

Measuring Sexual Identity on Federal Health Surveys

Heather Ridolfo, Ph.D.
Kristen Miller, Ph.D.
Aaron Maitland, MS
National Center for Health Statistics
Questionnaire Design Research Laboratory

Background

- Need to understand health disparities among sexual minority groups
- Development of sexual identity question for the National Health Interview Survey

Patterns of poorer health have been noted for lesbian, gay, bisexual, and transgendered people, leading to the establishment of a reduction of health disparities by sexual identity groups as a federal priority. To meet this objective we need a valid measure of sexual identity on our federal health surveys.

Sexual identity, however, is complex concept that is not easy to measure on a survey questionnaire. In this presentation I am going to give an overview of the work that the Questionnaire Design Research Laboratory at NCHS has done over the last several years to evaluate our current sexual identity measures and the work we are currently doing to develop a new, more valid measure of sexual identity.

Known challenges to validity

- Comprehension of terms
- Commonality of terms
- Fluidity of sex/gender
- Fluidity of SI
- Current definitions add confusion
- “Other” categories
- Sub-cultural effects on question response process

Over the years we have conducted a number of studies to evaluate different measures of sexual identity. To evaluate these measures we used cognitive interviewing, which is a qualitative method used widely in the survey methodology field to identify potential response errors related to question design, as well as socio-cultural factors that might impact the question response process. And based on these studies we have drawn the following conclusions. First, we know from cognitive interviewing that some groups of the population have difficulty comprehending terms used to describe sexual identities. In particular, less-educated, and Hispanic respondents are often unfamiliar with the terms heterosexual and bisexual.

Additionally, the terms used in these questions are not universal. Meaning that different subgroups of the population use different labels to describe themselves. For example, we have evidence from cognitive interviews that some subgroups of the population prefer terms such as queer, to gay or homosexual, while others prefer heterosexual over straight, or conversely straight over heterosexual.

The way sexual identity is currently measured is also gender dependent. Meaning that the questions that we currently use presume the respondent is either male or female, and that their sex or gender is static over time. For respondents' who sex is ambiguous, or those who have transitioned genders over time it can be difficult for these respondents to select a response from the responses that are traditionally provided.

Just as one's gender can be fluid over time, so can one's sexual identity. Respondents in cognitive interviews sometimes wished to report that their sexual identity had evolved over time, a single question will not capture this.

In order to accommodate for some of these comprehension and reporting issues survey designers have done a couple of things. First, in order to address comprehension problems, some surveys have added the term straight to the heterosexual category. This seems to help comprehension somewhat. Additionally, some surveys have added definitions to the response categories. However, this practice of adding definitions can create more harm than good. This was found to be the case during the evaluation of one question, which used attraction to define sexual identity. Some respondents that we interviewed felt attracted to the same sex, but did not identify as gay or lesbian, and therefore they found this question difficult to answer.

Survey designers have also added categories such as “other” “something else, “not sure” to their sexual identity questions to address problems caused because of a lack of common terminology used to express one's sexual identity, and the fluidity of sexual identities over time. However, during cognitive testing we found this is not necessarily a sound solution. Instead, what ends up happening is respondents who do not understand the terms within the question, respondents who prefer alternative terms to describe their sexual identity, respondents who are transgendered, and respondents who are in a transitional identity end up selecting these other categories,, leaving researchers with a hodgepodge if you will of respondents, and these “other” categories are not very useful from an analytic standpoint.

Finally, we have found that problems that I just identified are not random but rather systematic. Meaning that these types of problems are more likely to occur among less educated respondents, those who are transgendered, and some racial and ethnic minority groups, particularly Hispanic women.

Findings from NSFG & NHANES

- High missing data rates for the sexual identity question
- Missing data is systematic

A couple of the questions that we have evaluated over the years have been fielded on our health surveys. Using data from these surveys we were able to examine the magnitude of some of the problems that we uncovered in the cognitive interviews. One of these problems that we were able to examine was the prevalence of respondents who were not able to fit themselves into the categories of sexual identity provided. In other words, the prevalence of missing data for these questions. So using data from the National Survey of Family Growth and the National Health and Nutrition Examination Survey, we examined the prevalence of missing data. Responses of “something else” “not sure” “don’t know” or “refused” were all coded as missing data. What we found was that we have high missing data rates for this measure on both surveys, and the missing data was once again not random, but rather systematic.

