August 2024

Sampling and Weighting Methodology for the 2024 New York Youth Tobacco Survey (NY YTS)

Prepared for

New York State Department of Health
Corning Tower, Room 1055
Albany, NY 12237-0676

Prepared by

[redacted]

Sampling Plan Overview

This document describes the sampling plan for the 2024 New York Youth Tobacco Survey (NY YTS). The 2024 NY YTS sample is designed to collect data from 71 responding schools and approximately 7,900 responding students across grades 6—12.

A probability-based study design will result in a set of respondents that are representative of New York's racially, ethnically, culturally, and geographically diverse student population. The sample design will yield an analytical dataset that will be used to make state-level population estimates and contain at least 500 responding students in each of the following domains defined by geographic region crossed with race/ethnicity category.[1]

- New York City and Non-Hispanic White/other race
- New York City and Non-Hispanic Black
- New York City and Hispanic
- Balance of state[2] and Non-Hispanic White/other race
- Balance of state and Non-Hispanic Black
- Balance of state and Hispanic

The sampling methodology is based on procedures developed by the Centers for Disease Control and Prevention (CDC) for the State Youth Tobacco Surveys (YTS). Some of the text in this document, starting with the section "Description of Sampling Methodology", was adapted from from the Youth Tobacco Survey (YTS) Methodology Report prepared for the CDC Office on Smoking and Health by Burton Levine.[3]

Frame Creation

The sample frame is a list of all eligible schools from which the 2024 NY YTS school sample will be drawn. Eligible students are those enrolled in grades 6—12 in New York State public or private schools. Each school in the frame has several pieces of information including:

- BEDSCode—a string of 12 digits, the first 2 identify the county, the first 6 identify the school district the last 6 identify the school
- School name
- Number of students by grade in the 2020-2021 school year
- Type of school—private or public
- County (FIPS code and name)
- For public schools—the percentage of total students that are Hispanic, Black/African American, and Asian American

[1] These are the same sample size goals applied to the 2022 NY-YTS study design.

[2] *Balance of state* contains all New Yorkers that do not reside in NYC.

[3] Office on Smoking and Health. State Youth Tobacco Survey (YTS) Methodology Report. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2018. This report was not published by the CDC, but can be found here: http://www.tacenters.emory.edu/documents/netconference_docs/SE2018/State%20YTS%20Methodology%20Report%202018

Private school frame
The private school data were downloaded from this webpage: http://www.p12.nysed.gov/irs/statistics/nonpublic/.
The file used is called "enrollment-by-gender-and-grade-nonpublic-2021-22.xlsx." In this file there are 1,444 private schools containing one or more 6th—12th grade students.

Public school frame
The public school data come from this webpage: https://data.nysed.gov/downloads.php.
The file used is called "ENROLL2019_20200228.acccdb." In this file there are 2,792 public schools containing one or more 6th—12th grade students.

Schools with fewer than 45 eligible students will not be sampled and are removed from the sampling frame prior to selecting the sample. The following are the advantages of removing schools with fewer than 45 eligible students from the frame. The first advantage is to improve the precision of the estimates. The section "Adjusting for Schools with Very Small Enrollment" describes increasing the size measure for schools with small enrollment so that the sampling fraction is equal for all students in a stratum. Keeping the sampling fraction within each stratum equal for all students reduces the weight variation and improves the precision of the estimates. Including schools with fewer than 45 students increases the need for this adjustment, which results in more schools being selected, with a probability that is not proportional to their size. Second, increasing the number of small schools requires collecting data on more students in the larger schools, which reduces precision and increases the burden in the larger schools. The cut-off of 45 students was selected to balance the coverage error with the improvements in efficiency. Not including schools with fewer than 45 students results in a coverage error of 0.3% students. This is a very small coverage error.

Table 1 displays the quantity of schools by stratum for all schools in the frame and the schools with 45 or more 6th—12th graders.

