



# How to Analyze YRBS Sexual Minority Data

September 2024

---

Where can I get more information? Visit [www.cdc.gov/yrbs](http://www.cdc.gov/yrbs) or call 800-CDC-INFO (800-232-4636).

---



### Introduction

**Purpose** “How to Analyze YRBS Sexual Minority Data” describes best practices for analyzing data from the sexual identity and sex of sexual contacts questions on national, state, local school district, and territorial Youth Risk Behavior Surveys (YRBSs).

---

**Contents**

- Introduction
  - Purpose
  - Contents
  - Background
- Analytic Considerations
  - Selecting sexual minority measures
    - Sexual identity
    - Sex of sexual contacts
  - Software requirements
  - Minimum sample size
  - Selecting YRBS data

---

**Background** Sexual identity and sex of sexual contacts may both be used to identify sexual minority youth. Sexual minority youth defined by sexual identity include those who identify as gay, lesbian, or bisexual, and may include those who describe their sexual identity some other way and those who are not sure about their sexual identity (questioning). Sexual minority youth defined by sex of sexual contacts include those who have had sexual contact with only the same sex or with both sexes.

Questions ascertaining sexual identity and sex of sexual contacts have been included in some state and school district YRBSs since 1995. Beginning in 2015, a question asking about sexual identity and a question asking about sex of sexual contacts was added for the first time to the standard questionnaire used by the states and local school districts and to the national YRBS questionnaire. In 2021, the sexual identity question was revised to add a few new response options.

This document is not intended to teach about sexual orientation and adolescents, nor to teach statistical methods or inference techniques. This document assumes readers are familiar with the YRBS and with the concepts referenced, for example, weighting methods, statistical software and programming for analyzing YRBS data, and beginning descriptive analyses.

---

## How to Analyze YRBS Sexual Minority Data

---

Readers unfamiliar with YRBS data should review:

- the latest [national data documentation](#),
  - [Youth Risk Behavior Surveillance—United States, 2019](#)
  - [Sexual Identity, Sex of Sexual Contacts, and Health-Related Behaviors Among Students in Grades 9-12 – United States and Selected Sites, 2015](#)
  - [Methodology of the Youth Risk Behavior Surveillance System—2013](#),
  - [Software for Analysis of YRBS Data](#),
  - [Combining YRBS Data Across Years and Sites](#), and
  - the [Frequently Asked Questions](#) page on the YRBS website.
-

## Analytic Considerations

### Selecting a sexual minority question for analysis

Sexual identity and sex of sexual contacts are two components of sexual orientation. A third component, sexual attraction, is not ascertained on the standard YRBS questionnaire or the national YRBS questionnaire. Each of these three components are separate constructs and should be considered independently. Discordance between sexual identity and sex of sexual contacts occurs, particularly among youth. For example, some youth who have had sexual contact with only the same sex or with both sexes may identify as heterosexual and some youth who have had sexual contact with only the opposite sex may identify as gay, lesbian, or bisexual or be not sure of their sexual identity. In addition, some youth who identify as heterosexual, gay, lesbian, or bisexual and some youth who are not sure of their sexual identity may not have had any sexual contact.

Researchers using the YRBS sexual minority data should thoughtfully consider whether sexual identity or sex of sexual contacts is the appropriate measure of sexual minority status for each specific research question.

### *Sexual Identity*

For sexual identity, combining data across years may be problematic. It is recommended that analysts carefully note the changes to the response options before choosing an analytic strategy.

In the 2021 and 2023 YRBS, sexual identity was ascertained with the question, “Which of the following best describes you?” Response options were “heterosexual (straight),” “gay or lesbian,” “bisexual,” “I describe my sexual identity some other way,” “I am not sure about my sexual identity (questioning),” and “I do not know what this question is asking”. For analysis, the response options can be categorized as shown in Table 1.

**Table 1. Sexual minority variable from sexual identity.**

Which of the following best describes you?	Category
<i>Heterosexual (straight)</i>	Heterosexual
<i>Gay or lesbian</i>	Gay, lesbian, or bisexual
<i>Bisexual</i>	
<i>I describe my sexual identity some other way</i>	Identify some other way/questioning
<i>I am not sure about my sexual identity (questioning)</i>	
<i>I do not know what this question is asking</i>	Missing

In the 2015, 2017 and 2019 YRBS, sexual identity was ascertained with the same initial question wording as in 2021 and 2023, but response options were “heterosexual (straight),” “gay or lesbian,” “bisexual,” and “not sure”.

---

## How to Analyze YRBS Sexual Minority Data

---

Please note that the “not sure” response option in 2015-2019 surveys may not be equivalent to the “Identify some other way” and/or “Questioning” responses in the 2021 and 2023 surveys. While students who responded “not sure” in the previous surveys may have truly been unsure of their sexual identity, it is also possible that this response option was selected by students who simply did not understand the question or the other response options. Therefore, it may be inappropriate and difficult to combine data from this variable in the 2015-2019 surveys with data from the 2021 and 2023 surveys.

Researchers should carefully consider the implications of including students who responded “not sure” (2015-2019) or identify some other way/questioning (2021, 2023) along with those who identified as gay, lesbian, or bisexual, in an attempt to further increase cell sizes for sexual minority youth. Researchers who choose to treat not sure respondents as sexual minority youth should acknowledge this possible misclassification in their study limitations.

In any year sexual identity was measured, researchers might consider combining “gay or lesbian” with “bisexual” to create a “sexual minority” category to increase cell sizes. Researchers should note that the prevalence of health risk behaviors associated with sexual orientation may vary between gay/lesbian students and bisexual students, which might be important to acknowledge.

Compared to the sex of sexual contacts question, the sexual identity question yields a higher analytic sample size because more states and districts included this question in their questionnaires.

