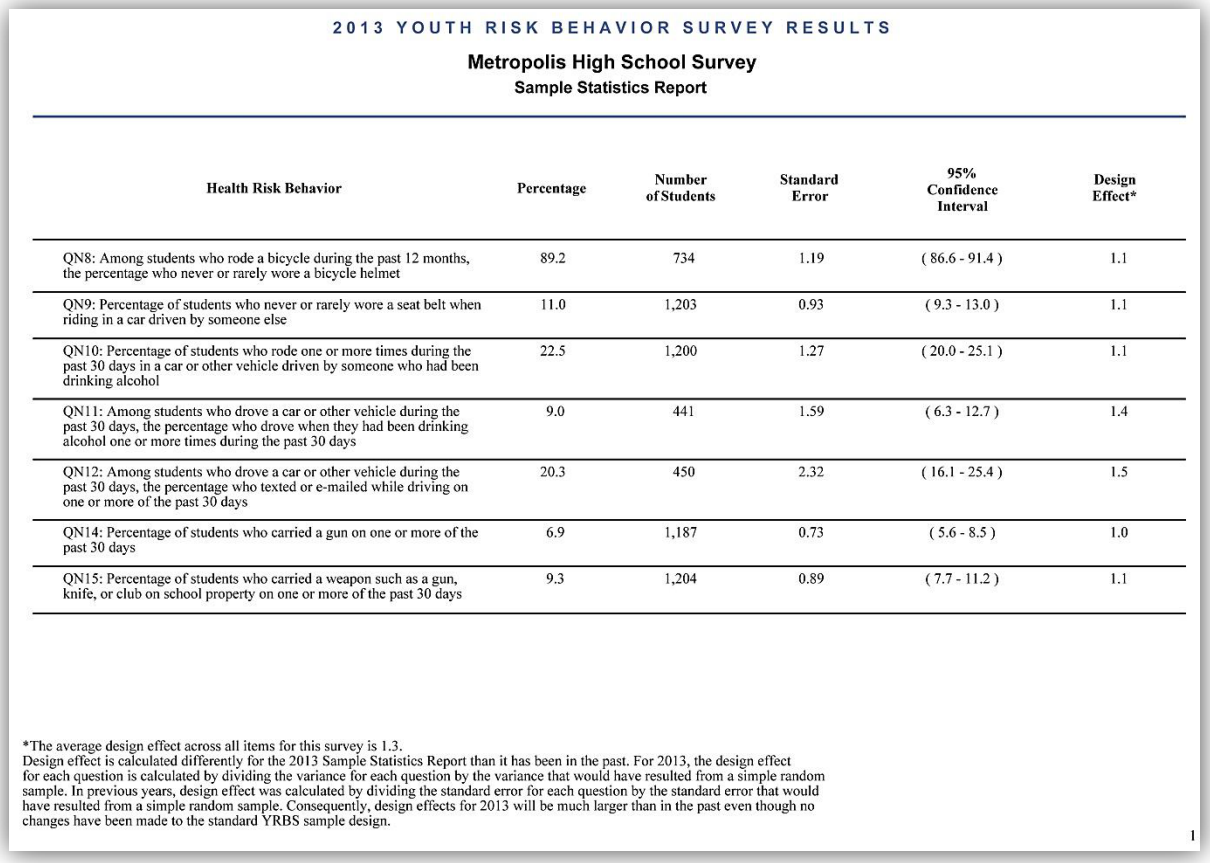
Sample Statistics Report Documentation

Purpose

The Sample Statistics Report is used to evaluate the efficiency of the sample design and the precision of your YRBS results.

Example



2013 YOUTH RISK BEHAVIOR SURVEY RESULTS

Metropolis High School Survey

Sample Statistics Report

Health Risk Behavior	Percentage	Number of Students	Standard Error	95% Confidence Interval	Design Effect*
QN8: Among students who rode a bicycle during the past 12 months, the percentage who never or rarely wore a bicycle helmet	89.2	734	1.19	(86.6 - 91.4)	1.1
QN9: Percentage of students who never or rarely wore a seat belt when riding in a car driven by someone else	11.0	1,203	0.93	(9.3 - 13.0)	1.1
QN10: Percentage of students who rode one or more times during the past 30 days in a car or other vehicle driven by someone who had been drinking alcohol	22.5	1,200	1.27	(20.0 - 25.1)	1.1
QN11: Among students who drove a car or other vehicle during the past 30 days, the percentage who drove when they had been drinking alcohol one or more times during the past 30 days	9.0	441	1.59	(6.3 - 12.7)	1.4
QN12: Among students who drove a car or other vehicle during the past 30 days, the percentage who texted or e-mailed while driving on one or more of the past 30 days	20.3	450	2.32	(16.1 - 25.4)	1.5
QN14: Percentage of students who carried a gun on one or more of the past 30 days	6.9	1,187	0.73	(5.6 - 8.5)	1.0
QN15: Percentage of students who carried a weapon such as a gun, knife, or club on school property on one or more of the past 30 days	9.3	1,204	0.89	(7.7 - 11.2)	1.1

*The average design effect across all items for this survey is 1.3.
Design effect is calculated differently for the 2013 Sample Statistics Report than it has been in the past. For 2013, the design effect for each question is calculated by dividing the variance for each question by the variance that would have resulted from a simple random sample. In previous years, design effect was calculated by dividing the standard error for each question by the standard error that would have resulted from a simple random sample. Consequently, design effects for 2013 will be much larger than in the past even though no changes have been made to the standard YRBS sample design.