Table 1. Quantity of schools by stratum

Stratum			All schools with 1 or more 6th—12th grade students		Schools with 45 or more 6th—12th grade students	
Region	**Black %**	**Grades**	**N**	**%**	**N**	**%**
Balance of state	>40	N/A	178	4.2	154	4.3
	<40	6-8	1,223	28.9	894	25.0
	<40	9-12	535	12.6	502	14.0
	<40	6-8 and 9-12	582	13.7	478	13.4
New York City	N/A	6-8	867	20.5	742	20.7
	N/A	9-12	560	13.2	543	15.2
	N/A	6-8 and 9-12	291	6.9	266	7.4
Total			**4,316**	**100.0**	**4,236**	**100.0**

Description of the Sample Allocation

Overview

A sample is drawn using a stratified two-stage cluster sample design where, within strata, schools are selected with probability proportional to school enrollment size; classes within schools are selected so that, within strata, the overall probability of selection of each student is equal. Every eligible student in schools with 45 or more eligible students has a chance of being selected. We aim to obtain data from approximately 71 responding schools and approximately 7,900 responding students. We assumed a 60% school response rate, a 97.4% school eligibility rate, and a 78.2% student response rate. Consequently, we selected 122 schools.

$$Schools\ selected = \frac{(\frac{School\ goal}{School\ response\ rate})}{(School\ eligiblity\ rate)} = \frac{(71/0.6)}{(0.974)} = 121.5$$

Stratification

The sampling frame is partitioned into 7 strata. Six of the strata are created by crossing the region (New York City/balance of state) with the school composition (has middle school only, has high school only, has both high school and middle school). In the balance of state region, an addition stratum is created containing all schools that have more than 40% African-American students. Only schools with less than 40% African-American students are assigned to the remaining 3 balance of state strata. The stratum containing the high density African-American schools will be oversampled to obtain the responent goal of 500 balance of state African-American students. With this design there are schools selected in Richmond County, consequently, there is no need to create strata to force schools to be selected in Richmond County.

School and student totals in the eight strata

Table 2 displays the quantity of schools and students by stratum for the frame, after removing schools with fewer than 45 students.

Table 2. Quantity of schools and students by stratum

Stratum			Schools		Students	
Region	Black %	Grades	N	%	N	%
Balance of state	>40	N/A	154	4.3	59,228	3.7
	<40	6-8	894	25.0	303,406	19.2
	<40	9-12	502	14.0	406,336	25.7
	<40	6-8 and 9-12	478	13.4	161,952	10.2
New York City	N/A	6-8	742	20.7	227,278	14.4
	N/A	9-12	543	15.2	303,885	19.2
	N/A	6-8 and 9-12	266	7.4	117,942	7.5
Total			**3,650**	**100.0**	**1,580,492**	**100.0**

Response rate assumptions
We assume that the school response rate is 60% and the student response rate, within a responding school, is 78.2%. Also, we assume that 97.4% of the schools selected will be eligible. This estimate is based on response rates for the 2022 New York State YTS.

Quantity of schools selected by stratum
Table 3 describes the sample allocation of schools to strata. The sampling rate, denoted by f_i, is the ratio of schools in the sample to schools in the frame within stratum *i*.

$$f_i = \frac{n_i^s}{n_i^f}$$

Where n_i^s is the sample school count in stratum *i* and n_i^f is the frame school count in stratum *i*. The "over/under-sample" is the ratio of the sampling rate in a stratum to the overall sampling rate. In the first stratum, the "over/under-sample" is 1.71. This means the quantity of schools selected is 1.71 times a proportional allocation. The sampling rate was selected to result in approximately 500 Black/African American student respondents in the balance of state region. The school allocation was scaled to result in 71 responding schools (see the values listed under the heading "Sample" and the subheading "n before NR adj"). The sample was divided by 0.6 to account for school level nonresponse. Finally, the number of schools were rounded to the nearest whole number. The last column in Table 3 contains the school allocation by stratum.

You will notice that the sample has 63 schools in New York City and 59 schools in the balance of state region. We expect New York City schools will have a lower response rate than balance of state schools. The expectation is that there will be an approximately equal number of responding schools in New York City and balance of state regions.

Table 3. Allocation of school sample

Stratum			Frame				Sample		
Region	Black %	Grades	n	%	Sampling rate *(f)*	Over/ under-sample	%	n before NR adj	n after NR adj
Balance of state	≥40	N/A	154	4.3	0.058	1.71	7.4	9	5
	<40	6-8	894	25.0	0.027	0.79	19.7	24	14
	<40	9-12	502	14.0	0.028	0.82	11.5	14	8
	<40	6-8 and 9-12	478	13.4	0.025	0.74	9.8	12	7
NYC	N/A	6-8	742	20.7	0.042	1.23	25.4	31	18
	N/A	9-12	543	15.2	0.041	1.19	18.0	22	13
	N/A	6-8 and 9-12	266	7.4	0.038	1.10	8.2	10	6
Total			**3,579**	**100.0**			**100.0**	**122**	**71**

