METHODOLOGY REPORT OF THE 2017 NATIONAL YOUTH TOBACCO SURVEY

Recommended Citation

Office on Smoking and Health. 2017 National Youth Tobacco Survey: 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, 2017.

PRINT ONLY

December 2017

For questions about this report, please email Sean Hu at shu@cdc.gov

Prepared for Centers for Disease Control and Prevention

Prepared by ICF, Rockville, Maryland

Centers for Disease Control and Prevention

Office on Smoking and Health

Atlanta, GA

December 2017

TABLE OF CONTENTS

APPENDICES

Chapter 1—NYTS Sampling Design

1.1 Overview of the National Youth Tobacco Survey (NYTS)

In conjunction with the State Youth Tobacco Survey (YTS), the National Youth Tobacco Survey (NYTS) was developed to provide the data necessary to support the design, implementation, and evaluation of state and national tobacco prevention and control programs (TCPs).[1,2] In addition, NYTS data supplement other existing surveys, such as the Youth Risk Behavior Surveillance System (YRBSS) by providing more comprehensive data on tobacco-related indicators for both middle school (grades 6–8) and high school (grades 9–12) students. Tobacco-related indicators included in the NYTS are: tobacco use (e.g., bidis, cigarettes, cigars, tobacco pipes, smokeless tobacco, snus, dissolvable tobacco products, hookahs, and electronic cigarettes); exposure to secondhand smoke; smoking cessation; minors' ability to purchase or obtain tobacco products; and, knowledge and attitudes about tobacco and familiarity with pro-tobacco and anti-tobacco media messages. NYTS data also serve as essential benchmarks against which TCPs can assess the extent of youth tobacco use. The NYTS provides multiple measures and data for six of the 20 tobacco-related Healthy People 2020 objectives (USDHHS, 2010): TU-2, TU-3, TU-7, TU-11, TU-18 and TU-19.

First conducted during fall 1999 and again during the springs of 2000, 2002, 2004, 2006, and 2009, then annually starting in 2011, the NYTS provides data that are representative of all middle school and high school students in the 50 states and the District of Columbia. Beginning in 2011, the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) have collaborated to administer the NYTS.

1.2 Overview of the 2017 NYTS Methodology

The 2017 NYTS employed a stratified, three-stage cluster sample design to produce a nationally representative sample of middle school and high school students in the United States. Sampling procedures were probabilistic and conducted without replacement at all stages and entailed selection of: 1) Primary Sampling Units (PSUs) (defined as a county, or a group of small counties, or part of a very large county) within each stratum; 2) Secondary Sampling Units (SSUs), (defined as schools or linked schools) within each selected PSU; and 3) students within each selected school. Participating students completed the survey via pencil and paper using a self-administered, scannable questionnaire booklet.

Participation in the NYTS was voluntary at both the school and student levels. At the student level, participation was anonymous. CDC's Institutional Review Board (IRB) requires that parents be given the opportunity to opt their student out of participating in the survey. Schools used either passive or active permission forms at their discretion.

[1] Centers for Disease Control and Prevention. (CDC) (2014). *Best Practices for comprehensive tobacco control programs-2014*. Atlanta, GA: US Department of Health and Human Services, Public Health Service, CDC.

[2] Centers for Disease Control and Prevention. Surveillance and Evaluation Data Resources for Comprehensive Tobacco Control Programs. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.

The final sample consisted of 241 schools, of which 185 participated, yielding a school participation rate of 76.8%. A total of 17,872 student questionnaires were completed out of a sample of 20,144 students, yielding a student participation rate of 88.7%. The overall participation rate, defined as the product of the school-level and student-level participation rates, was 68.1%.

A weighting factor was applied to each student record to adjust for nonresponse and for varying probabilities of selection. Weights were adjusted to ensure that the weighted proportions of students in each grade matched national population proportions.

The remainder of this report provides detailed information on the methodology used in the 2017 NYTS sample selection (Chapter 2), data collection (Chapter 3), and weighting of student response data (Chapter 4).

Chapter 2—NYTS Sampling Methods

2.1 Sample Design

The objective of the NYTS sampling design was to support estimation of tobacco-related knowledge, attitudes, and behaviors in a national population of public and private school students enrolled in grades 6 through 12 in the United States. More specifically, the study was designed to produce national estimates at a 95% confidence level by school level (middle school and high school), by grade (6, 7, 8, 9, 10, 11, and 12), by sex (male and female), and by race/ethnicity (non-Hispanic white, non-Hispanic black, and Hispanic). Additional estimates also were supported for subgroups defined by grade, by sex, and by race/ethnicity, each within school level domain; however, precision levels varied considerably according to differences in subpopulation sizes.

The universe for the study consisted of all public and private school students enrolled in regular middle schools and high schools in grades 6 through 12 in the 50 U.S. states and the District of Columbia. Alternative schools, special education schools, Department of Defense-operated schools, Bureau of Indian Affairs schools, vocational schools that serve only pull-out populations, and students enrolled in regular schools unable to complete the questionnaire without special assistance, were excluded. The NYTS employed a repeat cross-sectional design.

The sample was a stratified, three-stage cluster sample. Primary sampling units (PSUs) were stratified by race/ethnicity and urban versus non-urban ("rural"). PSUs were classified as "urban" (U) if they were in one of the 54 largest Metropolitan Statistical Areas (MSAs) in the United States; otherwise, they were classified as "non-urban" (R). Within each stratum, PSUs, defined as a county, a portion of a county, or a group of counties, were chosen without replacement. Table 2.1 presents key sampling design features.

Table 2.1 Key Sampling Design Features

Sampling Stage	Sampling Units	Stratification	Measure of Size (MOS)	Designed Sample Size
1	PSUs: Counties, portions of a county, or groups of counties	Urban vs. Non-urban (2 strata); Minority concentration (8 strata)	Aggregate school size in target grades	85 Counties, portions of a county, or groups of counties
2	Schools	Small, medium and large; High school vs. middle school	Aggregate eligible enrollment	220 SSUs (school) selections: 170 large schools (2 per PSU), 20 medium schools and 30 small schools
3	Classes/students			2 classes per grade in large high-minority schools; 1 class per grade otherwise

The sample was designed to provide target sample sizes in the key analytic subgroups of interest. More specifically, the study was designed to produce national estimates at a 95% confidence level with a margin of error of 5% by school level (middle school and high school), by grade (6, 7, 8, 9, 10, 11, and 12), by sex (male and female), and by race/ethnicity (non-Hispanic white, non-Hispanic black, and Hispanic). With the average design effects attained in the NYTS, the requirements translated to subgroup sample sizes of 1,200 or more. Sample sizes were sufficient to generate estimates with the required precision by grade as well as by sex and school level. Therefore, the precision requirements generally focus on racial/ethnic subgroups. As shown in Section 2.8, the requirements are met for the two key racial/ethnic subgroups—Hispanics and non-Hispanic blacks—at both the middle and high school levels.

2.2 Sampling Frame

As in previous cycles, the 2017 NYTS sample was based on a comprehensive sampling frame from multiple data sources to increase the coverage of schools nationally. The frame combined data files obtained from MDR Inc. (Market Data Retrieval Inc.) and from the National Center for Education Statistics (NCES). The MDR frame contained school information that included enrollments, grades, race distributions within the school, district and county information, and other contact information for public and non-public schools across the nation. The NCES frame sources included the Common Core of Data for public schools and the Private School Survey for non-public schools. This dual-source frame build method was piloted first in 2014 to build the frame for the National Youth Tobacco Survey.[3] Including schools sourced from the two NCES files resulted in a coverage increase among all public and non-public high schools of 23%. Most of the added schools were smaller schools. Efforts were made to ensure that each school was represented only once in the final sampling frame, even if the school showed up in both source files.

