METHODOLOGY REPORT OF THE 2021 NATIONAL YOUTH TOBACCO SURVEY

Recommended Citation

Office on Smoking and Health. 2021 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, 2021.

August 2021

For questions about this report, please email Sean Hu at fik4@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

August 2021

TABLE OF CONTENTS

CHAPTER 1—NYTS SAMPLING DESIGN 1

1.1 OVERVIEW OF THE NATIONAL YOUTH TOBACCO SURVEY (NYTS) 1

1.2 OVERVIEW OF THE 2021 NYTS METHODOLOGY 1

CHAPTER 2—NYTS SAMPLING METHODS 3

2.1 SAMPLE DESIGN 3

2.2 SAMPLING FRAME 4

2.3 SAMPLING UNITS AND MEASURE OF SIZE 5

2.4 PROJECTED SAMPLE SIZES 6

2.5 FORMING SAMPLING UNITS 7

2.6 STRATIFICATION 8

2.7 SAMPLE ALLOCATION AND SELECTION 9

2.8 SAMPLE SIZES ATTAINED IN THE SURVEY 10

2.9 SAMPLE VALIDATION 12

CHAPTER 3—NYTS DATA COLLECTION AND PROCESSING 14

3.1 SURVEY INSTRUMENT 14

3.2 EXTERNAL REVIEW AND APPROVALS 15

3.3 TECHNICAL ASSISTANCE PROVIDER (TAPS) STAFFING 15

3.4 RECRUITMENT PROCEDURES 15

3.5 SURVEY ADMINISTRATION 16

3.6 WEB-BASED DATA COLLECTION MANAGEMENT APPLICATION (DCMA) 17

3.7 DATA RECORDING 17

3.8 PARTICIPATION RATES 17

3.9 DATA MANAGEMENT 18

CHAPTER 4—WEIGHTING OF NYTS RESPONSE DATA 20

4.1 SAMPLING WEIGHTS 20

4.2 NONRESPONSE ADJUSTMENTS 23

4.3 POST-STRATIFICATION AND TRIMMING 25

4.4 ESTIMATORS AND VARIANCE ESTIMATION 29

APPENDICES

A. QUESTIONNAIRE

B. COMMON CORE OF DATA RACE AND ETHNICITY DEFINITIONS

C. STUDENT WEIGHT DETAIL

CHAPTER 1—NYTS SAMPLING DESIGN

1.1 OVERVIEW OF THE NATIONAL YOUTH TOBACCO SURVEY (NYTS)

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] Tobacco product-related indicators included in the NYTS are: tobacco product use (e.g., electronic cigarettes, cigarettes, cigars [including cigars, little cigars, and cigarillos], smokeless tobacco [chewing tobacco, snuff, or dip; snus, dissolvable tobacco products], hookahs, pipe tobacco, bidis, roll-your-own cigarettes, heated tobacco products, and nicotine pouches); exposure to secondhand smoke and e-cigarette aerosol; smoking cessation; minors' access to tobacco products; knowledge and attitudes about tobacco; and familiarity with pro-tobacco advertisements and anti-tobacco media messages. National estimates based on NYTS data also serve as essential benchmarks against which TCPs can compare the extent of youth tobacco product use in their own states and communities. The NYTS provides multiple measures and data for seven of the 18 tobacco-related Healthy People 2030 objectives (USDHHS, 2020): TU-4, TU-5, TU-6, TU-7, TU-8, TU-9, and TU-22.

First conducted during the fall of 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 2021 NYTS METHODOLOGY

The 2021 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.

After being conducted via paper and pencil questionnaires since its inception in 1999, the NYTS began using electronic data collection methods starting in 2019. The 2019 and 2020 cycles were conducted using a tablet-based administration with offline data collection in schools, supported by trained survey administrators. Using this methodology during the 2020/2021 school year was not feasible given state and local emergency COVID-19 protocols (e.g., distance and hybrid learning models, restricted visitor access). Instead, the 2021 NYTS was administered as a 100% online survey, supported virtually by trained technical assistance providers (TAPs). Students participated in the survey while at school or at home during a designated class period as part of a classroom

[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.

activity. Using a school-issued or personal internet-connected device, students logged into a secure website and watched a brief 2-minute instructional video before completing the survey. Students or whole classes unable to participate during the designated class period were asked to take the survey at the next possible opportunity. Due to the changes in the methodology for the 2021 administration, results of the 2021 NYTS cannot be compared with previous NYTS survey results that were primarily conducted on school campuses.

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.

Survey administration initiated on January 18, 2021 and concluded on May 21, 2021. The final sample consisted of 508 schools, of which 279 participated, yielding a school participation rate of 54.9%. A total of 20,413 student questionnaires were completed out of a sample of 25,149 students, yielding a student participation rate of 81.2%. The overall participation rate, defined as the product of the school-level and student-level participation rates, was 44.6%.