NR adj —nonresponse adjustment

Quantity of students selected by stratum
Table 4 describes the allocation of students sampled by stratum. The sampling rate, denoted by f_i, is the ratio of the sampled students to the frame count of students.

$$f_i = \frac{n_i^s}{n_i^f}$$

Where n_i^s is the sample count of students in stratum *i* and n_i^f is the frame count of students in stratum *i*. The "over/under sample" is the ratio of the sampling rate in a stratum to the overall sampling rate. In the first stratum, the "over/under-sample" is 2.02. This means the quantity of sample selected is 2.02 times a proportional allocation. The sampling rate was selected to result in approximately 500 Black/African American student respondents in the balance of state region. The sample allocation was scaled to result in approximately 7,900 responding students (see the values listed under the heading "Sample" and the subheading "sample n before NR adj"). The sample was divided by 0.69 (0.6 * 0.976 * 0.78) to account for the school level response (0.6), the school eligibility rate (0.976) and the student level response rate (0.78). Also, the sample was divided by 1.037; this accounts for the fact that the observed schools enrollment is smaller than the enrollment reported on the frame. Finally, the number of students was rounded to the nearest whole number. The last column in Table 4 contains the student allocation by stratum.

Table 4. Allocation of student sample

Stratum			Frame				Sample		
Region	Black %	Grades	n	%	Sampling rate *(f)*	Over/ under-sample	%	n before NR adj	n after NR adj
Balance of state	≥40	N/A	59,228	3.7	0.023	2.02	7.6	1,358	598
	<40	6-8	303,406	19.2	0.008	0.72	13.9	2,487	1,095
	<40	9-12	406,336	25.7	0.008	0.72	18.6	3,330	1,467
	<40	6-8 and 9-12	161,952	10.2	0.008	0.72	7.4	1,327	585
NYC	N/A	6-8	227,278	14.4	0.015	1.28	18.4	3,302	1,454
	N/A	9-12	303,885	19.2	0.015	1.28	24.6	4,415	1,945
	N/A	6-8 and 9-12	117,942	7.5	0.015	1.28	9.6	1,714	755
Total			**1,580,027**	**100.0**			**100.0**	**17,933**	**7,899**

NR adj —nonresponse adjustment

Description of Sampling Methodology

First Stage—selecting the schools
Within strata, schools are selected using systematic sampling with a random start and probabilities proportional to size (PPS). Probabilities of school selection are proportional to a measure of size (MOS) that is based on the enrollment of 6th—12th graders for each school. Except for very large and very small schools, the measure of size is exactly equal to the enrollment in the target grades. Prior to sampling, schools in the frame are sorted by stratum and in descending order of size.

Determining certainty schools
There were are no certainty school in the 2024 NY YTS. However, the decription of certainty schools is included to describe what might have been done.

Depending on the number of schools that are to be sampled, some very large schools could be selected with certainty. As each "certainty" school is selected, it is removed from the frame of eligible schools. Within each stratum, an initial sampling interval is calculated by dividing the sum of the enrollments of all the schools in the sample frame by the number of schools desired in the sample. Schools that have an enrollment greater than or equal to the initial sampling interval are treated as certainty schools and are removed from the sample frame. Certainty schools are always included in the sample. Each time a school is selected with certainty and removed from the frame, the sampling interval is recomputed based on the enrollment of the schools remaining on the sampling frame and on the number of schools remaining to be selected.

$$Revised\ sample\ interval_i = \frac{Total\ school\ enrollment\ in\ revised\ frame_i}{Adjusted\ number\ of\ schools_i}$$

Sampling of certainty schools continues until the enrollment of the largest school remaining in the frame is less than the revised sampling interval. At this point, sampling of certainty schools is complete.

Sampling Fraction for Certainty Schools
$f_{school} = 1$

Determining noncertainty schools
If more schools are needed in the sample after the certainty schools have been selected, they are sampled from the schools remaining in the frame. The sampling procedure for these "noncertainty schools" includes adjustments to the measure of size for schools that have very small enrollments. This procedure ensures that, within strata, each student has the same probability of selection for the sample and that this probability is equal to the overall sampling rate.

Noncertainty schools are selected using systematic sampling. Probabilities for these schools are proportional to school enrollment. Special adjustments are made to MOS for very small schools.