Certain schools were removed from the frame prior to drawing the sample following a stepwise process. The first step excluded non-eligible schools by category to remove schools such as Department of Defense schools, vocational schools, and adult education schools. This resulted in the exclusion of 3.94% of schools (2.87% public and 8.16% private) and 0.68% of students. Next, schools were removed that had fewer than 40 students in the eligible grades, resulting in the exclusion of 14.53% of schools (8.26% public and 39.06% private) and 1.35% of students.

2.3 Sampling Units and Measure of Size

A three-stage cluster sample design was used to produce a nationally representative sample of students in grades 6–12 who attend public and private schools The first-stage sampling frame consisted of PSUs made up of counties, groups of smaller, adjacent counties, or parts of larger counties. For the second stage of sampling, secondary sampling units (SSUs) were defined as a physical school that can supply a full complement of students in grades 6 through 8 (middle school) or 9 through 12 (high school) or a school created by linking component physical schools together to provide all grades for the level.

[3] Redesigning National School Surveys: Coverage and Stratification Improvement using Multiple Datasets. William Robb, Kate Flint, Alice Roberts, Ronaldo Iachan, ICF International, FEDCASIC, March 2014

The sampling stages may be summarized as follows:

- Selection of PSUs—Eighty-five PSUs were selected from 16 strata with probability proportional to the total number of eligible students enrolled in all eligible schools located within a PSU.
- Selection of schools—At the second sampling stage, a total of 170 large schools or SSUs were selected from the 85 sample PSUs. The sample PSUs are subsampled to support the selection of small schools, 30 small schools from 15 subsample PSUs (one school for each level), and medium schools, 20 medium schools from 10 subsample PSUs (one school for each level). This resulted in a total of 220 SSUs (220=170+20+30). The PSU subsamples were selected with simple random sampling, and the schools were drawn with probability proportional to the total number of eligible students enrolled in a school.
- Selection of students—Students were selected via whole classes whereby all students enrolled in any one selected class were chosen for participation. Classes were selected from course schedules provided by each school so that all eligible students had only a single chance of selection.

Schools were stratified into small, medium, and large schools based on their ability to support less than one or two class selections per grade. Small SSUs contained fewer than 28 students at any grade level, medium SSUs contained between 28 and 55 per grade, and large SSUs contained at least 56 students at each grade level. Two classes per grade in large schools and one class per grade in the remaining schools were selected. The threshold for double class sampling was based on the simulation study to ensure that the required numbers of minority students (black and Hispanic students) were achieved per school level.

The sampling approach utilized PPS sampling methods with the measure of size (MOS) defined as the count of final-stage sampling units, students in intact classrooms. Coupled with the selection of a fixed number of units, the design resulted in an equal probability of selection for all members of the universe (i.e., a self-weighting sample). These conditions were approximated for the NYTS resulting in the attainment of a roughly self-weighting sample.

The MOS also was used to compute stratum sizes and PSU sizes. By assigning an aggregate measure of size to the PSU, the sample allocated to the PSU was in proportion to the student population.

The third, and final, sampling stage selected classes within each grade of a sample SSU. All students in a selected class were then selected for the survey.

2.4 PROJECTED SAMPLE SIZES

This section describes the planned sample sizes developed by the design, while Section 8 discusses the sample sizes actually attained in the survey. The NYTS sample size calculations were based on the following assumptions:

- The main structure of the sampling design is consistent with the design used to draw the sample for prior cycles of the NYTS.
- The design included the selection of two large SSUs within each sample PSU, and an additional 20 medium and 30 small schools from subsample PSUs.

Across 11 previous cycles of the NYTS that had concluded at the time of the 2017 NYTS design, school participation had averaged 84.7%, with a low of 72.5%. Student participation had averaged 82.8% with a low of 87.4%. The combined response rate (student x school) averaged 76.2%. Historical participation rates at both school and student levels guided the sampling design and sample sizes. In calculating the sample sizes, a combined rate of 72% was conservatively assumed. Table 2.2 presents a detailed derivation of the sample sizes *planned* for the 2017 NYTS based on these assumptions.

Table 2.2 Planned Sample Sizes for the 2017 NYTS

PSU	Size	# of SSUs	Number of Schools Sampled	# of Classes per School	# of Students per Class	# of Sampled Students prior to Attrition	Combined School X Student 72% Response Rate
85	Large HS	85	Double classes: 39	8	28	8,758	6,306
			Single classes: 46	4	28	5,141	3,701
	Large MS	85	Double classes: 39	6	28	6,569	4,730
			Single classes: 46	3	28	3,856	2,776
	Large Total	**170**				**24,324**	**17,513**
10 (sub-sample)	Medium HS	10	10	4	28	1,120	806
	Medium MS	10	10	3	28	840	605
	Medium Total	**20**				**1,960**	**1,411**
15 (sub-sample)	Small HS	15	15	3.8	18.1	1,030	742
	Small MS	15	15	2.8	11.2	475	342
	Small Total	**30**				**1,505**	**1,084**
	Overall Total	**220**				**27,789**[4]	**20,008**

Eighty-five PSUs were selected, with two large SSUs ("full" schools) selected from each PSU for a total of 170 large SSUs. The estimated sample yield from these large schools was 24,324 students

[4] Note that this was the anticipated number of students in all sampled schools, and the actual number of sampled students was derived only from participating schools (and is thus considerably lower).

before school and student non-response, leading to an expected total 17,513 participating students in large schools after accounting for non-response.

To provide adequate coverage of students in small schools (those with an enrollment of less than 28 students in any grade) 20 medium SSUs from a subsample of 10 PSUs, and 30 small SSUs from a subsample of 15 PSUs were selected. The expected yield was 1,411 from medium schools and 1,084 students from small schools. In total, the number of participating students was 20,008.

Within each school, one class was selected from each grade to participate in the survey except in high minority schools, where two classes per grade were selected. Note that the set of high-minority schools defined for double class sampling is necessarily a subset of the large schools that can support such double class sampling. For the 2017 NYTS, double class selection in schools with higher enrollments of black students was implemented to enhance the black student yields. The cutoff used in the 2017 NYTS sampling was 3% percent black students or more in a school.

2.5 Forming Sampling Units

2.5.1 Forming primary sampling units (PSUs)

In defining PSUs, several issues were considered:

- Each PSU should be large enough to contain the requisite numbers of schools and students by grade, and small enough so as not to be selected with near certainty.
- Each PSU should be compact geographically so that field staff could go from school to school easily.
- PSUs should be consistent with school and school district definitions (i.e., should not cross or split districts).
- PSUs are defined to contain at least four middle and five high schools.

Generally, counties were equivalent to PSUs, with two exceptions:

- Low population counties were combined to provide sufficient numbers of schools and students.
- High population counties were divided into multiple PSUs so that the resulting PSUs would not be selected with certainty.

The PSU frame was screened for PSUs that no longer met the above criteria. The frame was adjusted by re-combining small counties/PSUs as necessary to ensure sufficient size while maintaining compactness. Near-certainty PSUs were split using an automated procedure built into the sampling program.

2.5.2 Forming secondary sampling units (SSUs)

Single schools represented their own SSU if they had students in each of grades 6 through 8 or in grades 9 through 12. Schools that did not have all eligible grades for the level were grouped together to form a SSU. Linked schools were treated as single schools during sampling.

2.6 STRATIFICATION

The PSUs were organized into 16 strata, based on urban/rural location and proportion minority enrollment.