NSFG Cycle 6 (2002)

Do you think of yourself as...

Heterosexual

Homosexual

Bisexual

or Something else

The first question we examined was from the 2002 National Survey of Family Growth. This question asked: Do you think of yourself as heterosexual, homosexual, bisexual, or something else. From the cognitive tests we know this question is particularly problematic as less-educated respondents report being unfamiliar with the term heterosexual and bisexual, and therefore report that their sexual identity is something else. Additionally, some respondents who are gay do not identify with the term homosexual and therefore will answer “something else” and finally respondents who are transgendered do not feel their experiences are adequately captured with these traditional response options and therefore also select the something else category.

NSFG Cycle 6 (2002)

Table 1. Sexual identity by gender.

| Sexual Identity | Men | Women | All |
|-----------------|------------|------------|------------|
| Heterosexual | 90.1% | 90.1% | 90.4% |
| Homosexual | 2.3 | 1.3 | 1.8 |
| Bisexual | 1.9 | 2.5 | 2.2 |
| Missing | <u>5.7</u> | <u>5.4</u> | <u>5.6</u> |
| | 100 | 100 | 100 |
| N | (3574) | (6109) | (9683) |

Note. Missing data= something else, refused, and don't know responses

The first thing we did was look at the distribution of sexual identities for men, women, and all respondents. As you can see the distributions of sexual identities are fairly similar across men and women, but the finding that is most alarming is that there is missing data for 5.6% of the sample.

NSFG Cycle 6 (2002)

Table 2. Sexual identity by gender.

| Sexual Identity | Men | Women | All |
|-----------------|------------|------------|------------|
| Heterosexual | 95.5% | 96.0% | 95.8% |
| Homosexual | 2.5 | 1.4 | 1.9 |
| Bisexual | <u>2.0</u> | <u>2.6</u> | <u>2.3</u> |
| | 100 | 100 | 100 |
| N | (3574) | (6109) | (9683) |

This next table shows what the distribution looks like when we drop missing data from the sample, which is how most data users will likely use this measure. As you can see in dropping the missing data, the percentage of heterosexuals in the sample becomes inflated. We went from about 91% of the sample is heterosexual to about 96%. So using this measure we are losing a significant percentage of the sample due to missing data, and our percentage of homosexuals and bisexuals remain small.

2002-2008 NHANES

Do you think of yourself as...

Heterosexual or straight (attracted to men/women)

Homosexual or lesbian/gay (attracted to women/men)

Bisexual (attracted to men and women)

Something else

Not sure

Similar patterns with found when we looked at data from the National Health and Nutrition Examination Survey. The question used on the NHANES survey asks: Do you think of yourself as heterosexual or straight, homosexual or lesbian/gay, bisexual, something else, or not sure. And as you can see here they provide definitions with these categories which focus solely on attraction. We know from our cognitive tests that this is problematic as individuals who are attracted to the same or opposite sex do not necessarily identify as gay or straight respectively and thus have difficulty responding to this question. The same comprehension problems the I described earlier with the response options were also found with this question. One good thing about this question though is they did add the term "straight" to the heterosexual category, as this was the term this group of respondents most preferred during cognitive testing. This should improve comprehension.

2002-2008 NHANES

Table 3. Sexual identity by gender.

| Sexual Identity | Men | Women | All |
|---------------------------|------------|------------|------------|
| Heterosexual or straight | 95.4% | 93.2% | 94.3% |
| Homosexual or gay/lesbian | 2.3 | 1.5 | 1.9 |
| Bisexual | 1.3 | 3.3 | 2.3 |
| Missing | <u>1.1</u> | <u>2.0</u> | <u>1.6</u> |
| | 100 | 100 | 100 |
| N | (3697) | (4201) | (7898) |

Note: Missing data = something else, not sure, don't know, refused and don't know responses

As you can see here the missing data rates are much smaller than the missing data rates found in the NSFG data.

2002-2008 NHANES

Table 4. Sexual identity by gender.

| Sexual Identity | Men | Women | All |
|---------------------------|------------|------------|------------|
| Heterosexual or straight | 96.4% | 95.1% | 95.7% |
| Homosexual or gay/lesbian | 2.3 | 1.5 | 1.9 |
| Bisexual | <u>1.3</u> | <u>3.4</u> | <u>2.4</u> |
| | 100 | 100 | 100 |
| N | (3632) | (4087) | (7719) |

And as with the NSFG, when we drop the respondents with missing data from the sample the percentage of heterosexuals in the sample increases slightly, while the percentage of gays/lesbians, and bisexual remains about the same. But what is important about this missing data is that in both the NSFG and the NHANES this missing data was not random. I'm going to demonstrate what I mean by this next, using data from the NHANES.