Adjusting for schools with very small enrollments
For noncertainty schools with very small enrollments, it is possible that the probability of selection based on enrollment is so small that the overall probability of selection for students in these schools may be less than the required overall rate. Specifically, this happens when the school enrollment is so small that, even if all students were selected with certainty, their probability of selection would not be equal to f_i, the overall probability in stratum *i*. Therefore, an adjustment is made to the measure of size

for small schools so that students from these schools will have the required overall probability of selection. Essentially, the probability of selection for the small schools is increased so that selecting students with certainty from these schools will match the overall probability of selection for students. This has the effect of slightly decreasing the probabilities of selection for the larger, noncertainty schools. The number of students sampled from the small schools is thus slightly larger than it would be ordinarily.

After the certainty schools have been removed from the sampling frame, within strata, the remaining schools are ordered according to decreasing enrollment. It is then necessary to determine a minimum measure of size (MINMOS) for each school. When a school's enrollment falls below $MINMOS_i$, the actual enrollment is replaced by $MINMOS_i$. The procedure for determining $MINMOS_i$ is relatively simple. For the first and largest school remaining in the frame $MINMOS_i = 0$ (remember the schools are sorted in descending order by size). Thereafter, each succeeding school will have:

$$MINMOS_i = \frac{f \times \ sum\ of\ the\ enrollment\ of\ all\ schools\ preceding\ school\ i}{number\ of\ schools\ remaining\ to\ be\ selected - \begin{pmatrix} f \times (number\ of\ schools\ in\ the\ adjusted\ frame - \\ the\ number\ of\ schools\ preceding\ school\ i + 1) \end{pmatrix}}$$

Here the subscript *i* indexes the order of the schools within a stratum. Remember, the schools are ordered in decreasing size. MINMOS is calculated separately for each stratum. The stratum indexing was not included to simplify the equation.

Selection of noncertainty schools is carried out using systematic sampling with a random start and an adjusted school sampling interval that uses a total enrollment based on the revised MOS_i.

$$Adjusted\ school\ sampling\ interval_i = \frac{Total\ enrollment\ based\ on\ revised\ MOS_i}{Number\ of\ noncertainty\ schools\ required_i}$$

The selection procedure uses implicit stratification based on school enrollment. This procedure helps to ensure that schools of varying sizes are selected and helps to control the precision of estimates.

Second Stage Sampling Fraction (Probability of selection within school)

For certainty schools all students have an overall probability of selection equal to the overall sampling rate.

$f_{class} = f$

For noncertainty schools, the within-school sampling probabilities are based on the adjusted school sampling interval and on the adjusted school probability of selection so that, within strata,

$$Adjusted\ school\ probability\ (f_{school}) = \frac{Adjusted\ school\ measure\ of\ size}{Adjusted\ school\ sampling\ interval}$$

and

$$Within - school\ interval = \frac{Adjusted\ school\ probability}{Overall\ sampling\ fraction}$$

The second stage probability for selecting classes within a school is calculated by dividing the overall sampling fraction by the school sampling fraction.

The product of the first and second stage sampling fractions equals the overall sampling rate.

This two-stage sampling procedure yields an overall sample with probability of selection for each student equal to the overall sampling fraction.

Class selection
Within each selected school, the within school sampling interval is applied to a random start.
For example, for a school with 987 students and a within school sampling interval of 12, the random start is a random number between 1 and the sampling interval (12), say, 4. The following classes are selected for the survey: 4, 16, 28, 40, 52, 64, … . (4 + 12 = 16, 16 + 12 = 28, etc.) The school coordinator orders the 8th, 10th, and 12th grade homeroom classes. If there are a total of 45 classes, the 4th, 16th, 28th, and 40th classes are selected based on the ordering. It is expected that only a portion of the total list of classes will exist and be available. [redacted] will not know how many classes are available beforehand, so, the sampling team will pick a larger number of classes than will be used.

Response rates
Table 5 displays the school, student, and overall response rates.

Table *5*. School, student, and overall response rates

Domain	schools selected	schools responded	School response rate (%)	Students selected	Students responded	Student response rate (%)	Overall response rate (%)
All Schools	122	78	63.9	10,671	8,537	80.0	51.1

Weighting

School Base Weight

Each school's base weight is the inverse of the product of the probabilities of selection. The following is the formula for $Prob_{i,j}$, the probability of selection for each school sampled.

$$Prob_{i,j} = \frac{\alpha_i * MOS_{i,\mathrm{j}}}{\sum_{\forall j} MOS_{i,j}}$$

Where α_i is the number of schools selected in stratum *i*.