- If the percentage of Hispanic students in the PSU exceeded the percentage of non-Hispanic black students, then the PSU was classified as Hispanic. Otherwise it was classified as black.
- If the PSU was within one of the 54 largest MSAs in the United States, it was classified as "urban," otherwise it was classified as non-urban, or "rural".
- Hispanic urban and Hispanic rural PSUs were classified into four density groupings depending upon the percentages of Hispanic students in the PSU.
- Non-Hispanic black urban and non-Hispanic black rural PSUs were also classified into four groupings depending upon the percentages of black students in the PSU.

The density grouping bounds were computed using an optimization algorithm[5] that is refreshed each cycle to reflect changes in the racial/ethnic distribution of the student population. The boundaries or cutoffs changed as the frequency distribution ("*f*") for the racial groupings changed from one survey cycle to the next. Table 2.3 presents the stratum boundaries used in the 2017 NYTS.

Table 2.3 Stratum Boundaries: Minority Percentage Cutoffs

Minority Concentration	Density Group	Bounds	
		Urban	Non-Urban (Rural)
Black	1	0%-26%	0%-20%
	2	>26%-40%	>20%-34%
	3	>40%-54%	>34%-54%
	4	>54%-100%	>54%-100%
Hispanic	1	0%-26%	0%-24%
	2	>26%-42%	>24%-48%
	3	>42%-58%	>48%-68%
	4	>58%-100%	>68%-100%

As described earlier, SSUs were stratified into three sizes for small, medium, and large schools.

2.7 SAMPLE ALLOCATION AND SELECTION

The 2017 NYTS was designed to select a sample of 85 PSUs. The PSUs initially were allocated to strata proportional to student enrollment. For this cycle, a nearly proportional PSU allocation was achieved, resulting in gains in sampling efficiency. Table 2.4 shows the actual allocation of the PSU sample to the 16 strata defined by minority density and urban status, alongside a

[5] The cumulative square root of "*f*" method developed by Dalenius and Hodges (1959) "Minimum variance stratification". J American Statistical Association, 54, 88-101.

proportional allocation. The initial proportional allocation was modified slightly to ensure that all strata contained at least two PSUs to facilitate accurate variance estimation.

Table 2.4 First-Stage Strata and Frame PSU Distribution

Predominant Minority	Urban/Non-Urban	Density Group Number	Stratum Code	Student Population	Number of Sample PSUs (Revised)
Non-Hispanic Black	Urban	1	BU1	2,720,181	7
		2	BU2	975,490	4
		3	BU3	908,299	4
		4	BU4	516,712	5
	Non-urban (R)	1	BR1	3,937,157	8
		2	BR2	1,503,403	4
		3	BR3	1,026,612	4
		4	BR4	313,063	5
Hispanic	Urban	1	HU1	3,530,556	8
		2	HU2	2,429,442	7
		3	HU3	1,865,988	5
		4	HU4	2,106,242	4
	Non-urban (R)	1	HR1	4,427,215	12
		2	HR2	1,284,402	3
		3	HR3	988,655	3
		4	HR4	523,491	2

The sample was selected with probabilities proportional to size (PPS) methods at the first and second stages. With PPS sampling, the selection probability for each PSU is proportional to the PSU's measure of size. Systematic sampling procedures were applied to the stratified frame to select a PPS sample of PSUs:

- Selected 85 PSUs with a systematic random sampling within each stratum. The method applied within each stratum was a sampling interval computed as the sum of the measures of size for the PSUs in the stratum, divided by the number of PSUs to be selected in the stratum.

- Subsampled PSUs for the small school (15 PSUs) and medium school (10 PSUs) sampling of two schools per level in each subsample PSU.

2.8 Sample Sizes Attained in the Survey

The 2017 NYTS attained the target sample sizes in the key analytic subgroups of interest. Tables 2.5a–d show the number of participating students in subgroups defined by gender, grade, and race/ethnicity. Table 2.5d, about race/ethnicity distribution, is presented in two different ways: 1) using the original variable allowing for multiple races and including missing data, and 2) using the

imputed variable developed for post-stratification which includes complete data. By either measure, the sample led to more than 4,500 Hispanic students. It also led to 3,757 black students using the imputed variable and nearly 3,000 black student using the original variable.

Table 2.5a Subgroup Sample Sizes: Number of Participating Students

What is your sex?				
Sex (Q2)	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Male	8881	50.19	8881	50.19
Female	8815	49.81	17696	100.00
Frequency Missing = 176				

Table 2.5b Subgroup Sample Sizes: Number of Participating Students

What grade are you in?				
Grade (Q3)	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
6th	2524	14.20	2524	14.20
7th	2565	14.43	5089	28.63
8th	2473	13.91	7562	42.55
9th	2583	14.53	10145	57.08
10th	2637	14.84	12782	71.92
11th	2575	14.49	15357	86.41
12th	2391	13.45	17748	99.86
Ungraded or other grade	25	0.14	17773	100.00
Frequency Missing = 99				

Table 2.5c Subgroup Sample Sizes: Number of Participating Students

RECODE: Race/Ethnicity[6]				
RACE_M	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NH-White	7532	44.09	7532	44.09
NH-Black	2983	17.46	10515	61.55
Hispanic	4614	27.01	15129	88.56
NH-Asian	728	4.26	15857	92.82
NH-AI/AN	230	1.35	16087	94.16
NH-NHOPI	96	0.56	16183	94.73
Multiple Races	901	5.27	17084	100.00
Frequency Missing = 788				

Table 2.5d Subgroup Sample Sizes: Number of Participating Students[7]

Race Categories for Post-stratification				
Imputed Race/Ethnicity	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Asian/Native Hawaiian/Other PI	998	5.58	998	5.58
Black or African American	3757	21.02	4755	26.61
Hispanic/Latino	4706	26.33	9461	52.94
American Indian/Alaska Native	356	1.99	9817	54.93
White	8055	45.07	17872	100.00

Note: The multiple race categories are Hispanic, non-Hispanic (NH) white, non-Hispanic black, non-Hispanic Asian, non-Hispanic American Indian or Alaskan Native (AIAN), and non-Hispanic Native Hawaiian or Pacific Islander (NHOPI).

2.9 Sample Validation

Following the sample draw, each district and school were called to verify the correct information for each entity.

District validation included confirmation of the following:

- District name
- Name and title of 2016-2017 district superintendent
- District street address used for overnight deliveries, with city name and ZIP code

School validation included confirmation of the following:

- School is operational

[6] This variable is based on the original race/ethnicity question in the survey. It allows for multiple races and has missing values.

[7] There were 788 missing values for the Race/ethnicity variable in Table 2.5c. The variable in Table 2.5d imputed those values and also assigned a unique race/ethnic classification to multiple race respondents.

- School name and relationship to identified district (if applicable)
- Name and title of 2016-2017 school principal
- School street address used for overnight deliveries, with city name and ZIP code
- Grade levels served during 2016-2017 school year
- Approximate school enrollment
- At least a cumulative total enrollment of 40 students in the grades for which the school was selected
- School is a traditional "brick and mortar" school with traditional school-aged students who are not adults and who attend classes in person throughout the academic year
- School has its own unique student body
- School does not exclusively serve a specialized student population such as English Language Learners or Special Education students

Chapter 3—NYTS Data Collection and Processing

3.1 Survey Instrument

The NYTS collects data on key short-term, intermediate, and long-term tobacco prevention and control outcome indicators. The 2017 survey instrument included 88 questions. The first five questions collected student demographic information and the rest measured a comprehensive set of tobacco-related topics (Appendix A). Specific areas covered by the survey included: prevalence of tobacco product use; knowledge of and attitudes toward tobacco use; pro- and anti-tobacco media and advertising; minors' access to tobacco products; nicotine dependence; cessation attempts; exposure to second-hand smoke; harm perceptions; exposure to tobacco product warnings; and tobacco use prevention school curricula.

