. No date. "Assessing Sexual Identity & Behavior in California's Diverse Population" UCLA Center for Health Policy Research.
- Grant, LA, Mottet, JD, Tanis, J., Herman, J, Harrison, J, Keisling, M. 2010. "National Transgender Discrimination Survey Report on Health and Health Care." National LGBTQ Task Force. http://www.thetaskforce.org/static_html/downloads/resources_and_tools/ntds_report_on_health.pdf
- Gruskin, E, Hart, S, Gordon, N, and Ackerson, L. 2001. "Patterns of Cigarette Smoking and Alcohol Use among Lesbians and Bisexual Women Enrolled in a Large Health Maintenance Organization." *American Journal of Public Health* 91 (6): 976-9.
- Ingraham, N, Pratt, V, and Gorton N. 2015. "Counting Trans* Patients: A Community Health Center Case Study." *Transgender Studies Quarterly* 2 (1): 136-147.
- Institute of Medicine of the National Academies, Committee on Lesbian, Gay, Bisexual, and Transgender Health Issues and Research Gaps and Opportunities. 2011. "The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding." *National Academies Press*, March 31. <http://www.nationalacademies.org/hmd/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx>

- Interagency Working Group on Measuring Relationships in Federal Household Surveys (MRFHS). 2014. Improved Measurement of Household Relationships in Federal Surveys. Accessed online at http://fcsm.sites.usa.gov/files/2014/04/MRFHS_StatisticalPolicyWorkingPaper201408.pdf.
- Jans, M, Viana, J, Grant, D, Cochran, S, Lee, A, and Ponce, N. 2015. "Trends in Sexual Orientation Missing Data over a Decade of the California Health Interview Survey." *American Journal of Public Health* 105 (5): E43-E50.
- Joloza, T, Evans, J., and O'Brien, R. 2010. "Measuring Sexual Identity: An Evaluation Report." Office for National Statistics, United Kingdom.
- Kelly, M. 2016. Sex versus Gender Categorization. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. 1–3
- Kim, H, and Fredriksen-Goldsen, K. 2013 "Nonresponse to a Question on Self-Identified Sexual Orientation in a Public Health Survey and Its Relationship to Race and Ethnicity." *American Journal of Public Health* 103(1): 67-69.
- Kreider, R and Lofquist, D. 2014. "Matching Survey Data with Administrative Records to Evaluate Reports of Same-Sex Married Couple Households." SEHSD Working Paper 2014-36. Suitland, MD: U. S. Census Bureau.
- Krosnick, A, Narayan, S, and Smith, R. 1996. "Satisficing in Surveys: Initial evidence." In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29-44). San Francisco: Jossey-Boss. <https://pprg.stanford.edu/wp-content/uploads/1996-Satisficing-in-surveys-Initial-evidence.pdf>
- Lev, A. 2016. Gender Dysphoria. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. 1–3.
- Lewis, JM, Bates, N, and Streeter, M. 2015. "Measuring Same-Sex Couples: The What and Who of Misreporting on Relationship and Sex." SEHSD Working Paper 2015-12. Suitland, MD: U. S. Census Bureau.
- Little, Roderick J.A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. New York, N.Y.: Wiley, 1987.
- Lombardi, E, and Banik, S. 2016. "The Utility of the Two-Step Gender Measure within Trans and Cis Populations." *Sexuality Research and Social Policy* 13(3): 288–296.
- Magliozzi, D, Saperstein, A, and Westbrook, L. 2016. "Scaling Up: Representing Gender Diversity in Survey Research." *Socius: Sociological Research for a Dynamic World* 2: 1-11.
- Martinez, M, Henderson, A, and Luck, J. 2016. "Cognitive Pretesting of the National Crime Victimization Survey Supplemental Victimization Survey." U.S. Census Bureau (unpublished report)
- McCabe, SE, Hughes, T, Bostwick, W, Morales M, and Boyd, C. 2012. "Measurement of Sexual Identity in Surveys: Implications for Substance Abuse Research." *Archives of Sexual Behavior* 41 (3): 649-57.
- Miller, K. 2001. "Cognitive Testing of the NHANES Sexual Orientation Questions." National Center for Health Statistics. Q-Bank. Retrieved from <http://www.cdc.gov/qbank>.