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 2021 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 product-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 and ethnicity (non-Hispanic White, non-Hispanic Black, and Hispanic; Appendix B). Additional estimates also were supported for subgroups defined by grade, by sex, and by race and ethnicity, each within school level domain; however, precision levels varied 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 racial/ethnic status and urban versus rural. PSUs were classified as "urban" if they were in one of the 54 largest Metropolitan Statistical Areas (MSAs) in the United States; otherwise, they were classified as "non-urban." 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	100 Counties, portions of a county, or groups of counties
2	Schools	Small, medium and large; High school vs. middle school	Aggregate eligible enrollment	450 SSUs (school) selections: 320 large schools, 80 medium schools and 50 small schools
3	Classes/students			2 Classes per grade in half of large schools; 1 class per grade otherwise

Acronyms: PSU = Primary Sampling Unit; SSU = Secondary Sampling Unit

As described in Section 1.2, the first stage of sampling selected PSUs within each stratum for a total of 100 sample PSUs. At the second sampling stage, a total of 450 Secondary Sampling Units

(SSUs), or schools, were selected from the sample PSUs, as follows: two large schools were selected per sample PSU, one per level (middle or high); an additional large school for each level was selected in a subsample of 60 PSUs, for a total of 320 large SSUs. An additional 80 medium SSUs and 50 small SSUs were selected from subsample PSUs, for a total of 450 sample SSUs (450 = 320 + 80 + 50). 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.

Depending on the average design effects, target subgroup sample sizes are between 1,200 and 1,700. The NYTS design has experienced lower design effects with less oversampling over the last few cycles (due to proportional allocation and enrollment size measures). Compared to previous cycles, the NYTS sampling design has had both lower effects on unequal weighting and smaller clustering effects. These factors lead to lower design effects, particularly for subgroups. Smaller design effects have, in turn, led to smaller variances and improved precision.

An appropriate sample size can enable generation of estimates with the required precision by grade, as well as by sex and school level. Therefore, the precision requirements generally focused on racial/ethnic subgroups within school level. The targets of n=700 students per racial and ethnic minority group by school level (1,400 total per group) correspond to prevalence estimates within +/- 5% for confidence intervals at 95% confidence for all key racial and ethnic subgroups when broken down by school level.

The prevalence estimates presented in Chapter 4 show that for all key racial and ethnic subgroups, prevalence estimates are within +/- 5% for confidence intervals at 95% confidence (i.e., standard errors are less than 2.5%). Standard errors are less than 2.5% for all estimates for Black and Hispanic students at the middle school and high school level.

2.2 Sampling Frame

As in previous cycles, the 2021 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 and ethnicity 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 NYTS.[3] Including schools sourced from the two NCES files resulted in substantial coverage increase among all public and non-public high schools.[4] 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.

[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

4 The coverage increase has ranged from 6.6% to 12.7% in recent sampling frames.

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 3.9% of schools (2.8% of public schools and 8.0% of private schools) and 1.1% of students. Lastly, schools were removed that had fewer than 40 students enrolled across eligible grades, resulting in the exclusion of 20.4% of schools (12.8% public and 42.6% private) which had been eligible after the other exclusions. This exclusion of schools with fewer than 40 students led to the exclusion of only 1.03% of students of those in eligible schools.[5]

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, 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.

Schools were stratified into small, medium, and large schools based on their ability to support less than one, one or two class selections per grade. Small SSUs contained fewer than 28 students at any grade level, and large SSUs contained at least 56 students at each grade level. The remaining schools were classified as medium sized.

The sampling stages may be summarized as follows:

- Selection of PSUs—One hundred PSUs (from approximately 1,257 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 320 large schools, or SSUs, were selected from the sample PSUs. Additionally, as described in Section 2.1, we selected 80 medium schools and 50 small schools, resulting in a total of 450 sample SSUs (450 = 320 + 80 + 50).
- 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.

The sampling approach utilized probability proportional to size (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

[5] Note that the two exclusions are additive so that 2.2% of students are excluded from the frame.

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. We selected two classes per grade in large schools and one class per grade in the remaining schools. The threshold for double class sampling was based on a simulation study to ensure that the required numbers of students in specified racial and ethnic minority groups were achieved per school level.

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 2.8 discusses the sample sizes attained in the fielded 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 120 large, 80 medium and 50 small schools from subsample PSUs.

Across 16 previous cycles of the NYTS that had concluded prior to the 2021 NYTS design, school participation had averaged 80.8% with a low of 49.9%.[6] Student participation had averaged 89.5% with a low of 85.9%. The combined response rate (student x school) averaged 72.4% with a low of 43.6%. Historical participation rates at both school and student levels guided the sampling design and sample sizes. In calculating the sample sizes for the 2021 NYTS, we made our approach more robust by assuming a conservative combined rate (student x school) of 42.5%, substantially lower than the historical overall response rate. The main reason is to account for higher levels of anticipated school refusals due to COVID-19 precautions in the 2020/2021 school environment. A secondary reason is that the student participation rate needs to be adjusted to account for a growing number of ineligible students. The number of ineligible students needs to be subtracted from the net number of students available for selection in participating schools. Table 2.2 presents a detailed derivation of the sample sizes *planned* for the 2021 NYTS based on these assumptions.

