Methods

Sample Selection

Respondents were selected using a two-stage cluster sampling design to obtain a representative sample of public school students at the state level in grades 6 through 8 and grades 9 through 12. The selection process was coordinated by the Centers for Disease Control and Prevention (CDC), Office of Smoking and Health (OSH), and the organization working with them, Northrop-Grumman Information Technology. In the first stage, sample frames for public high schools and public middle schools were developed using 2009/10 enrollment data from the Pennsylvania Department of Education enrollment files. Each school was listed with its enrollment by race, sex and grade. From these sample frames, 87 eligible high schools and 87 eligible middle schools were selected, using a probability proportionate to size selection method. In the second stage, lists of classes were obtained from each participating school, and one to four classes were selected randomly from these class rosters.

Certain public schools were not included because of the challenges of working with the schools. These included juvenile detention schools, special education schools, part-time vo-tech schools, cyber schools and schools that had fewer than 30 students enrolled. Private and religious schools were not included in the survey.

Questionnaire and Survey Procedures

The Pennsylvania Youth Tobacco Survey (PA YTS) instrument uses a core set of 50 questions and 31 supplemental questions developed by CDC/OSH and implemented by most states, as well as nationally. By using these core questions, the analysis of Pennsylvania's youth tobacco use can be compared with other states. In addition to the core questions and supplemental questions, three questions about organ donation and dental visits were added to the core questions, and analyses of these questions are not included in this report. A copy of the questionnaire is included in Appendix A.

School districts and schools were sent recruitment letters in July 2010, and follow-up phone calls were subsequently made to the schools to encourage their participation.

Survey procedures were developed to protect students' privacy by allowing for anonymous and voluntary participation. Before the surveys were conducted, local parental permission procedures were followed. The survey was conducted by Department of Health primary contractors, who were provided training to conduct the surveys in the schools. Students completed a self-administered questionnaire in the classroom, recording their responses in the questionnaire booklet. No identifying information was collected.

School and Student Response Rates

	School			Student			Overall
	Selected	Participated	Participation Rate	Selected	Participated	Participation Rate	Participation Rate
Middle	87	66	75.9%	2,931	2,519	85.9%	65.2%
High	87	66	75.9%	2,844	2,320	81.6%	61.9%

Weighting and Analysis

A weight has been associated with each student's responses to reflect the likelihood of sampling each student and to reduce bias by compensating for differing patterns of nonresponse. The weight used for estimation is given by:

$$W = W1 * W2 * f1 * f2 * f3 * f4$$

W1 = the inverse of the probability of selecting the school
W2 = the inverse of the probability of selecting the classroom within the school
f1 = a school-level nonresponse adjustment factor calculated by school size category (small, medium, large).
f2= a class adjustment factor calculated by school
f3 = a student-level nonresponse adjustment factor calculated by class
f4 = a post stratification adjustment factor calculated by gender, race and grade

SUDAAN® was used to compute prevalence and 95% confidence intervals (CIs). The confidence intervals were used to determine the differences among subgroups at the $p<0.05$ level. Differences between prevalence estimates were considered statistically significant if the 95% confidence intervals did not overlap.

Percentages were not calculated and shown for subgroups of the population with sample size was less than 50. The method used to determine the reliability of percentages calculated from sample sizes of 50 or more consisted of comparison of the relative standard error of the calculated percentage with the relative standard error of the same percentage outcome for a simple random sample. If the relative standard error for the percentage being tested was smaller than the relative error of the same percentage outcome for simple random sample, then the calculated percentage was considered reliable.

Sample Characteristics

The table below compares the distribution by sex, grade and race/ethnicity of students who participated in the survey, with the distribution by the same demographics of the student enrollment for the 2009/10 school year, the most current school year for which enrollment was available.

	Middle School Survey Participants		High School Survey Participants		2009/10 Middle School Enrollment		2009/10 High School Enrollment	
	Number*	Percent	Number*	Percent	Number	Percent	Number	Percent
Male	1,258	50.2%	1,191	51.7%	205,912	51.4%	293,120	51.0%
Female	1,250	49.8%	1,112	48.3%	195,044	48.6%	281,471	49.0%
6th Grade	833	33.1%	-	-	131,579	32.8%	-	-
7th Grade	802	31.9%	-	-	133,412	33.3%	-	-
8th Grade	879	35.0%	-	-	135,965	33.9%	-	-
9th Grade	-	-	666	28.8%	-	-	149,704	26.1%
10th Grade	-	-	609	26.4%	-	-	147,507	25.7%
11th Grade	-	-	550	23.8%	-	-	139,767	24.3%
12th Grade	-	-	486	21.0%	-	-	137,613	23.9%
Black	308	12.4%	263	11.5%	60,768	15.2%	87,717	15.3%
Hispanic	122	4.9%	145	6.4%	30,533	7.6%	39,030	6.8%
White	1,875	75.5%	1,766	77.3%	293,281	73.1%	427,594	74.4%
Other	178	7.2%	110	4.8%	16,374	4.1%	20,250	3.5%
Total	2,519		2,320		400,956		574,591	

*Numbers do not add to total due to missing responses.
Note: Hispanic can be of any race; the racial groups of white and black exclude Hispanic ethnicity.

Limitations

This is a school-based survey and can only be generalized to the population of public school students in Pennsylvania. It cannot be generalized to home-schooled students, students who dropped out of school and students who are in special needs schools outside of their school districts.

Another limitation is that the data is self-reported, and students might not give honest responses about their behaviors and attitudes. It cannot be known whether students will over-report or under-report their behaviors. However, the survey questions have been analyzed and have been shown to be generally accurate.