Methods

Sample Selection

Respondents were selected using a two-stage cluster sampling design to obtain a representative sample of students at the state level in grades 6 through 8 and grades 9 through 12. The selection process was coordinated by the Center 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 2006-2007 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, 90 high schools and 84 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.

Questionnaire and Survey Procedures

The Pennsylvania Youth Tobacco Survey (PA YTS) instrument uses a core set of 83 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 that of other states. In addition to these core questions, 14 questions about diet and exercise were added to the survey. However, analyses of these questions are not included in this report, because the same questions were asked on the Youth Risk Behavioral Survey (YRBS), which was successfully completed in the spring of 2009 and is a more standard source for those data. A shorter instrument was used for middle school students. A copy of the questionnaires is included at the end of this report.

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

School district offices and schools were sent recruitment letters in June 2008, and follow-up phone calls were made in October 2008 to the schools to encourage their participation.

School and Student Response Rates

	Schools			Students			Overall
	Selected	Participated	Participation Rate	Selected	Participated	Participation Rate	Participation Rate
Middle	84	71	84.5%	2,829	2,384	84.3%	71.2%
High	90	76	84.4%	3,133	2,456	78.7%	66.2%

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 non-response. 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 non-response adjustment factor calculated by school size category (small, medium, large).
f2=	a class adjustment factor calculated by school
f3 =	a student-level non-response adjustment factor calculated by class
f4 =	a post stratification adjustment factor calculated by gender, race and grade

SUDAAN® was used to compute prevalence estimates and 95 percent 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 percent confidence intervals did not overlap.

During the analysis, if students answered that they did not know or were not sure, those students were removed from the calculations. If the question asked about smoking habits and allowed a response that the student did not smoke or that the question did not pertain to them, then those students were usually removed from the calculations. This is indicted by a phrase such as, 'of smokers', or 'of students who tried to quit.'

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 2006-2007 school year, the most current school year for which enrollment was available.

	Middle School Sample		High School Sample		2006-2007 Middle School Enrollment		2006-2007 High School Enrollment	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Male	1,208	50.7%	1,257	51.2%	220,162	51.5%	306,116	51.1%
Female	1,176	49.3%	1,199	48.8%	207,146	48.5%	293,284	48.9%
6th Grade	834	35.0%	-	-	137,238	32.1%	-	-
7th Grade	798	33.5%	-	-	143,683	33.6%	-	-
8th Grade	736	30.9%	-	-	146,387	34.3%	-	-
9th Grade	-	-	670	27.3%	-	-	161,416	26.9%
10th Grade	-	-	634	25.8%	-	-	155,550	26.0%
11th Grade	-	-	594	24.2%	-	-	145,260	24.2%
12th Grade	-	-	543	22.1%	-	-	137,174	22.9%
Black	338	14.2%	344	14.0%	69,593	16.3%	89,648	15.0%
Hispanic	221	9.3%	228	9.3%	28,754	6.7%	33,960	5.7%
White	1,682	70.6%	1,745	71.1%	317,684	74.3%	460,648	76.9%
Other	130	5.5%	131	5.3%	11,277	2.6%	15,144	2.5%
Total	2,384		2,456		427,308		599,400	

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 or students in special needs schools outside of their school districts.

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