[6] Data collection during the 2020 NYTS was truncated due to school closures in response to COVID-19, resulting in lower school and student response rates than historical averages.

Table 2.2 Planned Sample Sizes for the 2021 NYTS

PSU	Size	# of SSUs	Number of Schools Sampled	# of Classes per School	# of Students per Class	# of Sampled Students prior to Attrition	# of Participating Students Based on 42.5% Response Rate
100 (with 60 PSUs subsampled to provide additional large schools)	Large HS	160	Double classes: 80	8	25	16,000	6,800
			Single classes: 80	4	25	8,000	3,400
	Large MS	160	Double classes: 80	6	25	12,000	5,100
			Single classes:80	3	25	6,000	2,550
	Large Total	**320**				**42,000**	**17,850**
40 (sub-sample)	Medium HS	40	40	4	25	4,000	1,700
	Medium MS	40	40	3	25	3,000	1,275
	Medium Total	**80**				**7,000**	**2,975**
25 (sub-sample)	Small HS	25	25	4	25	2,500	1,063
	Small MS	25	25	3	25	1,875	797
	Small Total	**50**				**4,375**	**1,859**
	Overall Total	**450**				**53,375**	**22,684**

One-hundred PSUs were selected, with two large SSUs ("full" schools) selected from each PSU and one additional large SSU per level selected from 60 subsampled PSUs for a total of 320 large SSUs. The estimated sample yield from these large schools was 42,000 students before school and student non-response, leading to an expected total 17,850 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) 80 medium SSUs from a subsample of 40 PSUs, and 50 small SSUs from a subsample of 25 PSUs were selected. The expected yield was 7,000 from medium schools and 4,375 students from small schools. In total, the expected number of participating students was 22,684.

Within each school, one class was selected from each grade to participate in the survey except in large schools with high racial and ethnic minority populations, where two classes per grade were selected. Note that the set of schools with high racial and ethnic minority populations defined for double class sampling is necessarily a subset of the large schools that can support such double class sampling. For the 2021 NYTS, we implemented double class selection for half of large schools (randomly selected) to ensure sufficient student yields.

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 to control the number of school districts contacted and recruited.
- Recent data should be available to characterize each PSU.
- 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 an SSU. Linked schools were treated as single schools during sampling.

2.6 STRATIFICATION

The PSUs were organized into 16 strata, based on urban/non-urban location and proportion racial and ethnic 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," for simplicity).
- 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[7] 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 and ethnic groupings changed from one survey cycle to the next. Table 2.3 presents the stratum boundaries used in the 2021 NYTS.

[7] The cumulative square root of "*f*" method developed by Dalenius and Hodges.

Table 2.3 Stratum Boundaries: Minority Percentage Cutoffs

Minority Concentration	Density Group	Bounds	
		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 2021 NYTS was designed to select a sample of 100 PSUs. The PSUs were initially 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 racial and ethnic minority density and urban status, alongside a proportional allocation. The initial proportional allocation was slightly modified 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/Rural	Density Group Number	Stratum Code	Student Population	Number of Sample PSUs (Revised)
Non-Hispanic Black	Urban	1	BU1	1,714,661	6
		2	BU2	1,853,137	8
		3	BU3	261,841	3
		4	BU4	507,699	5
	Non-urban	1	BR1	2,017,742	7
		2	BR2	1,163,987	4
		3	BR3	974,876	6
		4	BR4	501,663	6
Hispanic	Urban	1	HU1	3,333,126	8
		2	HU2	2,948,746	6
		3	HU3	2,354,412	5
		4	HU4	2,375,099	6
	Non-urban	1	HR1	5,501,371	18
		2	HR2	1,903,271	5
		3	HR3	972,022	4
		4	HR4	815,139	3

The sample was selected with 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 100 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 additional large schools (60 PSUs), medium schools (40 PSUs) and small schools (25 PSUs); sampling of two schools per level in each subsample PSU.

2.8 Sample Sizes Attained in the Survey

The 2021 NYTS attained the target sample sizes in the key analytic subgroups of interest. Tables 2.5a–d[8],show the number of participating students in subgroups defined by gender, grade, and race and ethnicity. Table 2.5d about race and ethnicity distribution, is presented in two different ways: 1) using the original variable allowing for multiple races and including missing data, and 2) using

[8] Percents may not add to exactly 100% due to rounding.

the variable whereby all respondents are categorized into a single race/ethnic group. The sample led to 5,056 Hispanic students and 3,446 Black students using the single-race variable.