- Miller, K and Ryan, JM. "Design, Development and Testing of the NHIS Sexual Identity Question." *National Center for Health Statistics* (2011): 1-33.
- Mohr, J and Kendra, M. 2011. "Revision and Extension of a Multidimensional Measure of Sexual Minority Identity: The Lesbian, Gay, and Bisexual Identity Scale." *Journal of Counseling Psychology* 58 (2): 234.
- National Center for Education Statistics. Forthcoming. "Cognitive Testing Items HSLs:09 Second Follow-Up – Field Test Version Findings."
- National Opinion Research Center (NORC). 2016. "Testing Alternative Response Options for Spanish Translations of Sexual Identity Items for National Surveys." (unpublished report)
- O'Connell, M and Feliz, S. 2011. "Same-Sex Couple Household Statistics from the 2010 Census." SEHSD Working Paper 2011-26. Suitland, MD: U. S. Census Bureau.
- Poston, D, and Chang, Y. nd. "Patterns of Gay Male and Lesbian Partnering in the Metropolitan Areas of the United States in 2010." International Union for the Scientific Study of Population. http://iussp.org/sites/default/files/event_call_for_papers/Poston-Chang-IUSSP-2013-FINAL_0.pdf
- Reisner, SL, Conron, KJ, Tardiff, LA, et al. 2014. "Monitoring the Health of Transgender and Other Gender Minority Populations: Validity of Natal Sex and Gender Identity in a U.S. National Cohort of Young Adults." *BMC Public Health* 14: 1224.
- Ridolfo, H, Miller, K, and Maitland, A. 2012. "Measuring Sexual Identity using Survey Questionnaires: How Valid are our Measures?" *Sexuality Research & Social Policy* 9 (2): 113-124.
- Ridolfo, H., Perez, K, and Miller, K. (2010). Testing of Sexual Identity and Health Related Questions: Results of Interviews Conducted May-July 2005. Q-Bank. Retrieved from <http://www.cdc.gov/qbank>.
- Ridolfo, H, and Schoua-Glusberg, A. (2009). Testing of NHANES ACASI Sexual Behavior Questions: Results of Interviews Conducted November 2008–February 2009. Q-Bank. Retrieved from <http://www.cdc.gov/qbank>.
- Robles, R, Fresan, A, Vega-Ramirez, H, et al. 2016. "Removing Transgender Identity from the Classification of Mental Disorders: A Mexican Field Study for ICD-11." *The Lancet Psychiatry* 3 (9): 850-859.
- Saperstein, A and Westbrook, L. 2015. "Gender Identity and Diversity." Time-sharing Experiments for the Social Sciences (TESS) Studies and Data. Accessed online. <http://www.tessexperiments.org/data/sapersteinBR44.html>
- Saewyc, E, Bauer, G, Skay, C, Bearinger, L, Resnick, M, Reis, E, and Murphy, A. "Measuring Sexual Orientation in Adolescent Health Surveys: Evaluation of Eight School-based Surveys." *Journal of Adolescent Health* 35, no. 4 (2004): 345-e1.

- Scheim, AI, Bauer, GR, and Shokoohi, M. (2016). "Heavy Episodic Drinking among Transgender Persons: Disparities and Predictors." *National Institutes of Health*.
[http://www.drugandalcoholdependence.com/article/S0376-8716\(16\)30250-2/fulltext?rss=yes](http://www.drugandalcoholdependence.com/article/S0376-8716(16)30250-2/fulltext?rss=yes)
- Sexual Minority Assessment Research Team (SMART). 2009. "Best Practices for Asking Questions about Sexual Orientation on Surveys." *Williams Institute at UCLA School of Law*, November.
<http://williamsinstitute.law.ucla.edu/wp-content/uploads/SMART-FINAL-Nov-2009.pdf>
- Shoemaker, P, Eichholz, M, and Skewes, E. 2002. "Item Nonresponse: Distinguishing Between Don't Know and Refuse." *International Journal of Public Opinion Research*, 14(2): 193-201.
- Singer, E. "Introduction: Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70, no. 5 (2006): 637-45. Accessed July 2016. doi:10.1093/poq/nfl034.
- Smirnova, M and Scanlon, P. 2013. "2013 Census Test Cognitive Interview Report." *Survey Methodology Paper 2013-18*. Suitland, MD: U. S. Census Bureau.
- Spade, D 2008. "Documenting Gender." *Williams Institute at UCLA School of Law*, April 30.
<http://williamsinstitute.law.ucla.edu/wp-content/uploads/Spade-Documenting-Gender-Apr-2008.pdf>
- Stern, MJ, Michaels, S, Milesi, C, et al. 2016. "Task 1.32.a: Develop and Test Sexual and Gender Minority Status (LGBT) Items Final Summary Report." (unpublished report)
- The GenIUSS Group. 2014. *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys*. J.L. Herman (Ed.). Los Angeles, CA: The Williams Institute.
- Tourangeau, R, Rips, LJ, and Rasinski, K. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- VanKim, N, Padilla, J, Joseph, L, and Goldstein, A. 2010. "Adding Sexual Orientation Questions to Statewide Public Health Surveillance: New Mexico's Experience." *American Journal of Public Health* 100 (12): 2392-6.
- Vrangalova, Z, and Savin-Williams, RC. 2012. "Mostly Heterosexual and Mostly Gay/Lesbian: Evidence for New Sexual Orientation Identities." *Archives of Sexual Behavior* 41(1): 85-101.
- Westbrook, L, and Saperstein, A. 2015. "New Categories are Not Enough: Rethinking the Measurement of Sex and Gender in Social Surveys." *Gender and Society* 29(4): 534-560.
- World Health Organization (WHO). "Gender." World Health Organization. Accessed July 2016.
<http://www.who.int/gender-equity-rights/understanding/gender-definition/en/>