2002-2008 NHANES

Table 5. Distribution of missing data by education

| Missing data | Men | | | Women | | |
|--------------|-----------------------------|-------|--------|-----------------------------|-------|--------|
| | <HS | HS | >HS | <HS | HS | >HS |
| Yes | 3.4% | 0.8% | 0.4% | 5.7% | 2.1% | 1.1% |
| No | 96.7 | 99.2 | 99.6 | 94.3 | 98.0 | 99.0 |
| | 100 | 100 | 100 | 100 | 100 | 100 |
| | (976) | (995) | (1724) | (976) | (918) | (2306) |
| | Chi-square = 42.34, p<.0001 | | | Chi-square = 50.29, p<.0001 | | |

As you recall from the cognitive interview findings, less educated respondents were more likely to have comprehension problems with this question. And as you can see here, education influences the percentage of missing data for this question. For both men and women the amount of missing data decreases as education increases.

2002-2008 NHANES

Table 6. Distribution of missing data by gender and ethnicity.

| Missing data | Men | | Women | |
|--------------|-----------------------------|--------------|-----------------------------|--------------|
| | Hispanic | Non-Hispanic | Hispanic | Non-Hispanic |
| Yes | 3.4% | 0.6% | 5.8% | 1.3% |
| No | 96.6 | 99.4 | 94.2 | 98.7 |
| | 100 | 100 | 100 | 100 |
| N | (1103) | (2594) | (1247) | (2954) |
| | Chi-square = 42.10, p<.0001 | | Chi-square = 46.06, p<.0001 | |

Note. Missing data = something else, not sure, refused, and don't know responses

Additionally, during the cognitive interviews we found Hispanic women were more likely to report that they were not familiar with the terms used in this question than other groups. Using the NHANES data we were able to look at the magnitude of this problem. What we found was missing data was high among both Hispanic men and women. As shown in Table 6, missing data rates are significantly different between Hispanics and Non-Hispanics. The percentage of missing data is more than five times higher for Hispanic men compared to Non-Hispanic men. The percentage of missing data is almost 4.5 times higher for Hispanic women compared to Non-Hispanic women.

2004-2008 NHANES

Table 7. Distribution of missing data by language and ethnicity (for men only).

| Missing data | Hispanic English interview | Hispanic Spanish interview | Non-Hispanic English interview |
|--------------|----------------------------------|----------------------------------|--------------------------------------|
| Yes | 0.7% | 5.5% | 0.7% |
| No | <u>99.3</u> | <u>94.5</u> | <u>99.3</u> |
| | 100 | 100 | 100 |
| N | (373) | (434) | (1965) |

Note. Chi-square = 70.13, $p < .0001$

In addition to ethnicity, language also influences the percentage of missing data. As illustrated in Table 7 the percentage of missing data is nearly eight times higher for Hispanic men who completed the interview in Spanish compared to others who completed the interview in English.

2004-2008 NHANES

Table 8. Distribution of missing data by language and ethnicity (for women only).

| Missing data | Hispanic English interview | Hispanic Spanish interview | Non-Hispanic English interview |
|--------------|----------------------------------|----------------------------------|--------------------------------------|
| Yes | 3.3% | 9.9% | 1.4% |
| No | <u>96.7</u> | <u>90.1</u> | <u>98.6</u> |
| | 100 | 100 | 100 |
| N | (496) | (426) | (2169) |

Note. Chi-square = 43.85, $p < .0001$

Table 8 also shows that language influences the percentage of missing data among women. The percentage of missing data is significantly higher for women who completed the interview in Spanish compared women who completed the interview in English.

These same patterns were found in the 2002 NSFG.

Implications

- High missing data
- Inflated numbers for heterosexuals, small numbers of gays, lesbians, and bisexuals
- Systematically dropping less-educated, and Hispanics from the sample.
- Should not assume that those who are dropped from the sample are straight

What does this all mean? It confirms what I said at the start of this presentation – Sexual identity is a particularly complex concept and it is very challenging to develop a single measure that is both meaningful and comparable across the various socio-cultural groups in the US. However, because the LGBT population represents such a small minority, it is especially important that, as much as possible, that error be reduced (or at least understood) so that impending data can be accurately interpreted.