Table 2.5a Subgroup Sample Sizes: Number of Participating Students

What is your sex?				
Q2	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Not Answered	121	0.59	121	0.59
Not Displayed	5	0.02	126	0.62
Male	10368	50.79	10494	51.41
Female	9919	48.59	20413	100.00

Table 2.5b Subgroup Sample Sizes: Number of Participating Students

What grade are you in?				
Q3	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Not Answered	71	0.35	71	0.35
Not Displayed	16	0.08	87	0.43
6th	3371	16.51	3458	16.94
7th	3198	15.67	6656	32.61
8th	3194	15.65	9850	48.25
9th	3097	15.17	12947	63.43
10th	2542	12.45	15489	75.88
11th	2595	12.71	18084	88.59
12th	2281	11.17	20365	99.76
Ungraded or other grade	48	0.24	20413	100.00

Table 2.5c Subgroup Sample Sizes: Number of Participating Students

RECODE: Race/Eth - multiple group				
RACE_M	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
<Missing>	631	3.09	631	3.09
NH-White	9232	45.23	9863	48.32
NH-Black	3280	16.07	13143	64.39
Hispanic	5056	24.77	18199	89.15
NH-Asian	851	4.17	19050	93.32
NH-AI/AN	223	1.09	19273	94.42
NH-NHOPI	62	0.30	19335	94.72
Multiple Races	1078	5.28	20413	100.00

Table 2.5d Subgroup Sample Sizes: Number of Participating Students

RECODE: Race/Eth - no multiple group				
RACE_S	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
<Missing>	631	3.09	631	3.09
NH-White	10104	49.50	10735	52.59
NH-Black	3446	16.88	14181	69.47
Hispanic	5056	24.77	19237	94.24
NH-Asian	889	4.36	20126	98.59
NH-AI/AN	225	1.10	20351	99.70
NH-NHOPI	62	0.30	20413	100.00

Note: This variable is named *race_s* in the public use data set. 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 2020-2021 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

- School name and relationship to identified district (if applicable)
- Name and title of 2020-2021 school principal
- School street address used for overnight deliveries, with city name and ZIP code
- Grade levels served during 2020-2021school year
- Approximate school enrollment
- At least a cumulative 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, meaning it does not draw its population from surrounding schools in order to provide specialized instruction
- 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 product prevention and control outcome indicators. The 2021 survey instrument included 166 questions. The 2021 NYTS represented the third cycle the study was conducted using electronic data collection methods rather than traditional paper-and-pencil (PAPI) and the first cycle that was conducted 100% online. The web survey was created using ColdFusion[9] and all data were stored in a MS SQL Server. To take the web survey, students navigated to a dedicated URL, nyts.cdc.gov, and entered a randomly generated, five-digit access code. Each access code was tied in a backend database to its associated school and classroom to facilitate tracking and calculate class and school response rates.

The survey followed a skip-pattern logic based on the student's responses to questions about ever and current tobacco product use behaviors. To improve students' sense of privacy, only one question was displayed on each screen so that responses to prior questions were not susceptible to observation. Students were given one class period (approximately 35-45 minutes) to complete the survey. Students who could not take the survey on the planned date for administration were asked to take the survey at the next possible opportunity.

The length of interview (LOI) was captured for each record and was calculated as the time lapse between the date/time of the first response and the date/time of the last response given. LOI ranged from 1 second to 23 days 59 minutes and 56 seconds, with an average of 26 minutes 16 seconds. After exclusion of outliers[10], the average survey completion time was 22 minutes 10 seconds.

The first five questions on the survey 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 product use; exposure to pro- and anti-tobacco media and advertising; minors' access to tobacco products; nicotine dependence; cessation attempts; exposure to second-hand smoke; harm perceptions; and exposure to tobacco product warnings. At the beginning of each tobacco product section, a description of the product (with example brands) and generic images of specific tobacco products were provided to assist with product recognition and increase the accuracy of student data. Students could refer back to this description and the images as needed as they answered related questions. The 2021 NYTS also included socio-demographic questions about family affluence, depression and anxiety, and sexual orientation and gender identity (SOGI).

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 partners, 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.

[9] https://www.adobe.com/products/coldfusion-family.html
[10] The average completion time was calculated after dropping outliers with survey duration lengths greater than 80 minutes (n=216).

3.2 External Review and Approvals

Three bodies reviewed and approved the instrumentation, processes, privacy and security elements, and sampling design of the 2021 NYTS: the Office of Management and Budget (OMB), ICF's Institutional Review Board (IRB) and CDC's Institutional Review Board (IRB).

With the transition to an electronic data collection format for the 2019 NYTS, the Security Assessment and Authorization (SA&A) approval and Enterprise Performance Life Cycle (EPLC) review remained valid for the 2021 NYTS cycle. The SA&A is a formal methodology for testing and evaluating the security controls of the system to ensure that it is configured properly to meet the security mandated by the Federal Information Security Management Act (FISMA). EPLC is a framework to enhance the Department of Health and Human Services (HHS) IT governance through rigorous application of sound investment and project management principals, in conjunction with industry's best practices.