$MOS_{i,\mathrm{j}}$ is the size measure (eligible students) in *j*th school in the *i*th stratum.

$\sum_{\forall j} MOS_{i,j}$ is the sum of the size measures in all the schools in the *i*th stratum.

The base weight assigned to each selected school is the inverse of $Prob_{i,j}$.

$$W_{i,j}^{School} = \frac{1}{Prob_{i,j}}$$

where $W_{i,j}^{School}$ is the base weight of the *j*th school, in the *i*th stratum.

School level nonresponse adjustment

The relationship between response propensity and the following school characteristics were investigated:

- stratum
- number of eligible students
- school has high school students
- school has middle school students
- NYC/rest-of-state
- Public/private
- School student population—percent black
- School student population—percent Hispanic
- School student population—percent White or another race

The only school characteristic that was corrected with response propensity was private/public; the p-value was 0.0008. Table 6 displays the nonresponse adjustment applied to the public and private schools

Table *6*. School, student, and overall response rates

Public / Private	Schools selected	Schools responded	School response rate (%)	Nonresponse adjusment
Public	108	76	70.4	1.42
Private	14	2	14.3	7

$$Adj_{i,j}^{School\ NR} = \begin{cases} 1.42 \text{ for public schools} \\ 7 \text{ for private schools} \end{cases}$$

Student Base Weight

The probability of selecting a class within each selected school is:

$$Prob_{i,j,k} = \frac{\beta_{i,j} * mean\ students\ per\ class_{i,j}}{MOS_{i,j}}$$

Where $Prob_{i,j,k} =$ the probability of selecting the k^{th} class, in the j^{th} school, in the i^{th} stratum.

$\beta_{i,j} =$ Number of classes selected in the j^{th} school, in the i^{th} stratum.

$MOS_{i,j} =$ The number of eligible students in the i^{th} stratum and j^{th} school.

All students are selected within a class, so the probability of selecting a student is equal to the probability of selecting the student's class.

The student weight is:

$$W_{i,j,k}^{Student} = \frac{1}{Prob_{i,j,k}}$$

Student nonresponse adjustment

Within weighting classes, the student nonresponse adjustment is the ratio of sampled students to responding students.

Within each stratum, the student nonresponse adjustment is the ratio of the students selected in the stratum to the students responding in the stratum.

$$Adj_i^{Student\ NR} = \frac{Students\ selected\ in\ stratum_i}{Students\ responding\ in\ stratum_i}$$

Table 7 displays, by stratum and overall, the number of students selected and responding, and the student nonresponse adjustment.

Table *7*. Student Nonresponse Adjustment

Stratum			Eligible students	Responding students	Nonresponse adjusment
Region	Black %	Grades			
Balance of state	≥40	N/A	1,055	812	1.30
	<40	6-8	1,037	881	1.18
	<40	9-12	2,230	1,833	1.22
	<40	6-8 and 9-12	998	845	1.18
NYC	N/A	6-8	1,829	1,541	1.19
	N/A	9-12	2,778	2,154	1.29
	N/A	6-8 and 9-12	744	471	1.58
Total			**10,671**	**8,537**	**N/A**

Multiply the nonresponse adjusted school weight with the nonresponse adjusted student weight

The nonresponse-adjusted student weight is the product of the school weight, the school nonresponse adjustment, the class weight, and the student nonresponse adjustment.

$$W_{i,j,k,l}^{NR\ adj\ student} = W_{i,j}^{School} * Adj_{i,j}^{School\ NR} * W_{i,j,k}^{Class} * Adj_{i}^{Student\ NR}$$

Where, $W_{i,j,k,l}^{NR\ adj\ student}$ is the nonresponse adjusted student weight, for the *l*th student, in the *k*th class, in the *j*th school, in the *i*th stratum.
The nonresponse adjusted student weight is the input weight used in the calibration procedure. The nonresponse adjusted student weight is sometimes referred to as the design weight.

Calibration
Calibration constrains the sum of the weights for specific groups to equal population totals obtained from a source external to the survey. Calibration can reduce coverage bias and nonresponse bias. The source of the calibration totals was the sampling frame. The creation of the sampling frame is described in the section titled *Frame Creation*.

The calibration categories are stratum, grade, sex, and private/public school by race. The calibration procedure constrained the sum of the weights to the marginal distribution of each calibration category. This type of calibration procedure is sometimes called proportional iterative fitting, and sometimes it is called raking.