Historically, experts within CDC's Office on Smoking and Health (OSH), Epidemiology Branch have taken the lead on the NYTS questionnaire design. Working in concert with a variety of local, state, and federal stakeholders, including representatives from FDA, CDC reviews the questionnaire prior to each cycle to identify and remove redundancies, examine the most relevant indicators, and obtain guidance and suggestions for new items on the questionnaire.

3.2 External Review and Approvals

Two bodies reviewed and approved the instrumentation, processes, privacy and security elements, and sampling design of the 2017 NYTS: ICF's Institutional Review Board (IRB) and CDC's IRB. Office of Management and Budget review and approval was not required for the 2017 cycle of NYTS because the existing approval was not set to expire until January 31, 2018.

3.3 Recruitment Procedures

The schools selected to participate in the 2017 NYTS were located in 32 different states. Recruitment began in June 2016 with calls to state departments of education and health. Letters of support were obtained from various state agencies and used in mailings to districts and schools. District- and school-level recruitment began in September 2016. Before public or diocesan schools were contacted, verbal or written agreement was first obtained by their district or diocese; private schools were approached directly. A date for survey implementation was selected to optimize the efficiency of data collection while accommodating school schedules. In selecting a date, convenience to the school and its calendar were considered. Additionally, an effort was made to schedule groups of schools from the same school district or PSU around the same time to facilitate efficient travel to and survey implementation within selected schools. Recruiters used a secure web-based calendar to facilitate communication and to avoid scheduling two schools for the same data collector on the same day.

3.4 Survey Administration

Survey administration in the schools began on February 13, 2017, immediately after data collector training, and continued until June 14, 2017. Each data collector visited an average of three schools per week. While the details of each data collection varied, there were six core steps followed for every school: 1) pre-contact call with the principal or lead contact prior to arrival at the school; 2) entry meeting with the principal or lead contact; 3) entry meeting with teacher or group of teachers

prior to survey administration; 4) survey administration; 5) post-survey meeting with the teacher or teachers; and 6) post-survey meeting with the principal or lead contact prior to leaving the school. Most survey administrations could be completed in one day, while at other times, due to the number of classes selected or alternating block schedules, the data collector needed to return for a second day. Procedures were designed to protect students' privacy by assuring that student participation was anonymous and voluntary. Students completed a self-administered scannable questionnaire booklet via pencil and paper.

3.4.1 Data Collection Staffing

In order to minimize the amount of data collector travel between home and school assignments, hiring was done geographically across the country, with greater numbers of data collectors in those areas with higher concentrations of sampled schools. Data collectors were recruited from a pool of previously trained data collectors. Data collector training was conducted on February 8-10, 2017.

Key components of the training included the following:

- Pre-contact activities with the schools
- Entry and exit meetings with school officials
- Data collection protocols
- Follow-up activities
- Communication with headquarters staff

3.4.2 Field Procedures

After schools had been recruited, classes selected, and a date scheduled, each school received a packet of pre-survey materials. These materials included all the information necessary to prepare the school for data collection. Teacher packets contained the parental permission forms to be distributed to all students in the selected classes prior to data collection. The timing of these pre-survey packet mailings was determined in part by the type of permission form being used by the school. Passive parental permission forms, or forms returned only if the parents do not want their child to participate, were sent approximately one week prior to the scheduled date of data collection in the majority of schools. Active parental permission forms (i.e., forms that must be returned with the parent's signature in order for the child to participate) were sent out at least two weeks prior to the scheduled date of data collection for schools that require active consent. Follow-up calls were made to the selected schools to answer any questions and to make sure materials were received and distributed to selected classes and students.

On a weekly basis, data collectors received mailings containing their assignments for the coming week, travel, and logistics information. In addition to these mailings, boxes of survey supplies were sent to data collectors, either to the data collector's home or hotel. These boxes contained all supplies necessary for completing the data collection, including questionnaires, survey envelopes, field forms, and pencils. Data collectors were supplied with extra materials for emergency packs as well, which they carried with them at all times.

3.4.3 Classroom Selection

Students were selected for participation by default via the selection of whole classes (i.e., all students enrolled in a selected class were eligible to take the survey). The frames from which classes were chosen were constructed so that eligible students had one, and only one, chance of being selected. However, at times the specific method of selecting classes varied from school to school, according to how a school's class schedule was structured. Typically, classes were selected from a list of required core courses such as English, social studies, math, or science. Among middle school students, and among high school students in a few states, physical education and/or health also were considered core courses. However, in a small number of schools, it was difficult to develop an appropriate frame using this approach. Therefore, in these schools, classes were selected by using a time of day (e.g., second period) when all eligible students were scheduled to be attending a class of one kind or another as the frame, and randomly selecting from all classes held at this time. Lastly, in some schools, homerooms were used as the frame for class selection.

3.5 Web-Based Case Management System

For multiple cycles of the NYTS, a web-based case management system has been utilized to help centralize the management of the study, facilitate information exchange with field staff, and allow all members of the project management, recruitment, and supervisory teams and field staff access to information necessary to implement the study. The system is designed with differing levels of access depending on the user's role on the study. The system's main functions include generating invitation letters, tracking recruitment progress, scheduling data collection, and tracking school and student response rates.

3.6 Data Receipt, Processing, and Scanning

Data were shipped directly from the field to the contractor's headquarters for processing and scanning. Shipments were immediately logged in as received and checked for completeness (i.e., data from all expected schools and classes were accounted for). The status of each school's received data was logged into the case management system described in Section 3.5. Survey booklets were counted and discrepancies with the reported number of completed booklets reconciled. Booklets were also individually reviewed, page by page, to make sure bubbles were sufficiently filled in; stray pencil marks, comments, or illustrations that could interfere with scanning were erased; defaced booklets or those completed in something other than a No. 2 pencil were transcribed; and that at least one question had been answered. This process was repeated for each class within a school. As make-up surveys were received back from teachers, the process was again repeated and all accompanying documentation for that school was updated. Data were subsequently scanned using an optical scanner.

3.7 Participation Rates

Participation rates for the NYTS were calculated at the school and student levels. The ultimate goal for weightable data was to have the product of the two participation rates equal to or greater than 60%.[8]

3.7.1 School-level Participation Rates

At the school level, 242 schools were selected across 166 districts in 32 states. During sample validation, 15 schools were deemed to be ineligible and were replaced; one school was deemed ineligible but was not replaced. The remaining 241 schools were considered eligible for participation in the study.

In total, 185 schools (76.8%) participated in the study. The remaining 56 schools were considered refusals. Of these refusals, 21 of them were due to their district refusing to grant access to their schools to discuss participation and 35 were due to refusals at the school level. The most common reasons given for a refusal at the district or school level were loss of instructional time and standardized testing.

3.7.2 Student-level Participation Rates

Initial student-level participation rates were calculated from the field as data collectors completed survey administration each day. However, as data were reviewed, further refinements were made to: 1) revise the number of eligible students based on available documentation, 2) correct mathematical errors, 3) review counts of completed surveys, and 4) incorporate make-ups as they were received from schools from students and classes that did not participate on the initial day of survey administration.

After visual editing, incorporating make-ups, and receiving missing classes, the final student participation rate for the 2017 NYTS was 88.72%. Overall, 20,144 eligible students were invited to participate in the survey, and 17,872 did so. Table 3.1 below shows the number of eligible students, participants, and participation rates for the NYTS.

Table 3.1 Overall NYTS 2017 Student Participation Rate

	# Eligible	# Completed	Participation %
NYTS Participating Students	20,144	17,872	88.7

The 2017 NYTS survey attained an actual school participation rate of 76.8% and a student participation rate of 88.7%. The overall participation rate, the product of the school-level and student-level participation rates, was 68.1%, thus considered sufficient for weighting purposes.