3.3 Technical Assistance Provider (TAPs) Staffing

The role of the Technical Assistance Provider (TAP) was developed in response to anticipated complications due to COVID-19 that prohibited data collectors from conducting in-person survey administration. TAPs provided 100% virtual support to schools and teachers before, during, and after survey administration to 1) ensure teachers had received all the necessary materials to administer the survey, 2) answer any questions schools contacts and/or teachers may have prior to, during, or after survey administration, 3) ascertain that parental consent was properly obtained prior to the scheduled survey administration date, and 4) provide remote IT support, if needed. To ensure schools in various time zones would be adequately supported during school hours, TAPs were hired geographically across the country so that every time zone with sampled schools had at least one TAP local to that part of the country. TAPs were recruited from a pool of previously trained data collectors. A remote training for TAPs was conducted on January 5-6, 2021.

Key components of the training included the following:

- Pre-and post-survey communications with the schools and teachers
- Protocols specific to the type of instructional model (e.g., in-person, exclusively distance learning, hybrid) expected to be in place at the time of survey administration
- Orientation to student and teacher portals
- IT troubleshooting
- Communication with headquarters staff

3.4 Recruitment Procedures

The schools selected to participate in the 2021 NYTS were located in 37 states. Recruitment began in October 2020 with calls to state departments of education and health to inform them of the survey effort and sampled schools in their state. After notification at the state level, district- and school-level recruitment began. For public or diocesan schools, verbal or written agreement was first obtained by their district or diocese, respectively, before contact was made with the school. However, private schools were approached directly. A date for survey implementation was selected that was convenient to the school, its academic calendar, and, in some instances, the

anticipated date for return to in-person instruction. Recruiters and TAPs used a secure web-based calendar to facilitate communication and adjust survey dates upon request by the school.

3.5 Survey Administration

Survey administration began on January 18, 2021 after a comprehensive TAP training and continued through May 21, 2021. While the details of each data collection varied, there were six core steps followed for every school:

1) Conduct pre-contact call with the principal or lead contact to confirm instructional model (e.g., in-person, exclusively distance learning, hybrid)
2) Send tailored communications and survey materials to selected teachers in hard copy, electronically, or both, depending on instructional model
3) Outreach directly to teachers to confirm receipt of materials, verify intentions to administer the survey on the scheduled date, confirm parental consent procedures were followed, and provide additional survey instructions
4) Virtually monitor survey activities and respond to requests for technical support, as needed
5) Follow-up with teachers regarding student response rates and class enrollment
6) Report final progress to school contact and thank them for their school's participation

Procedures were designed to protect students' privacy by assuring that student participation was anonymous and voluntary. Using a school-issued or personal internet-connected device, students logged into a secure website, watched a brief 2-minute instructional video, and responded to a question regarding their location (e.g., classroom, home, other location) before completing the survey. All surveys were submitted directly to the secure SQL server.

3.5.1 Field Procedures

After schools had been recruited, classes selected, and a date for survey administration scheduled, each school with in-person or hybrid instructional models received a hard-copy mailing with pre-survey materials containing instructions for the school contact and packets for the teacher of each selected class. Teacher packets contained the parental permission forms to be distributed to all students in the selected classes prior to data collection. School with exclusively distance learning models received their materials electronically. They, in turn, distributed the parental permission forms to students electronically following local established channels for teacher/parent communications.

The timing of these pre-survey materials was determined in part by the type of permission form being used by the school; this decision was made by the school district or individual school. Passive parental permission forms (i.e., forms returned only if the parents do not want their child to participate) were sent approximately two weeks 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 for the child to participate) were sent out four weeks prior to the scheduled date of data collection for schools that require active consent. TAPs conducted follow-up calls and emails to the selected schools to answer any questions and to make sure materials were received and distributed to selected classes and students.

3.5.2 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 or advisory periods were used as the frame for class selection.

3.6 Web-Based Data Collection Management Application (DCMA)

For multiple cycles of the NYTS, a web-based data collection management application (DCMA) has been utilized to help centralize the management of the study, facilitate information exchange with project staff, and allow all members of the project management, recruitment, supervisory teams, and remote 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, registering student records submitted to the central repository, and tracking school and student response rates.

3.7 Data Recording

Preliminary student participation rates were calculated based on class enrollment numbers provided by teachers of selected classes and the number of surveys received in the central repository. If teachers reported a different number of expected completes than what was received in the central repository, a TAP followed up to resolve discrepancies and determine additional strategies to maximize student participation. As additional surveys were received after the initial survey administration date, the DCMA automatically updated the number of actual records received and participation reporting was revised accordingly.

3.8 Participation Rates

Participation rates for the NYTS were calculated at the school and student levels.

3.8.1 School-level Participation Rates

At the school level, 508 schools were selected across 243 districts in 37 states. During sample validation, 30 schools were deemed to be ineligible and were replaced.

In total, 279 schools (54.9%) participated in the study. The remaining 229 schools were considered refusals. Of refusals, 106 of them were due to district-level refusals to allow contact with schools

to discuss participation and 123 were school level refusals. The most common reasons given for a refusal at the district or school level were COVID-related concerns, loss of instructional time and burden to teachers/staff.