Appendices

A. Surveys Measuring Sexual Identity in Cited Work

Author(s)	Survey/method
Austin, Conron, et al., 2007	30 cognitive interviews of adolescents
Cahill et al., 2014	Survey of 301 English-speaking patients randomly selected from 4 CHARN-affiliated Community Health Centers (CHCs)
Case et al., 2006	Pilot study using Nurses' Health Study II
Clark et al., 2005	Cancer Screening Project for Women – Cognitive Test
Coffman et al., 2013	Experiment via Amazon Mechanical Turk (MTurk)
Dahlhamer et al., 2014	National Health Interview Survey (NHIS), 2013
Fisher et al., 2001	Needs assessment with LBT women in Washington, DC, metro area
Grant and Jans, n.d.	California Health Interview Survey (CHIS)
Gruskin et al., 2001	Health survey of Kaiser Permanente Mental Care program members in Northern California
Ingraham et al., 2015	Community Health Center Case Study
Jans et al., 2015	CHIS, 2003-2011
Joloza et al., 2010	UK Integrated Household Survey (IHS)
Kim and Fredriksen-Goldsen, 2013	Washington State Behavioral Risk Factor Surveillance System (BRFSS)
Martinez, Henderson, and Luck, 2016	National Crime Victimization Survey (NCVS) – cognitive interviews
McCabe, Hughes, Bostwick, Morales, and Boyd, 2012	Simple random sample of 19,370 undergraduates at large, Midwestern university
Miller, 2001	National Center for Health Statistics (NCHS) cognitive testing/evaluation study
Miller and Ryan, 2011	NCHS cognitive testing/evaluation study
Mohr and Kendra, 2011	Sample of university students in which the majority identified as lesbian/gay
National Center for Education Statistics, U.S. Department of Education, forthcoming	High School Longitudinal Study of 2009 (HSL:09)
Poston and Chang, 2015	NSFG, 2006-2008
Ridolfo and Schoua-Glusberg, 2009	NCHS cognitive testing/evaluation study
Ridolfo et al., 2010	NCHS cognitive testing/evaluation study
	NCHS cognitive testing/evaluation study
	National Survey of Family Growth (NSFG), 2002
	NSFG, 2006-2008
	2001-2008 National Health and Nutrition Examination Survey (NHANES)
Saewyc et al., 2004	Eight student surveys, 7th-12th grades, ages 12 to 20
Stern et al., 2016	Medicare Current Beneficiary Survey (MCBS) – cognitive interviews
VanKim et al., 2010	2007-2008 Behavioral Risk Factor Surveillance System (BRFSS)
VanKim et al., 2010	2003 New Mexico Adult Tobacco Survey (ATS)
Vrangalova and Savin-Williams, 2012	Online survey via Facebook focused on sexual morality

B. Sexual Identity Items Used in Cited Work

Authors, year	Survey (if applicable)	Question wording	Response options
Austin, Conron, et al., 2007	---	<p>1. Sexual identity: Which of the following best describes you?</p> <p>2. Sexual identity/attraction: Which of the following best describes your feelings?</p>	<p>_ Heterosexual (straight)</p> <p>_ Gay or lesbian</p> <p>_ Bisexual</p> <p>_ Not sure</p> <p>_ Completely heterosexual (attracted to persons of the opposite sex)</p> <p>_ Mostly heterosexual</p> <p>_ Bisexual (equally attracted to men and women)</p> <p>_ Mostly homosexual</p> <p>_ Completely homosexual (gay/lesbian, attracted to persons of the same sex)</p> <p>_ Not sure</p>
Cahill et al., 2014	---	Do you think of yourself as:	<p>1) Lesbian, gay or homosexual</p> <p>2) Straight or heterosexual</p> <p>3) Bisexual</p> <p>4) Something else, please describe _____</p> <p>5) Don't know</p>
Case et al., 2006	---	<p>Whether you are currently sexually active or not, what is your sexual identity or orientation?</p> <p>Please choose one answer:</p>	<p>1) Heterosexual</p> <p>2) Bisexual</p> <p>3) Lesbian, gay, or homosexual</p> <p>4) None of the above</p> <p>5) Prefer not to answer</p>
Clark et al., 2005	---	The next question is about how do you think of yourself? Please answer all that apply to you.	<p>1) Lesbian, gay or homosexual</p> <p>2) Straight or heterosexual</p> <p>3) Bisexual</p> <p>4) Transgendered</p> <p>5) Not sure</p> <p>6) You think of yourself in different terms (please describe)</p>
Coffman et al., 2013	---	Do you consider yourself to be heterosexual?	<p>Yes</p> <p>No</p>
		I consider myself to be heterosexual.	(Respondents were asked to determine if the statement applied to them)

Authors, year	Survey (if applicable)	Question wording	Response options
Dahlhamer et al., 2014	NHIS, 2013	Which of the following best represent how you think of yourself?	Gay (Lesbian or gay) Straight, that is, not gay (lesbian or gay) Bisexual Something else I don't know the answer
BJS, 2015	2016 Survey of Prison Inmates	Which of the following best represents how you think of yourself?	-Lesbian or gay -Straight, that is not lesbian or gay -Bisexual -Something else -You don't know the answer (Refused)
Grant and Jans, n.d.	2003-present California Health Interview Survey (CHIS)	Do you think of yourself as straight or heterosexual, as gay {, lesbian} or homosexual, or bisexual?	
Gruskin et al., 2001	---	What is your sexual orientation?	1) Heterosexual (straight) 2) Homosexual (gay, lesbian) 3) Transsexual
Ingraham, Pratt, and Gorton, 2015	---	How do you identify yourself, sexually?	1) Lesbian 2) Gay 3) Queer 4) Bisexual 5) Heterosexual 6) Celibate 7) Other ____ 8) Decline
		My sexual orientation is:	1) Lesbian 2) Gay 3) Queer 4) Bisexual 5) Heterosexual 6) Celibate 7) Other ____ 8) Decline
Jans et al., 2015	CHIS, 2003-2013	Do you think of yourself as straight or heterosexual, as [gay/gay, lesbian] or homosexual, or bisexual?	