3.8 Data Management

Scanned data were converted from separate school-specific ASCII files to a single national SAS dataset. Because the NYTS is administered via a paper-and-pencil booklet, there is no restriction

[8] Note that the recruitment goal for the combined school x student participation rate is 70%. This is in excess of the combined participation rate needed for confidence in weighting.

for the participant (e.g., providing two or more responses to a single response question) that prevents inconsistencies in the dataset. Therefore, CDC created a series of data-cleaning specifications that were applied to eliminate internal inconsistencies. These cleaning specifications also computed certain analytic variables and re-coded race/ethnicity values to match CDC-required classifications.

The survey data file preparation for weighting involved a series of data file linking steps. These steps ensured that the data files merged the school information compiled during frame construction, sample selection, replacement of ineligible schools, recruitment, and data collection using a common school identifier (QID).

Chapter 4—Weighting of NYTS Response Data

This section describes the procedures used to weight the NYTS data including:

- Sampling weights
- Nonresponse adjustments
- Post-stratification to national estimates by grade and weight trimming

This section focuses on the development of the weights for the student response data. The final student-level response data were weighted to reflect the initial probabilities of selection and nonresponse patterns, to mitigate large variations in sampling weights, and to post-stratify the data to known sampling frame characteristics. The section also describes the computation of weighted estimates and variance estimates.

Although the sample was designed to be approximately self-weighting, survey weights were necessary to produce unbiased estimates. The basic weights, or sampling weights, were computed on a case-by-case basis as the reciprocal of the probability of selection of that case. Below is a simple presentation of the basic steps in weight computation.

4.1 Sampling Weights

The base weight is the inverse of the probability of selection for each responding student. The base weight was adjusted to compensate for nonresponse, to alleviate excess weight variation, and to match the weighted data to known control totals. The base weight was computed by inverting the probabilities of selection at each stage to derive a stage weight. For each respondent, the stage weights were multiplied to form the overall sampling weight assigned to each student.

The NYTS computation of sampling weights began at the student sampling stage, and then moved to the school and PSU sampling stages. This sequence allowed the student sampling weights to incorporate adjustments for student nonresponse. These adjustments, described next, used enrollment data by sex and by grade collected for each participating school. Because the process began with the student weights within a given grade, school, and PSU, these weights are referred to as conditional.

4.1.1 Adjusted Conditional Student Weights

The adjusted conditional student weight is the student weight given the selection of the PSU, school, and grade. This weight is the product of the inverse of the probability of selection and a nonresponse adjustment within weighting classes based on grade and sex. Note that this step also includes an approach designed to limit the nonresponse adjustment factor, an early step to avoid extreme weights and hence to control the variability in the weights.

This three-step process is simplified algebraically and computed directly as the ratio of the number of enrolled students to the number of responding students in a given weighting class within a school. The weighting class definition is set dynamically so as to avoid extreme weights, as described next.

The student selection weight is denoted as W^{R}_{cklm}, where the subscripts *k*, *l*, and *m* refer to the school, PSU and stratum as before. The subscript *c* refers to the weighting class, described below. This weight was computed as below, where *N* is the number of enrolled students for each school (the counts are provided by the school during data collection by grade and sex) and *R* is the number of responding students in weighting class *c* within a given school:

$$W^{R}_{cklm} = \frac{N_{cklm}}{R_{cklm}}$$

The weighting class *c* was defined by a sequence of rules that depended on the number of responding students. This was to avoid large weights for classes with low numbers of respondents. This process operated entirely within schools.

Initially, the weighting class was defined by grade and sex within each school. If the weight for the class exceeds a maximum value, *C*, then weighting classes are combined. This cap *C* was computed using the following equation:

$$C_{cklm} = 2\frac{N_{cklm}}{\min(10, N_{cklm})}$$

The combination sequence first grouped males and females within a grade. Both the cap and the weight were then recomputed. If the weight still exceeded the cap, grades were combined. The process was repeated, and if the student weight still exceeded the cap, the school was taken as the weighting class.

This had the effect, within a school, of setting an upper limit on the weight of 2 in weighting classes with an enrollment of less than 10, and 20% of the enrollment in weighting classes with an enrollment of more than 10. Note that the cap could be exceeded, however, in the rare cases where the weighting class was collapsed to the school level.

4.1.2 School Sampling Weights

For large schools, the partial school weight was the inverse of the probability of selection of the school given that the PSU was selected:

$$W^{LS}_{klm} = \left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{LS}_{klm}}$$

For small schools, the partial school weight was:

$$W^{SS}_{klm} = (85/15)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{SS}_{klm}}$$

For medium schools, the partial school weight for both high schools and middle schools was:

$$W^{MS}{}_{klm} = (85/10)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{MS}{}_{klm}}$$

The overall weights for a given PSU, school and grade combination were the product of the adjusted PSU, school and grade-level weights.

4.1.3 Grade Sampling Weights

Grade selection occurred within linked schools where the grade was available in each of the linked schools, or school "components" that constitute the SSU. The partial weight for a grade, given the selection of the linked school containing it, was simply the inverse of the probability of selection described in Section 2.4. In a non-linked school, the weight was 1.0. The grade weight is denoted as W^{G}_{jklm}.

4.1.4 PSU Sampling Weights

The weight of the PSU was the inverse of the probability of selection of that PSU:

$$W^{P}{}_{lm} = \frac{1}{K_m}\left(\frac{MOS_{.m}}{MOS_{lm}}\right) = \frac{1}{P^{P}{}_{lm}}$$

For small and medium school selections, the supporting sample PSUs were drawn as a subsample. This PSU subsampling component of the PSU weight was accounted for in the school selection probability and corresponding weight.

4.1.5 Overall Sampling Weight

The overall sampling weight was formed as the product of the stage selection weights. This weight, W^{T1}, was then adjusted for nonresponse, trimmed, and poststratified to control totals, as described in the following sections. This weight was computed as:

$$\begin{cases} W^{T1}{}_{hijklm} = W^{P}{}_{lm} W^{LS}{}_{klm} W^{G}{}_{jklm} W^{R}{}_{hijklm} \\ W^{T1}{}_{hijklm} = W^{P}{}_{lm} W^{MS}{}_{klm} W^{G}{}_{jklm} W^{R}{}_{hijklm} \\ W^{T1}{}_{hijklm} = W^{P}{}_{lm} W^{SS}{}_{klm} W^{G}{}_{jklm} W^{R}{}_{hijklm} \end{cases}$$

for large, medium, and small schools, respectively, where the weights in the latter portions of the equations are defined in the preceding sections.

4.2 Nonresponse Adjustments

Nonresponse adjustment of weights is important to reduce potential bias incorporated into surveys from differences between responding and nonresponding students and schools included in the sample.

4.2.1 Student Nonresponse Adjustment

An adjustment for student nonresponse was made by sex and grade within school. With this adjustment, the sum of the student weights over participating students within a school matched the total enrollment by grade and sex in the school collected during data collection. This adjustment factor was capped in extreme situations to limit the potential effects of extreme weights on the precision of survey estimates.

In the 2017 NYTS cycle, the school nonresponse adjustment methods were refined to further minimize nonresponse bias potential. As opposed to the previous method, which created adjustment cells based on sampling strata, the new method defined nonresponse adjustment cells in a more tailored and systematic approach stemming from the non-response analysis. These analyses are detailed in the *2017 NYTS Nonresponse Bias Analysis* report.

Specifically, the definition of the most appropriate nonresponse adjustment weighting cells followed these steps:

1. Conduct bivariate analysis to identify key predictors of school nonresponse and student nonresponse.
2. Conduct multivariate logistic regression analysis, or response propensity models, including the subset of key predictors identified in No. 1 to identify significant predictors of non-response at both levels.
3. Develop nonresponse adjustment weighting cells based on the significant predictors while incorporating information about cell sizes and correlations between predictors.