3.8.2 Student-level Participation Rates

Initial student-level participation rates were calculated from the field as teachers reported enrollment information and submitted surveys registered in the central repository. In subsequent follow-ups between teachers and TAPs, further refinements were made to 1) revise the number of eligible students based on available documentation, 2) correct mathematical errors, 3) review counts of surveys received by the database, and 4) account for make-ups as they were received from students and classes that did not participate on the initial day of survey administration.

The final student participation rate for the 2021 NYTS was 81.2%. Overall, 25,149 eligible students from the 279 participating schools were invited to participate in the survey, and 20,413 did so. Table 3.1 below shows the number of eligible students, participants, and participation rates for the NYTS.

Table 3.1 Overall NYTS 2021 Student Participation Rate

	# Eligible	# Completed	Participation %
NYTS Participating Students	25,149	20,413	81.2%

The 2021 NYTS survey attained an actual school participation rate of 54.9% and a student participation rate of 81.2%. The overall participation rate, the product of the school-level and student-level participation rates, was 44.6%.

3.9 Data Management

To take advantage of the electronic format of the NYTS, the dataset was designed to be self-cleaning based on programming logic. However, to ensure accuracy, 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 and ethnicity values to match CDC-required classifications. Data "missingness" was categorized into one of four types: as a legitimate skip based on programmed logic, as item-level refusal if a question was presented to a student on-screen but not answered, as not answered because the student was never shown a question on screen (e.g., partial complete), or as recoded to missing due to edit checks. Missingness is distinguished in the data set as follows:

- .S – Legitimate skip
- .N – Displayed, not answered (item-level refusal)
- .Z – Not displayed (partial complete)
- .E – Missing due to edit check

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.

CHAPTER 4—WEIGHTING OF NYTS RESPONSE DATA

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

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

This chapter 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. This chapter closes with a description of the computation of weighted estimates and variance estimates. Details of the student weights are provided in Appendix C.

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 those large schools belonging to the 60 subsampled PSUs, the partial school weight was:

$$W^{LS}{}_{klm} = 2\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{LS}{}_{klm}}$$

For small schools, the partial school weight was:

$$W^{SS}{}_{klm} = (100/25)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{SS}{}_{klm}}$$

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

$$W^{MS}{}_{klm} = (100/40)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{MS}{}_{klm}}$$

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 post stratified 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 schools. 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 2021 NYTS cycle, nonresponse adjustment cells were defined in a tailored and systematic approach stemming from the non-response analysis. These analyses are detailed in the *2021 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 Step. 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 2021 cycle, school type, school affluence (MDR index), and dichotomized NCES locale (city versus non-city) were found to be predictive on nonresponse. Nonresponse adjustment cells were created using school level (high vs middle), school type (public vs non-public), affluence (above median vs below median, or high vs low) and dichotomized NCES locale (city vs non-city).

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. 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. If enrollment by grade and sex is not available for certain schools, only adjustments by grade or school level were performed.

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. The adjustment cells were defined differently for public and non-public schools. Non-public schools were divided by school level; public schools were divided by school level, affluence, and dichotomized NCES locale.

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
High school, non-public	245,440.08	7	1,063,436.95	26	26.923	4.333
High school, public, low affluence, city	1,756,526.31	17	3,116,274.84	34	50.000	1.774
High school, public, low affluence, non-city	4,353,522.73	45	6,181,055.23	63	71.429	1.420
High school, public, high affluence, city	1,265,427.50	15	3,606,948.88	36	41.667	2.850
High school, public, high affluence, non-city	5,833,814.85	46	10081424.07	78	58.974	1.728
Middle school, non-public	548,146.23	16	1,362,112.69	32	50.000	2.485
Middle school, low affluence, city	925,410.79	15	2,457,503.60	36	41.667	2.656
Middle school, low affluence, non-city	3,133,665.68	52	4,138,572.64	68	76.471	1.321
Middle school, affluence above median, city	319,192.58	8	2,385,137.98	34	23.529	7.472
Middle school, affluence above median, non-city	5,172,567.25	60	8,798,941.23	103	58.252	1.701
	23553713.99	**281**	**43191408.11**	**510**[11]		

The variables considered in the non-response analyses which led to non-response adjustment cells are more fully described in the non-response analysis report. The four variables used in non-response adjustment cells are school level (middle vs high school), school type (public vs non-public), affluence (low vs high affluence; from sampling frame (MDR) Index) and dichotomized NCES locale (city vs non-city).

[11] Two schools were span schools and are therefore counted twice as different sampling units, i.e., both as middle schools and high schools

4.3 POST-STRATIFICATION AND TRIMMING

The final steps in the weighting process include trimming and post-stratification. 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 post-stratification, or raking, procedures described next.

Post-stratification 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 2017-18 CCD and the 2017-18 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).[12] 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 post-stratum cell. Post-stratification variables, also shown in the two tables, are a) school type by grade and sex, and b) census region by grade and race and ethnicity. These dimensions reflect the iterations used in the raking procedures.