The measures we have now are producing high missing data rates, and furthermore this missing data is not evenly distributed across the sample of respondents. By dropping respondents with missing data from the sample we are systematically dropping particular groups of respondents, that is, those who are less educated and those who are Hispanic. This is problematic. In sum, there is more work that needs to be done to develop a better measure. Note: Anjani will present more recent data from NSFG which shows some improvement.

Moving forward...

- We are currently developing a sexual identity measure for the NHIS.
- ACASI

The QDRL is currently developing a new question that is proposed to be added to the National Health Interview Survey. Based on our many studies over the years we have determined that questions which ask about sexual identity are complex questions more than they are sensitive. Don't have any evidence it's sensitive from our qualitative work. Sexual identity is a complex concept because of the multiplicity of identities and how personal identity is tied to particular social locations and experience. It makes for a difficult question to ask in population surveys. Consequently we believe that asking this question in the traditional way of interviewer administered, it's not likely that we will not be able to get rid of the response error. Our goal is to develop a question for NHIS that can be used on an Audio-CASI instrument. Audio-CASI is a self-administered instrument, in which the respondent is administered the question on a computer screen while listening to a digitized recording of the question. Using this type of instrument will allow us to break down complex concepts in a consistent way for all respondents. Additionally we are creating definitions of the response options based on findings from cognitive interviews. These definitions presented on the Audio-CASI should improve question comprehension and response rates. Ultimately we would like the sexual identity measure on the NHIS to serve as a gold standard, to which other surveys can compare their estimates....

Do you currently consider yourself to be:

- Heterosexual or "Straight"
- Homosexual or Gay
- Bisexual
- Other
- Refused to answer
- Don't know

If R is unfamiliar with the terms "heterosexual" or "straight" say:
A heterosexual or straight man is someone who sees himself as forming intimate or loving relationships with women. He may also think of himself as being attracted to, or primarily attracted to, women. He may also engage in sexual behaviors with women.

If R is unfamiliar with the terms "homosexual or gay" say:
A homosexual or gay man is someone who sees himself as forming intimate or loving relationships with other men. He may also think of himself as being attracted to, or primarily attracted to, men. He may also engage in sexual behavior with other men. He might also form political, social, or cultural connections with a community of other homosexual men.

If R is unfamiliar with the term "bisexual" say:
A bisexual man is someone who sees himself as forming intimate or loving relationships with either men or women. He may also think of himself as being attracted to both men and women, and he may also engage in sexual behavior with either men or women.

Here is the male version of the question....

Do you currently consider yourself to be:

- Heterosexual or "Straight"
- Homosexual or lesbian
- Bisexual
- Other
- Refused to answer
- Don't know

If R is unfamiliar with the terms "heterosexual" or "straight" say:
A heterosexual or straight woman is someone who sees herself as forming intimate or loving relationships with men. She may also think of herself as being attracted to, or primarily attracted to, men. She may also engage in sexual behavior with men.

If R is unfamiliar with the terms "homosexual" or "lesbian" say:
A homosexual or lesbian woman is someone who sees herself as forming intimate or loving relationships with women. She may also think of herself as being attracted to, or primarily attracted to, women and she may engage in sexual behavior with women. She might also form political, social, or cultural connections with a community of other lesbians.

If R is unfamiliar with the term "bisexual" say:
A bisexual woman is someone who sees herself as forming intimate or loving relationships with either men or women. She may also think of herself as being attracted to both men and women, and she may also engage in sexual behavior with either men or women.

And the female version...

Moving forward...

- Conducting 75 English and 60 Spanish Interviews
 - Heterosexual at high risk for HIV
 - Men who have sex with men
 - Injecting drug users
- Will conduct additional 100 English and 100 Spanish interviews using ACASI

To evaluate this question we are currently conducted cognitive interviews in English and Spanish. This project has been piggybacked on another project, which focuses on HIV risk behavior and thus we were only able to test the sexual identity measure on these select groups. In the next round of testing we will conduct an additional 100 English and 100 Spanish interviews to cover populations not reached in the current study. Additionally, in the next round of testing we will administer the question on an ACASI system and evaluate the use of this method.

Questions?

Heather Ridolfo
hridolfo@cdc.gov