During the 2017 cycle, only the school type (public versus non-public) was found to be predictive on nonresponse. With only one predictive variable, multivariate analysis was not needed. Nonresponse adjustment cells were created using school type and region. Because of the small number of sample non-public schools, they were included as their own category in the nonresponse adjustment cells.

Typically, with multiple variables associated with school nonresponse, the subset of variables selected for defining weight adjustment cells is effectively reduced in two ways: 1) by eliminating variables with high pairwise correlations, and 2) limiting to variables and cells with adequate representation of participating schools. These steps were not needed in 2017 because only school type was significant.

Several weight adjustments were used to account for student and school nonresponse patterns. An adjustment for student nonresponse was made by sex and grade within school. With this adjustment, the sum of the student weights over participating students within a school matches the total enrollment by grade and sex in the school collected during data collection. This adjustment factor was capped in extreme situations to limit the potential effects of extreme weights on the precision of survey estimates.

The weights of students in participating schools were adjusted to account for nonparticipation by other schools. The adjustment factor (A_m) is the ratio of the sum of weighted MOS of all selected schools in the stratum over the sum of the weighted MOS for participating schools in a stratum. The adjustment factor was computed and applied to public and non-public schools separately.

The adjustment process used the following equations for the adjustment factor:

$$A_m = \frac{\sum_{k,l \in sampled\ schools}(W_{lm}^P * W_{klm} * MOS_{klm})}{\sum_{k,l \in participating\ schools}(W_{lm}^P * W_{klm} * MOS_{klm})}$$

The student weight adjusted for nonresponse was then:

$$W_3^S = W_2^S * A_m$$

Table 4.1 presents the nonresponse adjustment factors within each of the nonresponse adjustment cells defined by non-public school and public school by region.

Table 4.1 Nonresponse Adjustment Factors in Each Adjustment Cell

Weighting Class	Weight Sum Over Participants	Responding School Count	Weight Sum over all Sample	Sample School Count	Response Rate	Adjustment Factor
Non-public Schools	1,218,137.13	14	2,850,776.31	25	56.000	2.340
Public Schools, Region 1	4,281,074.24	31	4,597,158.84	35	88.571	1.073
Public Schools, Region 2	6,785,317.73	35	8,014,359.24	44	79.545	1.181
Public Schools, Region 3	8,629,885.92	71	11,288,173.40	92	77.173	1.308
Public Schools, Region 4	5,811,827.49	34	7,821,973.30	46	73.913	1.345
	26,726,242.50	**185**	**34,572,441.08**	**242**		

4.3 Poststratification and Trimming

The final steps in the weighting process include trimming and poststratification. Extreme variation in sampling weights can inflate sampling variances and offset the precision gained from a well-designed sampling plan. Nonresponse adjustments while minimizing bias can add additional variances. One strategy to compensate for these potential effects is to trim extreme weights and distribute the trimmed weight among the untrimmed weights. The trimming is an iterative procedure. It is possible to implement the iterative trimming in conjunction with the iterative poststratification, or raking, procedures described next.

Poststratification approaches capitalize on known population totals and percentages available for groups of schools and students. National estimates of racial/ethnic counts for poststratification were obtained from two sources described next. Private schools enrollments by grade and five racial/ethnic groups were obtained from the Private School Survey (PSS); public school enrollments by grade, sex, and five racial/ethnic categories were obtained from the Common Core

of Data (CCD). Both are produced by the National Center of Education Statistics (NCES); the most recent versions, the 2013–14 CCD and the 2013–14 PSS was used.

These databases were combined to produce the enrollments for all schools and to develop population counts to use as controls in the poststratification step. Iterative poststratification, or raking, methods allowed the use of additional poststratification variables and categories. The iterative approach allowed the simultaneous application of a trimming procedure (see, for example Iachan, 2010).[9] Trimming is designed to limit the variance increase that may follow from the bias-reduction raking methods. The trimming method capped the weights at the median plus four times the interquartile range of the weight distribution.

Tables 4.2 and 4.3 present the population control totals, which are also the sums of the weights in each poststratum cell separately for public and non-public schools by grade and sex and by grade and race/ethnicity, respectively, to reflect the iterations used in the raking procedures.

[9] Iachan, R. (2010, August). *A new iterative method for weight trimming and raking.* Paper presented at the American Statistical Association meeting, Vancouver, Canada.

Table 4.2 Sum of Final Weights vs. Control Total - by Public Flag, Grade and Sex

School Type	Grade	Gender	Number of Records	Weight Sum = Control Total
Public	6	Male	1233	1887889.41
Public	6	Female	1197	1796710.59
Public	7	Male	1195	1918191.20
Public	7	Female	1264	1826578.80
Public	8	Male	1200	1912595.69
Public	8	Female	1183	1828252.31
Public	9	Male	1180	2008314.37
Public	9	Female	1271	1884945.63
Public	10	Male	1197	1861769.58
Public	10	Female	1259	1789898.42
Public	11	Male	1239	1711214.99
Public	11	Female	1152	1678713.01
Public	12	Male	1116	1643345.33
Public	12	Female	1108	1628831.67
Private	6	Combined	106	260122.00
Private	7	Combined	125	259795.00
Private	8	Combined	112	259872.00
Private	9	Combined	160	250101.00
Private	10	Combined	191	247254.00
Private	11	Combined	205	243380.00
Private	12	Combined	179	237793.00

Table 4.3 Sum of Final Weights vs. Control Total - by Public Flag, Grade and Race

School Type	Grade	Race/Hispanic Origin	Number of Records	Weight Sum = Control Total
Public	6	Non-Hispanic Native American	92	39520.01
Public	6	Non-Hispanic Asian and Pacific Islander	107	184901.31
Public	6	Non-Hispanic Black	513	583490.27
Public	6	Hispanic	689	964836.55
Public	6	Non-Hispanic White	1029	1911851.86
Public	7	Non-Hispanic Native American	69	40144.14
Public	7	Non-Hispanic Asian and Pacific Islander	108	184802.31
Public	7	Non-Hispanic Black	485	595382.61
Public	7	Hispanic	660	968059.68
Public	7	Non-Hispanic White	1137	1956381.26
Public	8	Non-Hispanic Native American	44	40377.79
Public	8	Non-Hispanic Asian and Pacific Islander	108	187423.33
Public	8	Non-Hispanic Black	513	596541.72
Public	8	Hispanic	628	943768.85
Public	8	Non-Hispanic White	1090	1972736.31
Public	9	Non-Hispanic Native American	40	43307.51
Public	9	Non-Hispanic Asian and Pacific Islander	136	181737.57
Public	9	Non-Hispanic Black	561	653529.40
Public	9	Hispanic	641	981991.29
Public	9	Non-Hispanic White	1073	2032694.23
Public	10	Non-Hispanic Native American	39	39499.34
Public	10	Non-Hispanic Asian and Pacific Islander	138	182627.19
Public	10	Non-Hispanic Black	471	581745.81
Public	10	Hispanic	693	884308.87
Public	10	Non-Hispanic White	1115	1963486.79
Public	11	Non-Hispanic Native American	33	34995.73
Public	11	Non-Hispanic Asian and Pacific Islander	156	182576.00
Public	11	Non-Hispanic Black	538	514866.40
Public	11	Hispanic	672	784371.22
Public	11	Non-Hispanic White	992	1873118.65
Public	12	Non-Hispanic Native American	27	33758.82
Public	12	Non-Hispanic Asian and Pacific Islander	133	180238.60
Public	12	Non-Hispanic Black	499	483886.39
Public	12	Hispanic	605	728269.73
Public	12	Non-Hispanic White	960	1846023.46
Private	6	Combined	106	260122.00
Private	7	Combined	125	259795.00
Private	8	Combined	112	259872.00
Private	9	Combined	160	250101.00
Private	10	Combined	191	247254.00
Private	11	Combined	205	243380.00
Private	12	Combined	179	237793.00

For poststratification purposes, a unique race/ethnicity was assigned to respondents with missing data on race/ethnicity, those with an "Other" classification, and those reporting multiple races. For non-public schools, we did not poststratify by race/ethnic classifications – white and non-white. For public schools we used the full five categories.