[12] 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	Sex*	Number of Records	Weight Sum=Control Total
Public	6	Male	1503	1,947,924.89
Public	6	Female	1492	1,856,189.11
Public	7	Male	1575	1,920,121.91
Public	7	Female	1404	1,827,816.09
Public	8	Male	1536	1,912,080.43
Public	8	Female	1430	1,823,155.03
Public	9	Male	1424	2,002,535.13
Public	9	Female	1484	1,879,501.87
Public	10	Male	1234	1,893,367.30
Public	10	Female	1235	1,810,373.17
Public	11	Male	1187	1,783,595.77
Public	11	Female	1285	1,738,184.23
Public	12	Male	1130	1,717,522.26
Public	12	Female	1094	1,690,411.26
Private	6	Combined	385	254,851.00
Private	7	Combined	256	252,753.00
Private	8	Combined	255	254,508.54
Private	9	Combined	213	254,889.00
Private	10	Combined	84	250,741.54
Private	11	Combined	142	249,096.00
Private	12	Combined	65	244,189.48

*Sex is combined for private schools due to small cell sizes.

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

Census Region	Grade	Race/Hispanic Origin	Number of Records	Weight Sum = Control Total
Northeast	6	Combined[13]	162	415,523.32
Northeast	6	Non-Hispanic Black	39	86,745.53
Northeast	6	Hispanic	75	133,417.15
Northeast	7	Combined	151	418,712.68
Northeast	7	Non-Hispanic Black	59	85,586.22
Northeast	7	Hispanic	67	128,800.10
Northeast	8	Combined	110	429,840.68
Northeast	8	Non-Hispanic Black	49	84,693.89
Northeast	8	Hispanic	53	124,588.44
Northeast	9	Combined	279	429,764.98
Northeast	9	Hispanic	172	221,204.02
Northeast	10	Combined	236	421,659.41
Northeast	10	Hispanic	135	208,068.59
Northeast	11	Combined	212	420,914.72
Northeast	11	Non-Hispanic Black	81	79,853.63
Northeast	11	Hispanic	95	107,284.65
Northeast	12	Combined	147	422,603.90
Northeast	12	Non-Hispanic Black	79	77,933.10
Northeast	12	Hispanic	85	99,937.00
Midwest	6	Combined	777	617,855.97
Midwest	6	Non-Hispanic Black	99	116,982.51
Midwest	6	Hispanic	119	112,706.51
Midwest	7	Combined	705	617,522.21
Midwest	7	Non-Hispanic Black	97	111,331.97
Midwest	7	Hispanic	115	109,222.82
Midwest	8	Combined	832	626,369.66
Midwest	8	Non-Hispanic Black	106	111,017.39
Midwest	8	Hispanic	114	107,145.94
Midwest	9	Combined	570	638,225.42
Midwest	9	Non-Hispanic Black	77	118,931.71
Midwest	9	Hispanic	89	109,560.87
Midwest	10	Combined	509	622,736.06
Midwest	10	Non-Hispanic Black	60	112,062.43
Midwest	10	Hispanic	85	103,957.51
Midwest	11	Combined	543	609,640.87
Midwest	11	Non-Hispanic Black	86	104,628.70
Midwest	11	Hispanic	73	96,923.43
Midwest	12	Combined	434	610,408.17
Midwest	12	Non-Hispanic Black	100	101,900.04

[13] For Northeast region grade 9 and grade 10, Non-Hispanic Black and Hispanic were collapsed due to cell size. The "Combined" category includes American Indian + Asian + White.

Census Region	Grade	Race/Hispanic Origin	Number of Records	Weight Sum = Control Total
Midwest	12	Hispanic	76	89,502.78
South	6	Combined	651	792,578.51
South	6	Non-Hispanic Black	394	363,665.63
South	6	Hispanic	365	423,494.86
South	7	Combined	599	790,881.27
South	7	Non-Hispanic Black	415	351,680.50
South	7	Hispanic	323	409,341.22
South	8	Combined	576	792,573.59
South	8	Non-Hispanic Black	433	344,856.86
South	8	Hispanic	312	399,009.55
South	9	Combined	526	825,891.15
South	9	Non-Hispanic Black	438	377,787.74
South	9	Hispanic	325	426,616.12
South	10	Combined	427	793,577.60
South	10	Non-Hispanic Black	335	349,946.57
South	10	Hispanic	213	390,866.83
South	11	Combined	407	766,118.94
South	11	Non-Hispanic Black	320	324,368.04
South	11	Hispanic	202	355,490.02
South	12	Combined	330	738,315.45
South	12	Non-Hispanic Black	355	307,965.35
South	12	Hispanic	170	315,158.20
West	6	Combined	338	508,970.09
West	6	Non-Hispanic Black	49	48,124.99
West	6	Hispanic	312	438,899.92
West	7	Combined	368	505,670.20
West	7	Non-Hispanic Black	54	47,309.12
West	7	Hispanic	282	424,632.68
West	8	Combined	311	509,287.52
West	8	Non-Hispanic Black	41	47,017.07
West	8	Hispanic	284	413,343.41
West	9	Combined	352	517,076.59
West	9	Non-Hispanic Black	20	48,771.84
West	9	Hispanic	273	423,095.57
West	10	Combined	255	500,098.97
West	10	Non-Hispanic Black	12	47,541.52
West	10	Hispanic	286	403,966.51
West	11	Combined	285	487,857.86
West	11	Non-Hispanic Black	14	44,757.57
West	11	Hispanic	296	373,037.57
West	12	Combined	239	485,680.96
West	12	Non-Hispanic Black	18	45,316.44
West	12	Hispanic	256	357,401.60

For poststratification purposes, a unique race and ethnicity was assigned to respondents with missing data on race and ethnicity, those with an "Other" classification, and those reporting multiple races.