Authors, year	Survey (if applicable)	Question wording	Response options
Joloza, et al., 2010	---	Face-to-face: Which of the options on this card best describes how you think of yourself? Just read out the number next to the description.	27) Heterosexual/straight 21) Gay/lesbian 24) Bisexual 29) Other
		Telephone: I will now read out a list of terms people sometimes use to describe how they think of themselves.	1) Heterosexual or straight 2) Gay or lesbian 3) Bisexual 4) Other Interviewer then reads the question a second time and asks the respondent to say 'yes' when they hear the option that best describes how they think of themselves.
Kim and Fredriksen-Goldsen, 2013	2003-2010 Washington State BRFS	Not stated	1) Heterosexual, that is straight 2) Homosexual, that is gay or lesbian 3) Bisexual 4) Other
Martinez, Henderson, and Luck, 2016	National Crime Victimization Survey (NCVS)	Which of the following best represents how you think of yourself?	-[Lesbian or] Gay -Straight, that is, not [lesbian or] gay -Bisexual -Something else -I don't know the answer (Refused)
McCabe, Hughes, Bostwick, Morales, and Boyd, 2012	---	How would you define your sexual identity? Would you say that you are...	(1) Only lesbian/gay (2) Mostly lesbian/gay (3) Bisexual (4) Mostly heterosexual (5) Only heterosexual (6) Other/specify versus (1) Heterosexual (2) Lesbian/gay (3) Bisexual (4) Refused

Authors, year	Survey (if applicable)	Question wording	Response options
Miller, 2001 Ridolfo and Schoua-Glusberg, 2009	2001-2008 NHANES	Do you think of yourself as	1) Heterosexual or straight (that is, sexually attracted only to women/men) 2) Homosexual or gay/lesbian (that is, sexually attracted only to men/women) 3) Bisexual (that is, sexually attracted to men and women) 4) Something else 5) Or you're not sure?
Miller and Ryan, 2011	---	Do you think of yourself as:	For men, 1) Gay 2) Straight, that is not gay For women, 1) Lesbian or gay 2) Straight, that is, not lesbian or gay 3) Bisexual 4) Something else 5) Don't know
		By something else, do you mean that...	1) You are not straight, but identify with another label such as queer, trisexual, omnisexual, or pan-sexual 2) You are transgender, transsexual or gender variant 3) You have not figured out your sexuality or are in the process of figuring it out 4) You do not think of yourself as having sexuality 5) You do not use labels to identify yourself 6) You made a mistake and did not mean to pick this answer 7) You mean something else.
Mohr and Kendra, 2011	The Lesbian, Gay, and Bisexual Identity Scale (LGBIS)	1. I prefer to keep my same-sex romantic relationships rather private. 2. If it were possible, I would choose to be straight. 3. I'm not totally sure what my sexual orientation is. 4. I keep careful control over who knows about my same-sex romantic relationships. 5. I often wonder whether others judge me for my sexual orientation.	For each statement, respondents are asked to choose one of the following: 1) Disagree strongly 2) Disagree 3) Disagree somewhat 4) Agree somewhat

Authors, year	Survey (if applicable)	Question wording	Response options
		<ol style="list-style-type: none"> 6. I am glad to be an LGB person. 7. I look down on heterosexuals. 8. I keep changing my mind about my sexual orientation. 9. I can't feel comfortable knowing that others judge me negatively for my sexual orientation. 10. I feel that LGB people are superior to heterosexuals. 11. My sexual orientation is an insignificant part of who I am. 12. Admitting to myself that I'm an LGB person has been a very painful process. 13. I'm proud to be part of the LGB community. 14. I can't decide whether I am bisexual or homosexual. 15. My sexual orientation is a central part of my identity. 16. I think a lot about how my sexual orientation affects the way people see me. 17. Admitting to myself that I'm an LGB person has been a very slow process. 18. Straight people have boring lives compared with LGB people. 19. My sexual orientation is a very personal and private matter. 20. I wish I were heterosexual. 21. To understand who I am as a person, you have to know that I'm LGB. 22. I get very confused when I try to figure out my sexual orientation. 23. I have felt comfortable with my sexual identity just about from the start. 24. Being an LGB person is a very important aspect of my life. 25. I believe being LGB is an important part of me. 26. I am proud to be LGB 	

Authors, year	Survey (if applicable)	Question wording	Response options
		27. I believe it is unfair that I am attracted to people of the same sex.	
National Center for Education Statistics, forthcoming	High School Longitudinal Study of 2009 (HSL:09)	Do you think of yourself as ...	1) Lesbian or gay 2) Straight, that is, not gay 3) Bisexual 4) Something else (please specify) 5) Don't Know
		Now look at this other way of asking the question. SHOWCARD. How would you answer if it was phrased like this?	2nd SHOWCARD: 1) Homosexual 2) Heterosexual 3) Bisexual 4) Something else (please specify) 5) Don't Know
Poston and Chang, 2015	NSFG, 2006-2008	Do you think of yourself as heterosexual or straight; as homosexual, gay, or lesbian; as bisexual; or as something else?	
Ridolfo and Schoua-Glusberg, 2009	2001-2008 NHANES	¿Se considera usted a si mismo	1) Heterosexual (le atraen las mujeres/los hombres) 2) Homosexual o gay/lesbiana (le atraen los hombres/las mujeres) 3) Bisexual (le atraen los hombres y las mujeres) 4) Alguna otra cosa 5) No está seguro(a)?
---	2002-2003 National Survey of Family Growth (NSFG)	Do you think of yourself as	1) Heterosexual 2) Homosexual 3) Bisexual 4) Something else?
Ridolfo et al., 2010	---	What sexual orientation do you consider yourself to be?	1) Heterosexual 2) Gay or lesbian 3) Bisexual 4) Other 5) Don't know
Ridolfo et al., 2010	---	Do you consider yourself to be	1) Heterosexual or straight 2) Gay or lesbian, 3) Bisexual?