The raking and trimming method ensured that final weights sum to the population control totals in each cell while also limiting the coefficient of variation (CV) of the weights. The CV=86.7% implies that the design-effect (DEFF) component due to unequal weighing effects is 1.75.[10]

4.4 Estimators and Variance Estimation

Weighted estimates of means, percentages and totals can be computed using the final weights included in the analysis file. If w_i is the weight of case i (the inverse of the probability of selection adjusted for nonresponse and poststratification adjustments) and x_i is a characteristic of case i (e.g., x_i=1 if student i smokes, but is zero otherwise), then the mean of characteristic x is estimated as $(\Sigma\, w_i x_i)/(\Sigma\, w_i)$. A weighted population total estimate is computed similarly as $(\Sigma\, w_i x_i)$. The weighted population estimates can be computed with the Statistical Analysis System (SAS) as well as with other statistical software.

These estimates are accompanied by measures of sampling variability, or sampling error, such as variances and standard errors, that account for the complex sampling design. These measures support the construction of confidence intervals and other statistical inference such as statistical testing (e.g., subgroup comparisons or trends over successive NYTS cycles). Sampling variances can be estimated using the method of general linearized estimators[11] as implemented in SAS survey procedures. These software packages must be used because they permit estimation of sampling variances for multistage stratified sampling designs. They also account for unequal weighting and for sample clustering and stratification.

The final weight files also include PSU and strata variables which support the analysis of clustered survey data and accurate variance estimation. As in previous cycles, a variable for "variance strata," was added which may differ from the design strata, to ensure that all variance strata had at least two PSUs.[12]

Tables 4.4–4.7 present weighted estimates and estimated standard errors for key outcome measures using the 2017 NYTS data. Sample SAS code is provided in Exhibit 4.1.

[10] The design effect due to unequal weighting may be expressed in terms of the cv of the weight as DEFF= 1 + cv**2.

[11] Skinner CJ, Holt D, and Smith TMF, Analysis of Complex Surveys, John Wiley & Sons, New York, 1989, 50.

[12] Specifically, two strata (coded 202 and 203) were combined into one variance stratum (203) because the original stratum "202" had only one PSU when analyzed at the middle school level.

Exhibit 4.1: Example SAS and SUDAAN Code for Generating Weighted Estimates and Standard Errors

SAS:

```
Proc Surveymeans Data=nyts2017 mean;
Var ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Class ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Stratum v_stratum2;
Cluster psu2;
Weight wt;
Domain HSMS HSMS*Sex HSMS*Race_S;
Title “NYTS 2017, Estimates by School Type, by School Type and Sex Cross-Classified, and by School Type and
Race/Ethnicity Cross-Classified”;
run;
```

SUDAAN:

```
Proc Descript Data=nyts2017 Filetype= SAS Design=WR;
Var ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Catlevel 1 1 1 1 1;
Nest v_stratum2 PSU2 / Missunit;
Weight wt;
Subgroup HSMS Sex Race_S;
Levels 2 2 3;
Tables HSMS HSMS*Sex HSMS*Race_S;
Title “NYTS 2017, Estimates by School Type, by School Type and Sex Cross-Classified, and by School Type and
Race Cross-Classified”;
Print Percent Sepercent / Style=NCHS;
run;
```

Table 4.4 Current Use Estimates[13] for Selected Tobacco Products for High School Students

Product	Overall %(SE)	Female %(SE)	Male %(SE)	White %(SE)	Black %(SE)	Hispanic %(SE)
CBIDIS	0.73% (0.11%)	0.61% (0.10%)	0.68% (0.17%)	0.60% (0.16%)	0.73% (0.23%)	1.10% (0.24%)
CCIGAR	7.74% (0.60%)	6.29% (0.69%)	8.98% (0.77%)	8.32% (0.79%)	8.46% (0.99%)	6.92% (0.88%)
CCIGT	7.66% (0.59%)	7.57% (0.73%)	7.57% (0.67%)	9.53% (0.80%)	3.82% (0.63%)	6.34% (0.96%)
CDISSOLV	0.68% (0.11%)	0.52% (0.13%)	0.82% (0.15%)	0.58% (0.15%)	0.47% (0.19%)	1.08% (0.21%)
CELCIGT	11.67% (1.02%)	9.93% (0.97%)	13.30% (1.22%)	14.35% (1.06%)	5.83% (0.86%)	10.27% (1.82%)
CHOOKAH	3.43% (0.36%)	3.28% (0.37%)	3.41% (0.47%)	2.76% (0.42%)	3.53% (0.57%)	4.93% (0.75%)
CPIPE	0.87% (0.10%)	0.54% (0.11%)	1.05% (0.15%)	0.73% (0.14%)	0.31% (0.10%)	1.35% (0.31%)
CROLLCIGTS	2.72% (0.32%)	2.22% (0.31%)	3.02% (0.43%)	2.93% (0.53%)	1.76% (0.32%)	2.68% (0.35%)
CSLT	4.10% (0.65%)	1.31% (0.30%)	6.72% (1.05%)	5.71% (0.94%)	1.47% (0.42%)	2.54% (0.61%)
CSNUS	2.01% (0.28%)	1.92% (0.30%)	2.09% (0.33%)	2.69% (0.42%)	0.81% (0.23%)	1.49% (0.32%)

[13] The estimates in tables 4.4 – 4.7 use the variable HSMS not provided on the public use file. Similar subgroup estimates can be produced using the variable Q3 which is included in that data file

Table 4.5 Current Use Estimates[14] for Selected Tobacco Products for Middle School Students

Product	Overall %(SE)	Female %(SE)	Male %(SE)	White %(SE)	Black %(SE)	Hispanic %(SE)
CBIDIS	0.31% (0.08%)	0.30% (0.10%)	0.31% (0.11%)	0.19% (0.08%)	0.36% (0.18%)	0.66% (0.24%)
CCIGAR	1.72% (0.25%)	1.46% (0.27%)	1.93% (0.38%)	1.51% (0.37%)	1.66% (0.37%)	2.50% (0.43%)
CCIGT	2.32% (0.23%)	2.25% (0.29%)	2.31% (0.36%)	2.00% (0.31%)	1.99% (0.46%)	3.62% (0.46%)
CDISSOLV	0.32% (0.08%)	0.19% (0.06%)	0.41% (0.14%)	0.10% (0.08%)	0.23% (0.15%)	0.84% (0.32%)
CELCIGT	3.61% (0.28%)	2.98% (0.33%)	4.10% (0.40%)	3.78% (0.53%)	2.65% (0.48%)	4.29% (0.57%)
CHOOKAH	1.36% (0.20%)	1.10% (0.20%)	1.60% (0.33%)	0.63% (0.19%)	1.51% (0.40%)	2.86% (0.49%)
CPIPE	0.45% (0.11%)	0.24% (0.08%)	0.63% (0.19%)	0.31% (0.12%)	1.04% (0.38%)	0.51% (0.19%)
CROLLCIGTS	1.36% (0.18%)	0.84% (0.16%)	1.75% (0.26%)	0.57% (0.15%)	1.45% (0.40%)	2.87% (0.46%)
CSLT	1.62% (0.24%)	0.81% (0.16%)	2.34% (0.42%)	1.61% (0.39%)	0.83% (0.26%)	2.43% (0.43%)
CSNUS	0.68% (0.12%)	0.66% (0.15%)	0.64% (0.17%)	0.57% (0.18%)	0.33% (0.16%)	1.25% (0.25%)

