

Youth Risk Behavior Surveillance System (YRBSS)



Combining YRBS Data Across Years and Sites

April 2023

Where can I get more information? Visit www.cdc.gov/yrbss or call 800–CDC–INFO (800–232–4636).



**Centers for Disease
Control and Prevention**
National Center for HIV, Viral
Hepatitis, STD, and TB Prevention

Introduction

Purpose “Combining YRBS Data Across Years and Sites” describes best practices for combining data from multiple Youth Risk Behavior Surveys (YRBSs).

Contents

- Introduction
- Purpose
- Contents
- Background
- General Guidance

Combining Data from the Same Site for Multiple Years

Combining Data from Multiple Sites from the Same Year

Combining Data from Multiple Sites for Multiple Years

Background Reasons to combine YRBS datasets include:

- to increase analytic sample size,
- to increase the represented area or population (i.e., a region of the country), and
- to allow for analysis of changes in prevalence of a behavior across time (i.e., trend analysis).

YRBS data from the national, state, territorial, tribal, and school district (district) surveys can be combined in three ways:

1. from the same site for multiple years,
2. from multiple sites for the same year, and
3. from multiple sites for multiple years.

This document is not intended to teach statistical methods or inference techniques and assumes the reader is familiar with the concepts referenced, for example, weighting methods, analysis of complex survey data, and statistical software and programming for analyzing YRBS data.

Combining YRBS Data Across Years and Sites

Readers unfamiliar with YRBS data should review:

- the latest [national data documentation](#),
 - [Methodology of the Youth Risk Behavior Surveillance System—2013](#),
 - [Software for Analysis of YRBS Data](#),
 - [Interpretation of YRBS Trend Data](#),
 - [Conducting Trend Analyses of YRBS Data](#), and
 - the [Frequently Asked Questions](#) page on the YRBS website.
-

General Guidance

Do not combine unweighted data with weighted data. The discussion in this document applies only to weighted YRBS data.

Do not combine *national* YRBS data with state, territorial, tribal, or district data.

Align data across and within years and sites prior to beginning analysis. Because each YRBS questionnaire is unique by year and site, review all related individual year and site documentation. Variable names and numbers may not be compatible, formats may not be compatible, and questions will be different across and within years and sites. Questions and their response options may differ across and within years and sites.

Use caution when combining data across different types of sites, for example, combining state data with district data (e.g., combining Massachusetts and Boston data). Doing so may introduce some overlap in the population overall and among some students specifically that may be difficult to interpret.

Finally, account for the combination of YRBS datasets in the design/nesting statement of your complex survey design software. PSUs or stratum values may be the same for multiple sites and/or years of data, so the design/nesting statements should account for this. In SUDAAN, it is possible to add another level to the NEST statement such as:

NEST sitenumber stratum psu / PSULEV=3;

or

NEST year stratum psu / PSULEV=3;

or

NEST year sitenumber stratum psu / PSULEV=4;

SUDAAN requires variables in the NEST statement to be numeric. The variable “sitenumber” is not provided but can be created by converting the variable “sitecode” to numeric values as needed. Sites need not be numbered

Combining YRBS Data Across Years and Sites

consecutively, but the number must be unique for each site. For the year variable, either “year” or “survyear” may be used even if the same variable is also used later in the same SUDAAN procedure. Consult your survey software package documentation for the correct syntax.

Combining Data from the Same Site for Multiple Years

Appropriate YRBSs Combining data from the same site for multiple years is appropriate for national, state, territorial, tribal, and district YRBSs.

Examples:

- National YRBS 1991 through 2021
- Vermont YRBS 2001 through 2011
- Boston YRBS 2019 and 2021

Weighting Implications and Adjustments

The implications for survey weights are different when combining national data sets versus combining state, territorial, tribal, or district data sets. Additionally, for state, territorial, tribal, and district data sets, the implications are different depending on the type of analysis.

Combining national YRBS data: The national YRBS data are weighted to the sample size; that is, the weighted count of respondents is equal to the unweighted count of respondents. Therefore, there is no need to adjust the weights when combining multiple national YRBS datasets.

Combining state, territorial, tribal, or district YRBS data: State, territorial, tribal, and district YRBS data are weighted to represent the number of high school students in the jurisdiction; for example, the weighted sample size for the Vermont 2011 YRBS is approximately equal to the number of high school students in Vermont in 2011.

Weights will need to be adjusted depending on the analysis planned for state, territorial, tribal, and district data.

- *For trend analysis:* If the combined dataset is to be analyzed by year, there is no need to adjust the survey weights in the datasets.
- *For an increase in sample size:* If the data sets are combined to increase the sample size, then the weights for each year of data should be divided by the number of data sets combined. For example, for a combination of Boston 2019 and 2021, divide the student weights from each year by two to get the average population of students across years 2019 and 2021. For Boston 2017, 2019, and 2021, divide the student weights by three to get the average population of Boston students across years 2017, 2019, and 2021.
- In all examples above, be sure to add a year variable to the NEST statement in SUDAAN.

Combining Data from Multiple Sites for the Same Year

Appropriate YRBSs Combining data from multiple sites for the same year is appropriate for state, territorial, tribal, and district YRBSs. It is never appropriate to combine national data with state, territorial, tribal, and/or district data.

Examples:

- Kentucky 2021, Illinois 2021, Michigan 2021, Colorado 2021.
- Boston 2019, Washington, D.C. 2019, New York City 2019.

Weighting Implications and Adjustments There is no need to adjust the weights. Each state, territory, tribal, and/or district YRBS is weighted to be representative of its jurisdiction. Thus, the combined dataset is representative of students in the combined jurisdictions. Be sure to add a numerical variable to the NEST statement in SUDAAN that uniquely identifies each site.

Combining Data from Multiple Sites for Multiple Years

Appropriate YRBSs Combining data from multiple sites for multiple years is appropriate for state, territorial, tribal, and district YRBSs. It is never appropriate to combine national data with state, territory, tribal, and/or district data.

Examples:

- Kentucky 2019 and 2021, Illinois 2019 and 2021, Michigan 2019 and 2021, Colorado 2019 and 2021.
- Boston 2019 and 2021, Washington, D.C. 2019 and 2021, New York City 2021.

Weighting Implications

- *For trend analysis:* If the data are to be analyzed by year, there is no need to adjust the survey weights in the combined data set.
 - *For an increase in sample size:* The weights for each site should be divided by the number of years of data being used from that site. In the example above, for Boston and Washington DC 2019 and 2021, divide the student weights in each data set by two to get the average population of Boston and Washington DC students across years 2019 and 2021.
 - Because only one year of New York City data is being used, there is no need to adjust the survey weights for students in the New York City data set.
 - In the SUDAAN NEST statement, be sure to add a year variable and a numerical variable that uniquely identifies each site.
-