Authors, year	Survey (if applicable)	Question wording	Response options
---	2006-2008 National Survey of Family Growth (NSFG)	Do you think of yourself as	1) Heterosexual or straight 2) Homosexual, gay,(or lesbian, 3) Bisexual 4) Or something else?
---	2006-2007 National Survey of Family Growth (NSFG)	When you say "something else," what do you mean?	
---	2006-2008 National Survey of Family Growth (NSFG)	¿Usted se considera...	1) Heterosexual 2) Homosexual 3) Bisexual 4) O alguna otra cosa?
---	2006-2007 NSFG	Cuando usted dice "alguna otra cosa", ¿qué se refiere?	
Saewyc et al., 2004	---	<p>1. Which of the following best describes your feelings:</p> <p>2. Many people say that they have different feelings about themselves when it comes to questions of being attracted to other people. Which of the following best describes your feelings? SAME RESPONSE OPTIONS AS #1</p> <p>3. People have different feelings about themselves when it comes to questions of being attracted to other people. Which of the following best describes your feelings? SAME RESPONSE OPTIONS AS #1</p> <p>4. How would you describe your sexual orientation/preference?</p> <p>5. How would you describe your sexual orientation? SAME RESPONSE OPTIONS AS #4</p>	<p>Response options for #1, 2, and 3: 100% heterosexual (attracted to persons of the opposite sex) Mostly heterosexual Bisexual (equally attracted to men and women) Mostly homosexual 100% homosexual ("gay/lesbian"; attracted to persons of the same sex)</p> <p>Response options for #4 and 5: Heterosexual—attracted to the opposite sex Bisexual—attracted to both sexes Homosexual (gay or lesbian)—attracted to same sex Not sure</p>
Stern et al., 2016	MCBS	Which of the following best represents how you think of yourself?	<p>-[Lesbian or] Gay -Straight, that is, not [lesbian or] gay -Bisexual -Something else -I don't know the answer</p>

Authors, year	Survey (if applicable)	Question wording	Response options
			-Refused
VanKim et al., 2010	---	Do you consider yourself to be ...?	1) Heterosexual or straight, 2) Homosexual (gay or lesbian), 3) Bisexual, 4) Other, 5) Don't know/not sure.
Vrangelova and Savin-Williams, 2012	---	Participants chose one of six labels to identify their sexual orientation:	1) Heterosexual 2) Mostly heterosexual 3) Bisexual 4) Mostly gay/lesbian 5) Gay/lesbian 6) Questioning/ uncertain. An "other" category was also presented.
---	2001 California Health Interview Survey (CHIS)	This next question is about your sexual orientation and I want to remind you again that your answers are completely confidential. Are you gay, {lesbian,} or bisexual? If yes, Is that gay, {lesbian} or bisexual?	

C. Sexual Attraction Items Used in Cited Work

Authors	Mode	Notes	Question wording	Response Options
Austin, Conron, et al., 2007	Cognitive testing based on retrospective talk-aloud method.	Interviews averaged 75 minutes, recorded and transcribed. 30 adolescents near a New England metropolitan area. Participants were of diverse age (15-21), gender identity, sexual orientation, and race-ethnicity.	Which of the following best describes your feelings?	<input type="checkbox"/> Completely heterosexual (attracted to persons of the opposite sex) <input type="checkbox"/> Mostly heterosexual <input type="checkbox"/> Bisexual (equally attracted to men and women) <input type="checkbox"/> Mostly homosexual <input type="checkbox"/> Completely homosexual (gay/lesbian, attracted to persons of the same sex) <input type="checkbox"/> Not sure
			People are different in their sexual attraction to other people. Which best describes your feelings?	<input type="checkbox"/> Only attracted to males <input type="checkbox"/> Mostly attracted to males <input type="checkbox"/> Equally attracted to males and females <input type="checkbox"/> Mostly attracted to females <input type="checkbox"/> Only attracted to females <input type="checkbox"/> Not sure
Clark et al., 2005	Cognitive interviews	<p>40 women participated in the cognitive-based interviews: 19 women who partner with women (WPW) and 21 women who partner with men (WPM). Participants had a mean age of 55.0 years, were predominantly White (39/40), highly educated (38/40 some college or more), working for pay (27/40), and currently insured (36/40).</p> <p>When asked to define the term physical attraction, answers included sexual, chemical, and erotic examples.</p>	<p>People differ in their physical attraction to other people. Which statement best describes your feelings? Would you say ...</p>	<input type="checkbox"/> Only physically attracted to women <input type="checkbox"/> Mostly physically attracted to women <input type="checkbox"/> Equally physically attracted to women and men <input type="checkbox"/> Mostly physically attracted to men <input type="checkbox"/> Only physically attracted to men <input type="checkbox"/> Not sure <input type="checkbox"/> None of the statements describe your feelings

Authors	Mode	Notes	Question wording	Response Options
		Although the definition of physical attraction included sexual intimacy for the majority of women, it also included other factors such as affection, affiliation, and emotional preference.		
Coffman, Coffman, and Ericson, 2013	Online survey	Study examines whether measurements are biased when responses to sexual orientation, behavior, and related opinions are private and anonymous. Experiment on 2,516 U.S. participants who were randomly assigned to a "best practices method" that used a computer and provided privacy and anonymity or assigned to a "veiled elicitation method" that further conceals individual responses.	Are you sexually attracted to members of the same sex?	(1) Yes (2) No
Saewyc et al., 2004	Face to face?	The questions assessed came from eight student surveys with varying sampling designs and subsequent demographic characteristics. In general, the surveys covered students in grades 7-12, ages 12-20. Three of the eight surveys were conducted by research staff in the classrooms; two were conducted by public health nurses in the classrooms; two were conducted by other teachers in the classrooms; and one was conducted by research staff in closed classrooms.	How would you describe your sexual orientation/preference?	Heterosexual—attracted to the opposite sex Bisexual—attracted to both sexes Homosexual (gay or lesbian)—attracted to same sex Not sure
			How would you describe your sexual orientation?	Heterosexual—attracted to the opposite sex Bisexual—attracted to both sexes Homosexual (gay or lesbian)—attracted to same sex Not sure
			Which of the following best describes your feelings:	I am only attracted to people of the same sex as mine, and I will only be sexual with persons of the same sex