[14] The estimates in tables 4.4–4.7 use the variable HSMS not provided on the public use file. Similar subgroup estimates can be produced using the variable Q3 which is available in that data file

Table 4.6 Ever Use Estimates[15] for Selected Tobacco Products for High School Students

Product	Overall %(SE)	Female %(SE)	Male %(SE)	White %(SE)	Black %(SE)	Hispanic %(SE)
EBIDIS	1.63% (0.18%)	1.37% (0.19%)	1.78% (0.25%)	1.56% (0.26%)	1.45% (0.31%)	1.98% (0.33%)
ECIGAR	20.03% (1.26%)	16.74% (1.15%)	23.04% (1.69%)	21.86% (1.35%)	21.18% (1.38%)	17.79% (2.11%)
ECIGT	23.38% (1.37%)	22.80% (1.44%)	23.56% (1.51%)	25.49% (1.55%)	17.79% (1.46%)	23.76% (2.54%)
EDISSOLV	1.53% (0.20%)	1.19% (0.20%)	1.79% (0.25%)	1.63% (0.30%)	0.37% (0.14%)	2.12% (0.37%)
EELCIGT	29.91% (1.55%)	28.28% (1.62%)	31.42% (1.70%)	33.84% (1.64%)	18.86% (1.61%)	30.37% (3.00%)
EHOOKAH	11.23% (0.76%)	11.21% (0.91%)	11.14% (0.86%)	10.25% (0.90%)	11.65% (1.17%)	14.28% (1.39%)
EPIPE	2.83% (0.25%)	2.08% (0.30%)	3.43% (0.38%)	3.27% (0.34%)	0.95% (0.30%)	2.98% (0.48%)
EROLLCIGTS	5.95% (0.48%)	5.66% (0.57%)	6.01% (0.55%)	6.73% (0.66%)	3.64% (0.61%)	5.47% (0.80%)
ESLT	10.00% (1.32%)	4.43% (0.82%)	15.29% (1.88%)	13.89% (1.92%)	2.88% (0.57%)	6.16% (1.18%)
ESNUS	5.61% (0.75%)	4.73% (0.51%)	6.40% (1.20%)	7.09% (1.13%)	1.96% (0.38%)	4.72% (0.73%)

[15] The estimates in tables 4.4–4.7 use the variable HSMS not provided on the public use file. Similar subgroup estimates can be produced using the variable Q3 which is available in that data file.

Table 4.7 Ever Use Estimates[16] for Selected Tobacco Products for Middle School Students

Product	Overall %(SE)	Female %(SE)	Male %(SE)	White %(SE)	Black %(SE)	Hispanic %(SE)
EBIDIS	0.73% (0.12%)	0.64% (0.17%)	0.80% (0.17%)	0.53% (0.17%)	0.82% (0.29%)	1.27% (0.33%)
ECIGAR	5.36% (0.51%)	4.41% (0.44%)	6.07% (0.71%)	4.49% (0.81%)	7.34% (0.98%)	6.80% (0.77%)
ECIGT	8.60% (0.62%)	7.44% (0.62%)	9.66% (0.88%)	8.27% (0.88%)	9.32% (1.11%)	9.81% (0.79%)
EDISSOLV	0.95% (0.17%)	0.54% (0.16%)	1.27% (0.27%)	0.81% (0.21%)	0.79% (0.36%)	1.43% (0.35%)
EELCIGT	11.16% (0.64%)	9.78% (0.77%)	12.39% (0.78%)	11.04% (0.90%)	10.75% (1.21%)	13.38% (0.94%)
EHOOKAH	3.33% (0.29%)	2.64% (0.29%)	3.92% (0.46%)	2.03% (0.37%)	4.39% (0.57%)	5.94% (0.59%)
EPIPE	1.08% (0.19%)	1.10% (0.28%)	1.06% (0.27%)	1.04% (0.31%)	0.91% (0.30%)	1.37% (0.38%)
EROLLCIGTS	3.00% (0.30%)	2.34% (0.34%)	3.43% (0.42%)	2.24% (0.36%)	3.22% (0.69%)	4.57% (0.60%)
ESLT	3.29% (0.47%)	1.65% (0.26%)	4.76% (0.81%)	3.85% (0.75%)	1.23% (0.29%)	3.53% (0.54%)
ESNUS	1.92% (0.32%)	1.62% (0.27%)	2.20% (0.43%)	2.03% (0.48%)	1.12% (0.30%)	2.27% (0.36%)

[16] The estimates in tables 4.4–4.7 use the variable HSMS not provided on the public use file. Similar subgroup estimates can be produced using the variable Q3, which is available in that data file.

APPENDIX A. QUESTIONNAIRE

Questionnaire only included in PDF version of this document.

APPENDIX B. STUDENT WEIGHT DETAIL

Students were selected from schools via the selection of intact class sections as described in Section 2.3. The student sampling weight was computed based on a ratio of enrolling to responding students described in Section 4.1.1. The purpose of this section is to show that the resulting student weight is equivalent to computing a student weight as the inverse of the selection probability—as are the other stage sampling weights—followed by two adjustments, one for nonresponse and another poststratifying to known enrollment totals.

For the purposes of clarity, subscripts denoting the sampling stages and weight class are omitted. The unsubscripted quantities presented are assumed to be within weight class *c*, as defined in Section 4.1.1.

The probability of selection of a class when there are C_{jklm} classes at grade j in school k, PSU_i, stratum m is just $1/C_{jklm}$ or $2/C_{jklm,}$ depending on whether 1 or 2 classes are taken in the school. All students in a selected class were chosen so the probability of selection of a student is the same as the class, as well as constant across students within a student weighting class. The initial selection probability is taken to be the inverse of this sampling probability.

A simplified notation, letting K represent the number of sampled class sections, would look like:

$$W = \frac{C}{K}$$

Nonresponse Adjustment

The nonresponse adjustment inflates the weight of the responding students to equal that of the sampled students. The adjustment was calculated as the sum of the weights for sampled students to the sum of the weights for responding students,

$$F_{\text{NR}} = \frac{\sum_{\text{Selected}} \text{W}}{\sum_{\text{Responding}} \text{W}} = \frac{n}{R}$$

where *n* represents the number of sampled students and *R* represents the number of responding students in the student weight class. Note that the equation simplifies to a ratio that does not involve *W*, as *W* is constant within the class.

Enrollment Ratio Adjustment

Next, the nonresponse adjusted student weights are ratio-adjusted to conform to known school enrollment totals for each grade and sex. The adjustment F_{ps} is computed as

$$F_{\text{ps}} = \frac{N}{\sum W'} = \frac{N}{R * W'}$$

where N is the number of enrolled students in the weight class, and

$$W' = W * F_{NR}$$

The fully adjusted student weight is computed as:

$$W'' = W' * F_{PS}$$

The simplified equation is as follows:

$$\begin{aligned} W'' &= W' * F_{PS} \\ &= W' * \frac{N}{R * W'} \\ &= \frac{N}{R} \end{aligned}$$

APPENDIX C. COMMON CORE OF DATA RACE/ETHNICITY DEFINITIONS

Non-Hispanic American Indian/Alaska Native—A person having origins in any of the original peoples of North and South America (including Central America) and who maintains cultural identification through tribal affiliation or community recognition.

Non-Hispanic Asian/Pacific Islander—A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, Thailand, Vietnam, Guam, the Philippine Islands, Samoa, and other Pacific Islands.

Non-Hispanic Black—A person having origins in any of the black racial groups of Africa; African American.

Hispanic—A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.

Non-Hispanic White—A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.