1

Inclusion Criteria

Sites with weighted data in 2013 receive the Sample Statistics Report. All 2013 standard questions and site added questions are included in the report.

Content

The Sample Statistics Report contains the following columns:

Column	Content
Health Risk Behavior	Every question on the questionnaire.
Percentage	Weighted prevalence estimate
Number of Students	The unweighted count of students
Standard Error	The standard error for each prevalence estimate.
95% Confidence Interval	Lower and upper limits of the 95% confidence interval for each prevalence estimate.
Design Effect	The variance calculated from your YRBS sample design divided by the variance calculated from a simple random sample.

Important Change for 2013

Design effect is calculated differently for the 2013 Sample Statistics Report than it has been in the past. For 2013, the design effect for each question is calculated by dividing the variance for each question by the variance that would have resulted from a simple random sample. In previous years, design effect was calculated by dividing the standard error for each question by the standard error that would have resulted from a simple random sample. Consequently, design effects for 2013 will be larger than in the past even though no changes have been made to the standard YRBS sample design.

How to Interpret the Sample Statistics Report

Overview: The results from surveys that use two-stage cluster samples such as the YRBS are affected by two types of errors: non-sampling error and sampling error. Nonsampling error usually is caused by procedural mistakes or random or purposeful student errors. For example, nonsampling errors would include surveying the wrong school or class, data entry errors, and poor question wording. While numerous measures were taken to minimize nonsampling error for the YRBS, nonsampling errors are impossible to avoid entirely and difficult to evaluate or measure statistically.

In contrast, sampling error can be described and evaluated. The sample of students selected for a YRBS is only one of many samples of the same size that could have been drawn from the population using the same design. Each sample would have yielded slightly different results had it actually been selected. This variation in results is called sampling error and it can be estimated using data from the sample that was chosen for your YRBS. The statistics in this report describe the sampling error for your YRBS.

If simple random sampling had been used to select students for your YRBS, it would have been possible to use straightforward formulas for calculating sampling errors. However, your YRBS sample design uses two-stage cluster sampling; thus it is necessary to use more complex formulas. Software packages like SUDAAN can be used to compute sampling errors.
Standard Error: A measure of sampling error that is often reported is the "standard error" of a particular statistic (mean, percentage, etc.), which is equivalent to the square root of the variance of the statistic across all possible

samples of equal size and design. Standard errors should be interpreted relative to the size of the observed prevalence estimate. Smaller standard errors indicate better precision of the prevalence estimate, while larger standard errors indicate less precision.

95% confidence interval: The 95% confidence interval uses standard error and sample size to determine a range around the observed prevalence estimate. This range is constructed in a way that indicates how likely it is to capture the true value of the percentage. If prevalence estimates were calculated for multiple samples of the population, we would expect that 95% of the 95% confidence intervals would encompass the true value of the population prevalence, and 5% of the 95% confidence intervals would not.[2] The width of the confidence interval gives an indication of the precision around the prevalence estimate; a very wide confidence interval indicates that the sample size should be increased to get a more precise estimate, while a narrow confidence interval indicates that the true prevalence is probably close to the prevalence estimate obtained from the sample. Confidence intervals may be asymmetric, particularly when the prevalence estimate is very small or very large.

Design Effect: This statistic can be used to assess the gain or loss of precision of prevalence estimates incurred by the sample design rather than a simple random sample. A design effect value of one indicates that the sample design is as efficient as a simple random sample; a value greater than one indicates a tendency for greater sampling error due to the use of a more complex and less statistically efficient design. Since the YRBS uses clusters of students in schools and classes to lower survey costs and reduce survey burden, it is not surprising that the design effect for prevalence estimates generated from YRBS data is much greater than one.

Notes

SUDAAN computes rates, means, or totals and their standard errors from the data collected in a complex multistage sample survey. The statistical approach used for computing the standard errors is a first-order Taylor Series linear approximation of the deviations of estimates from their expected values. For more details on the Taylor method, see Woodruff (1971). [3]

References

1. Research Triangle Institute. SUDAAN,® version 9.0.1 [software and documentation]. Triangle Park, NC: Research Triangle Institute; 2005.

2. Rosner, B.A. (2000) Fundamentals of Biostatistics, 5th ed, p. 182. Cambridge, MA: Harvard University Press

3. Woodruff, R.S. (1971). "Simple method for approximating variance of a complicated estimate." *Journal of the American Statistical Association* 66: 411-414.