Authors	Mode	Notes	Question wording	Response Options
				<p>I am strongly attracted to people of the same sex, and most of my sexual experiences will be with persons of the same sex as mine</p> <p>I am equally attracted to men and women and would like to be sexual with both</p> <p>I am strongly attracted to persons of the opposite sex, and most of my sexual experience will be with persons of the opposite sex</p> <p>I am only attracted to persons of the opposite sex, and I will only be sexual with persons of the opposite sex</p>
			When you think or daydream about sex, do you think about:	<p>Males</p> <p>Females</p> <p>Both</p>

D. Sexual Behavior Items Used in Cited Work

Authors, year	Survey (if applicable)	Question wording	Response options
Berg and Lien, 2006	GSS, 1991-2000	Not described	Not described
Case et al., 2006	Pilot study participants sampled from 1993 NHSII compared to NHSII 1995 cohort	<p>Whether you are currently sexually active or not, which response best describes who you have had sex with over your adult lifetime and over the last five years? Please choose one answer in each column.</p> <p>Please choose one answer:</p>	<p>(a) Over your adult lifetime: <input type="checkbox"/> Never had sex <input type="checkbox"/> Sex with a woman/women <input type="checkbox"/> Sex with a man/men <input type="checkbox"/> Sex with both men and women <input type="checkbox"/> Prefer not to answer</p> <p>(b) Over the last five years: <input type="checkbox"/> Never had sex <input type="checkbox"/> Sex with a woman/women <input type="checkbox"/> Sex with a man/men <input type="checkbox"/> Sex with both men and women <input type="checkbox"/> Prefer not to answer</p>
Poston and Chang, 2015	NSFG, 2006-2008	<p>Sexual behavior:</p> <p>Female 1: Counting all of your male partners, even those you had intercourse with only once, how many men have you had sexual intercourse with? Female 2: Thinking about your entire life, how many female sex partners have you had?</p> <p>Male 1: How many different females have you ever had intercourse with? This includes any female you had intercourse with, even if it was only once or if you did not know her well. Male 2: Thinking about your entire life, how many male sex partners have you had?</p>	<p>Female: Number of male partners Number of female partners</p> <p>Male: Number of female partners Number of male partners</p>
Vrangalova and Savin-Williams, 2012	---	Participants provided in two separate questions the total number of male and female partners with whom they have had a genital sexual experience, defined as including penile-vaginal penetration, oral sex, anal sex, and mutual masturbation.	<p>Number of male partners Number of female partners</p>

E. Gender Identity Items Used in Cited Work

Authors	Mode	Notes	Question wording	Response Options
Cahill et al., 2014	Self-administered (PAPI)		What is your current gender identity? (Check all that apply)	Male Female Female-to-Male (FTM)/Transgender Male/Trans Man Male-to-Female (MTF)/Transgender Female/Trans Woman Genderqueer, neither exclusively male nor female Additional Gender Category/(or Other), please specify
			What sex were you assigned at birth on your original birth certificate? (Check one)	Male Female Decline to Answer, please specify why
Deutsch et al., 2013	Electronic Medical Record		Current Gender Identity	Male Female Transmale/Transman/FTM Transfemale/Transwoman/MTF Genderqueer/Gender-non-conforming Different identity (please state)
			Sex assigned at birth	Male Female Other
BJS 2015	CAPI		What sex were you assigned at birth, on your original birth certificate?	(Male) (Female) (Refused) (Don't know)
			How do you describe yourself (select one)?	Male Female Transgender Do not identify as male, female, or transgender (Refused) (Don't know)

Authors	Mode	Notes	Question wording	Response Options
		This is a soft error in the CAPI system that allows the interview to verify the values were correctly recorded.	(If SEX ASSIGNED AT BIRTH does not equal CURRENT IDENTITY) Verify the values for PH3 and PH4 were correctly recorded.	Interviewer should verify the values were correctly recorded
Ingraham et al, 2015	Self-administered (PAPI)	Patient demographic information form	My gender identity is:	Female Male Trans (MTF) Trans (FTM) Genderqueer Other: _____ Decline
			My sex assigned at birth is:	Female Male Intersex Other: _____ Decline
		Patient demographic update form	My gender identity is:	Female Male Trans (MTF) Trans (FTM) Genderqueer Other: _____ Decline
Lombardi and Banik, 2016	Face-to-face (45) /Adobe Connect (5)	Part of cognitive interview; Participants were asked to read each question aloud, answered it, and explain why they answered the way that they did. Researchers used scripted, semi-structured, and spontaneous probes when appropriate.	What is your sex or gender? (Check ALL that apply)	Male Female Other: Please specify
			What sex were you assigned at birth? (Check one)	Male Female Unknown or Question Not Asked Decline to State

Authors	Mode	Notes	Question wording	Response Options
			Just to confirm, you were assigned {sex assigned at birth} at birth and now describe yourself as a {current gender}. Is that correct?	Yes No Don't know Refused
Martinez et al., 2016	Face-to-face (58) / Telephone (2)	Part of cognitive interview; In addition to asking the questions, interviewers also asked respondents probing questions that were designed to address specific testing issues for particular questions. The specific type of cognitive interview consisted of both concurrent and quasi-retrospective probing. After asking the sex question below, researchers probed about the question, "What sex were you assigned at birth, on your original birth certificate?"	On your original birth certificate, was your sex assigned as male or female?	Male Female Refused Don't Know
			Do you currently describe yourself as a man, woman, or transgender person?	Man Woman Transgender person None of these
Reisner et al., 2014			What sex were you assigned at birth, on your original birth certificate? (check one)	Female Male
			How do you describe yourself? (check one)	Female Male Transgender Do not identify as female, male, or transgender
Saperstein and Westbrook, 2014	Self-administered (electronic)	Order of items switched at random	What sex were you assigned at birth?	Male Female Intersex