The categories of the distribution used in the calibration are listed in Table 6. Private schools do not report race categories. Consequently, school type (private/public) was combined with race. Since private schools and public schools are in distinct calibration categories, the calibration constrains the sum of the weights to equal the number of students in both private and public schools. The same is true for the geographic regions balance-of-state and New York City. And the same is true for schools with only middle school students, schools with only high school students, and schools with both middle school and high school students.

Imputation
To apply calibration there cannot be any missing value for the calibration distributions. The following are the calibration distributions with the number and % missing, stratum (0, 0%), grade (63, 0.7%), sex (458, 5.4%), and race (121, 1.4%). Missing values of grade were imputed with the mode of the class. Missing values of sex and race were imputed with a hot-deck imputation method.

Table 6 contains the population and respondent distributions for the variables used in the calibration.

Table *6*. Population and respondent distributions for calibration variables

Distribution	Category			Population n	Population %	Respondents n	Respondents %
Stratum	Balance of state	Black ≥40		59,693	3.8	812	9.5
		Black <40	Grades 6-8	303,406	19.2	881	10.3
		Black <40	Grades 9-12	406,336	25.7	1,833	21.5
		Black ≥40	Grades 6-8 and 9-12	161,952	10.2	845	9.9
	New York City		Grades 6-8	227,278	14.4	1,541	18.1
			Grades 9-12	303,885	19.2	2,154	25.2
			Grades 6-8 and 9-12	117,942	7.5	471	5.5
Grade	6			222,379	14.1	863	10.1
	7			223,197	14.1	1,237	14.5
	8			220,560	14.0	1,494	17.5
	9			242,640	15.4	1,544	18.1
	10			235,580	14.9	1,380	16.2
	11			219,232	13.9	1,111	13.0
	12			216,904	13.7	908	10.6
Sex	Female			774,224	49.0	4,339	50.8
	Male			806,268	51.0	4,198	49.2
Public/ private school by race	Private school		195,126	12.3	174	2.0	5.7
	Public school	White NH	609,535	38.6	2,281	26.7	28.4
		Black NH	234,213	14.8	1,397	16.4	13.2
		Hispanic	367,218	23.2	3,222	37.7	33.1
		Asian NH	135,811	8.6	899	10.5	13.1
		Another race NH	38,589	2.4	564	6.6	6.6

NH—not Hispanic

Dataset containing the weights

The variables IMP_MALE and IMP_PRIVATE_RACE have the following coding:

```
proc format;
 value IMP_malef
  0="Female"
  1="Male";
 value IMP_private_racef
  1="Private school / all races"
  2="Public school / White NH"
  3="Public school / Black NH"
  4="Public school / Hispanic"
  5="Public school / Asian NH"
  6="Public school / Another race NH";
run;
```

Analysis

There are three variables that describe the study design, STRATUM, PSU, and WT_analysis. Their names are self-explanatory. The following is example SAS code that estimates the mean, 95% confidence interval, and design effect for the 4 outcomes (ever smoke cigarettes, currently smoke cigarettes, ever vape, currently vape) over two domains (middle school, high school).

```
data data_with_derived_vars;
 set datafile;
 by num_id;
 if       eversmk=1 then BL_eversmk=100;
  else if eversmk=2 then BL_eversmk=0;

 if       ever_vape3=1 then BL_ever_vape3=100;
  else if ever_vape3=2 then BL_ever_vape3=0;

 if 2<=days_vape3<=7 then BL_current_vape=100;
  else if ever_vape3=2 or days_vape3=1 then BL_current_vape=0;

 if 2<=dayssmoke<=7 then BL_current_smoke=100;
  else if eversmk=2 or dayssmoke=1 then BL_current_smoke=0;

 HS=IMP_grade in (9,10,11,12);
run;

proc surveymeans data= data_with_derived_vars;
 stratum stratum;
 cluster PSU;
 weight  WT_analysis;
 var     BL_eversmk BL_current_smoke BL_ever_vape3 BL_current_vape ;
 domain  HS;
run;
```

Table *6*. Example results

	Middle school		High school	
Outcome	Mean	95% CI	Mean	95% CI
Ever smoke	3.8	(2.4, 5.1)	8.7	(6.6, 10.8)
Current smoker	1.0	(0.3, 1.6)	2.5	(1.0, 4.0)
Ever vape	16.0	(12.6, 19.4)	28.4	(25.7, 31.1)
Current vaper	6.0	(4.5, 7.5)	13.0	(10.9, 15.3)