### *Sex of sexual contacts*

From 2015 to 2023, the questions used to measure sex of sexual contacts remained the same. Sex of sexual contacts was ascertained with the question, “During your life, with whom have you had sexual contact?” Response options were “I have never had sexual contact,” “females,” “males,” and “females and males.” No definition was supplied for sexual contact, so it is likely that students considered a range of sexual activities when responding to this question.

To create a sexual minority variable from the sex of sexual contacts question, it is necessary to cross sex with sex of sexual contacts. Table 1 shows this cross tabulation, and Figure 1 shows sample syntax for creating this variable in SAS. Sexual behavior questions require special consideration, as students with no sexual contact are excluded from analyses. Students who had sexual contact with the same sex only had different needs for birth control and condom use, and exclusions are applied for these students as noted in Table 2.

**Table 2. Sexual minority variable from sex and sex of sexual contacts.**

Sex	During your life, with whom have you had sexual contact?			
	<i>I have never had sexual contact</i>	<i>Females</i>	<i>Males</i>	<i>Females and males</i>
<i>Female</i>	No sexual contact*	Same sex only <sup>†,§</sup>	Opposite sex only	Both sexes
<i>Male</i>	No sexual contact*	Opposite sex only	Same sex only <sup>†</sup>	Both sexes

\* Excluded from analyses on sexual behaviors.

<sup>†</sup> Excluded from analyses on birth control use from 2015 to 2019. In 2021, the birth control question was revised and edit criteria were added to exclude students with same sex only contacts prior to data analysis (edit criteria are explained in the Data User’s Guide).

<sup>§</sup> Excluded from analyses on condom use.

**Figure 1. Sample SAS syntax for creating a sexual minority variable, “sexcontact” from sex and sex of sexual contacts.**

```

data one;
set yrbs2023;
*create four-level sex of sexual contacts;

if q2='1' then do;
  sex=2; *female;
  if q63='1' then sexcontact=4; *no sexual contact;
  else if q63='2' then sexcontact=2; *females/same;
  else if q63='3' then sexcontact=1; *males/opposite;
  else if q63='4' then sexcontact=3; *both males &
females;
end;

else if q2='2' then do;
  sex=1; *male;
  if q63='1' then sexcontact=4; *no sexual contact;
  else if q63='2' then sexcontact=1; *females/opposite;
  else if q63='3' then sexcontact=2; *males/same sex;
  else if q63='4' then sexcontact=3; *both males &
females;
end;
label sex="1: male 2: female"
sexcontact='1: opposite 2: same 3:both 4:no sexual contact';

run;

```

---

## How to Analyze YRBS Sexual Minority Data

---

It is possible to combine “same sex only” and “both sexes” in order to increase sample size. However, the prevalence of health risk behaviors associated with sexual orientation may vary between students who have had sexual contact with only the same sex and students who have had sexual contact with both sexes, which might be important to acknowledge.

Compared to the sexual identity question, the sex of sexual contacts question generally yields a lower analytic sample size because a sizeable proportion of students have had no sexual contact.

---

### Software Requirements

The YRBS uses a multi-stage cluster sample design. Statistical software used to analyze YRBS data should account for this design. Many packages with this capability are available.

Refer to [Software for Analysis of YRBS Data](#) for more information on five statistical packages (SUDAAN, SAS, STATA, SPSS, and Epi Info) that can be used to analyze YRBS data.

---

### Minimum Sample Size

The prevalence of sexual minority youth is low. As such, even in large, population-based samples like YRBS samples, the number of observations in sexual minority subgroups is low. It is important to be aware of denominator size whenever conducting analyses on YRBS sexual minority data. In general, CDC analyses of sexual minority YRBS data use 30 as the minimum denominator. Different cutoffs may be used for the number of observations in the denominator, but it is recommended to set a minimum denominator size *a priori* and to carefully consider results with relatively large standard errors or wide confidence intervals. Note that no minimum numerator size need be set. It is possible to have a numerator of 0 which is then reported as 0% with no confidence intervals.

The denominator may be further reduced if the research question calls for stratification by sex, race/ethnicity, or grade. To overcome this, one option is to combine data from multiple YRBSs (multiple years from the same site, multiple sites from the same year, or multiple years from multiple sites). Refer to [Combining YRBS Data Across Years and Sites](#) for more information about combining YRBS datasets. As stated before, the changes to sexual identity response options in 2021 might pose challenges when attempting to combine 2021 and 2023 data with previous years.

Although options to increase the analytic sample size of sexual minority students exist as described above, be aware that in some cases there simply will not be a large enough denominator for a robust analysis, particularly if stratifying the data is desired.

### **Selecting YRBS Data Sets**

Sexual minority questions were included in the national, state, local school district, and territorial YRBSs from 2015 to 2023 and have been included in some state, local school district, and territorial YRBSs since 1995. It is important to consider which data set(s) is most appropriate for each research question. The 2015 to 2023 national YRBSs are nationally representative and have a sample size around 15,000 students each; however, there may be a need for a larger analytic sample size. In that case, one solution is to combine data sets from many sites or combine many years of data from one or several sites. This method can increase the number of sexual minority observations, but it impacts the generalizability of the results. Researchers should consider the trade-off between sample size and generalizability when selecting which YRBS data sets to use. Please read [Combining YRBS Data Across Years and Sites](#) prior to merging data sets.

Data sets and documentation for the national YRBS are available online through [CDC's YRBS website](#). Additionally, many state, local school district, and territorial YRBS data sets are also available on the YRBS website, but some of these data sets are not available online. Access to data sets not available on the YRBS website must be requested from the sites. CDC can provide contact information as needed. To learn more about sexual minority data availability and for information about how to request state, district, territorial, and tribal government datasets, please visit the [YRBS FAQ website](#).

---