Authors	Mode	Notes	Question wording	Response Options
			What is your current gender?	Woman Man Transgender
Stern et al., 2016	Face-to-face	Part of cognitive interview; implemented retrospective probing for MCBS cognitive interviews, in which participants answer questions as normal and then interviewers probe with additional questions aimed at confirming comprehension of questions and terms, identifying sources of confusion or ambiguity in survey questions, and learning how participants arrive at answers.	What sex were you assigned at birth, on your original birth certificate?	Male Female
			How do you describe yourself?	Male Female Transgender Do not identify as female, male, or transgender
The GenIUSS Group 2014		Recommended measures for the two-step approach	What sex were you assigned at birth, on your original birth certificate?	Male Female
			How do you describe yourself? (check one)	Male Female Transgender Do not identify as female, male, or transgender
		Promising measure to replace the gender identity step in the two-step approach (more testing is needed)	What is your current gender identity?	Male Female Trans male/Trans man Trans female/Trans woman Genderqueer/Gender non-conforming Different identity (please state):

Authors	Mode	Notes	Question wording	Response Options
		Recommended measures for the single-item approach	Some people describe themselves as transgender when they experience a different gender identity from their sex at birth. For example, a person born into a male body, but who feels female or lives as a woman. Do you consider yourself to be transgender?	Yes, transgender, male to female Yes, transgender, female to male Yes, transgender, gender non-conforming No

F. Gender Identity Items in Federal Surveys

Single-item gender identity measures used in Federal surveys

Behavioral Risk Factor Surveillance System (BRFSS)	National Inmate Survey (NIS)	
<p>Do you consider yourself to be transgender?^{1,2}</p> <p><i>Yes, Transgender, male-to-female</i></p> <p><i>Yes, Transgender, female-to-male</i></p> <p><i>Yes, Transgender, gender non-conforming</i></p> <p><i>No</i></p> <p><i>Don't know/not sure</i></p> <p><i>Refused</i></p> <p>INTERVIEWER NOTE: If asked about definition of <i>transgender</i>: Some people describe themselves as transgender when they experience a different gender identity from their sex at birth. For example, a person born into a male body, but who feels female or lives as a woman would be transgender. Some transgender people change their physical appearance so that it matches their internal gender identity. Some transgender people take hormones and some have surgery. A transgender person may be of any sexual orientation—straight, gay, lesbian, or bisexual.</p> <p>INTERVIEWER NOTE: If asked about definition of <i>gender non-conforming</i>: Some people think of themselves as gender non-conforming when they do not identify only as a man or only as a woman.</p>	<p>Are you male, female, or transgender?</p> <p><i>Male</i></p> <p><i>Female</i></p> <p><i>Transgender</i></p> <tr> <th data-bbox="815 499 1424 569">Population Assessment of Tobacco and Health (PATH)</th> </tr> <p>Some people describe themselves as transgender when they experience a different gender identity from their sex at birth. For example, a person born into a male body, but who feels female or lives as a woman would be transgender. Do you consider yourself to be transgender?</p> <p><i>Yes</i></p> <p><i>No</i></p> <p><i>Don't know</i></p> <p><i>Refused</i></p> <p><i>Not Sure</i></p> <p>[If R answers YES] Do you consider yourself to be male-to-female, female-to-male, or non-conforming?</p> <p><i>Yes, Transgender, male to female</i></p> <p><i>Yes, Transgender, female to male</i></p> <p><i>Yes, Transgender, gender nonconforming</i></p> <p><i>No</i></p> <p><i>Not sure</i></p> <p><i>Don't know</i></p> <p><i>Refused</i></p> <p><i>Not Sure</i></p>	Population Assessment of Tobacco and Health (PATH)
Population Assessment of Tobacco and Health (PATH)		

Two-step gender identity measures used in Federal surveys

National Crime Victimization Survey (NCVS)	Survey of Prison Inmates (SPI)
<p>What sex were you assigned at birth, on your original birth certificate? <i>Male</i> <i>Female</i> <i>(Refused)</i> <i>(Don't know)</i></p> <p>Do you currently describe yourself as male, female or transgender? <i>Male</i> <i>Female</i> <i>Transgender</i> <i>None of these</i></p> <p>[If SEX ASSIGNED AT BIRTH does not equal CURRENT IDENTITY] Just to confirm, you were assigned {SEX ASSIGNED AT BIRTH} at birth and now describe yourself as {CURRENT IDENTITY}. Is that correct? <i>Yes</i> <i>No</i> <i>(Refused)</i> <i>(Don't know)</i></p>	<p>What sex were you assigned at birth, on your original birth certificate? <i>(Male)</i> <i>(Female)</i> <i>(Refused)</i> <i>(Don't know)</i></p> <p>[Field Interviewer Note: Did they tell you that you were born male or female?"]</p> <p>How do you describe yourself (select one)? <i>Male</i> <i>Female</i> <i>Transgender</i> <i>Do not identify as male, female, or transgender</i> <i>(Refused)</i> <i>(Don't know)</i></p> <p>IF (PH3 = 1 AND PH4 ne 1) OR (PH3 = 2 AND PH4 ne 2)] VERIFY THE VALUES FOR PH3 AND PH4 WERE CORRECTLY RECORDED</p>
National Adult Tobacco Survey (NATS)	
<p>What sex were you at birth? <i>Male</i> <i>Female</i></p> <p>Do you currently consider yourself to be: <i>Male</i> <i>Female</i></p>	

G. Scales Used to Measure Gender Expression

Authors, year	Scale wording	Response options
Correll et al., 2014	<p>First-order scale –</p> <p>“In general, how do you see yourself? Please answer on both scales below.”</p> <p>Separate scales for feminine and masculine.</p>	<p>Not at all – 1 2 3 4 5 6 Very – 7</p> <p>Same response options for feminine and masculine scales.</p>
	<p>Third-order scale –</p> <p>“In general, how do you think most people see you? Please answer on both scales below.”</p> <p>Separate scales for feminine and masculine.</p>	<p>Not at all – 1 2 3 4 5 6 Very – 7</p> <p>Same response options for feminine and masculine scales.</p>
Magliozzi et al., 2016	<p>First-order scale –</p> <p>“In general, how do you see yourself? Please answer on both scales below.”</p> <p>Separate scales for feminine and masculine.</p>	<p>Not at all 1 2 3 4 5 Very</p> <p>Same response options for feminine and masculine scales.</p>
	<p>Third-order scale –</p> <p>“In general, how do most people see you? Please answer on both scales below.”</p> <p>Separate scales for feminine and masculine.</p>	<p>Not at all 1 2 3 4 5 Very</p> <p>Same response options for feminine and masculine scales.</p>

H. Measures of Household Relationships

Figure 1. Example of a sex and relationship to householder edit check: 2016 National Health Interview Survey

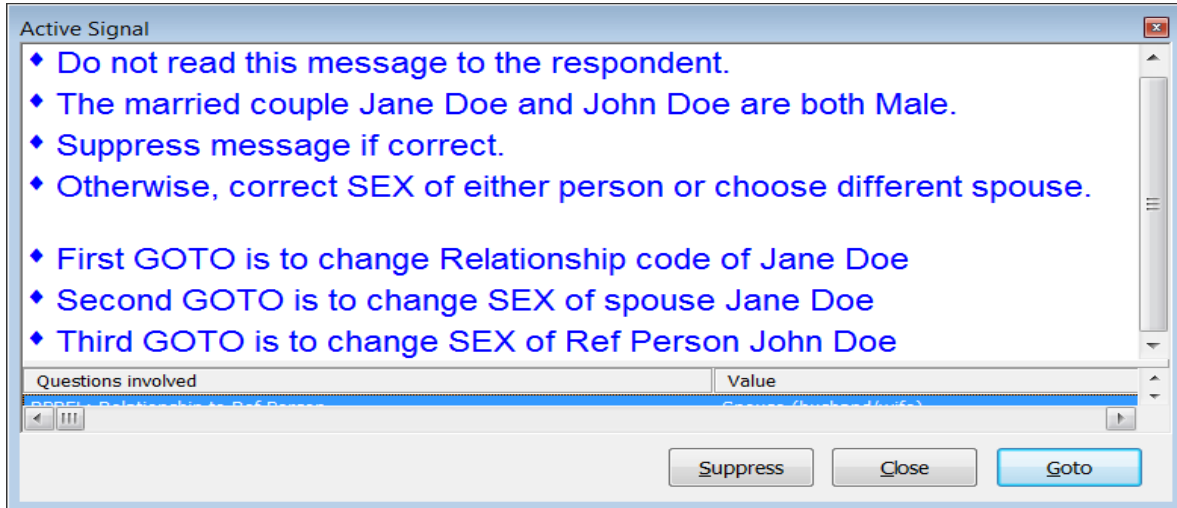


Figure 2. Relationship question used in the 2010 Decennial Census and 2010 American Community Survey

How is this person related to Person 1? Mark (X) ONE box.

- | | |
|---|--|
| <input type="checkbox"/> Husband or wife | <input type="checkbox"/> Parent -in-law |
| <input type="checkbox"/> Biological son or daughter | <input type="checkbox"/> Son-in-law or daughter-in-law |
| <input type="checkbox"/> Adopted son or daughter | <input type="checkbox"/> Other relative |
| <input type="checkbox"/> Stepson or stepdaughter | <input type="checkbox"/> Roomer or boarder |
| <input type="checkbox"/> Brother or sister | <input type="checkbox"/> Housemate or roommate |
| <input type="checkbox"/> Father or mother | <input type="checkbox"/> Unmarried partner |
| <input type="checkbox"/> Grandchild | <input type="checkbox"/> Other nonrelative |

Figure 3. Revised Relationship to Householder Question Tested in the 2013 American Housing Survey

How is this person related to Person 1? Mark (X) ONE box.

- | | |
|---|--|
| <input type="checkbox"/> Opposite-sex husband/wife/spouse | <input type="checkbox"/> Grandchild |
| <input type="checkbox"/> Opposite-sex unmarried partner | <input type="checkbox"/> Parent -in-law |
| <input type="checkbox"/> Same-sex husband/wife/spouse | <input type="checkbox"/> Son-in-law or daughter-in-law |
| <input type="checkbox"/> Same-sex unmarried partner | <input type="checkbox"/> Other relative |
| <input type="checkbox"/> Biological son or daughter | <input type="checkbox"/> Roomer or boarder |
| <input type="checkbox"/> Adopted son or daughter | <input type="checkbox"/> Housemate or roommate |
| <input type="checkbox"/> Stepson or stepdaughter | <input type="checkbox"/> Foster child |
| <input type="checkbox"/> Brother or sister | <input type="checkbox"/> Other nonrelative |