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=90.45% implies that the design-effect (DEFF) component due to unequal weighing effects is 1.82.[14]

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[15] 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.[16]

Tables 4.4–4.7 present weighted estimates and estimated standard errors for key outcome measures using the 2021 NYTS data. Sample SAS code is provided in Exhibit 4.1.

[14] The design effect due to unequal weighting may be expressed in terms of the cv of the weight as DEFF= 1 + cv**2.
[15] Skinner CJ, Holt D, and Smith TMF, Analysis of Complex Surveys, John Wiley & Sons, New York, 1989, 50.

Exhibit 4.1: Example SAS and SUDAAN Code for Generating Weighted Tobacco Product Use Estimates (ever use, current use) and Standard Errors*

```
SAS:
Proc Surveymeans Data=nyts2021 mean;
Var eelcigt ecigt ecigar eslt ehookah erollcigts epipe esnus edissolv ebidis ehtp epouch
        celcigt ccigt ccigar cslt chookah crollcigts cpipe csnus cdissolv cbidis chtp cpouch;
Class eelcigt ecigt ecigar eslt ehookah erollcigts epipe esnus edissolv ebidis ehtp epouch
        celcigt ccigt ccigar cslt chookah crollcigts cpipe csnus cdissolv cbidis chtp cpouch;
Stratum v_stratum2;
Cluster psu2;
Weight finwgt;
Domain SCHOOLTYPE SCHOOLTYPE*Sex SCHOOLTYPE*Race_S;
Title “NYTS 2021, Tobacco Product Use Estimates by School Type, by School Type and Sex Cross-Classified,
and by School Type and Race and ethnicity Cross-Classified”;
run;


SUDAAN:
Proc Descript Data=nyts2020 Filetype= SAS Design=WR;
Var eelcigt ecigt ecigar eslt ehookah erollcigts epipe esnus edissolv ebidis ehtp epouch
        celcigt ccigt ccigar cslt chookah crollcigts cpipe csnus cdissolv cbidis chtp cpouch;
Catlevel 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1;
Nest v_stratum2 PSU2 / Missunit;
Weight finwgt;
Subgroup SCHOOLTYPE Sex Race_S;
Levels 2 2 3;
Tables SCHOOLTYPE  SCHOOLTYPE*Sex SCHOOLTYPE*Race_S;
Title “NYTS 2020, Tobacco Product Use 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;
```

Example SAS and SUDAAN code will generate estimates of ever use and current (past 30-day use) of e-cigarettes, cigarettes, cigars, smokeless tobacco products (chewing tobacco, snuff, or dip), hookah tobacco, roll-your-own cigarettes, pipe tobacco, snus, dissolvable tobacco, bidis, heated tobacco products, and nicotine pouches.

Table 4.4 Current (past 30-day) Use Estimates for Selected Tobacco Products for High School Students[17]

Product	Overall %(SE) N=10,515	Female %(SE) N=5,172	Male %(SE) N=5,312	White %(SE) N=5,334	Black %(SE) N=1,857	Hispanic %(SE) N=2,684
Bidis	0.41% (0.08%) N=51	0.52% (0.14%) N=30	0.32% (0.09%) N=21	0.37% (0.10%) N=22	0.81% (0.31%) N=14	0.35% (0.14%) N=13
Cigars, little cigars, or cigarillos	2.11% (0.19%) N=224	1.53% (0.23%) N=96	2.62% (0.28%) N=127	2.06% (0.23%) N=103	4.45% (0.65%) N=79	1.17% (0.20%) N=31
Cigarettes	1.89% (0.20%) N=198	1.76% (0.25%) N=83	1.99% (0.30%) N=114	2.25% (0.26%) N=122	0.97% (0.35%) N=18	1.63% (0.26%) N=46
Dissolvable tobacco products	0.19% (0.06%) N=22	0.22% (0.09%) N=12	0.16% (0.07%) N=10	0.18% (0.07%) N=8	0.31% (0.21%) N=5	0.15% (0.08%) N=7
Electronic cigarettes	11.27% (0.80%) N=1116	11.86% (0.99%) N=576	10.72% (0.85%) N=537	14.45% (1.08%) N=760	5.87% (0.99%) N=87	7.60% (0.90%) N=214
Hookah or waterpipe	1.21% (0.18%) N=128	1.29% (0.30%) N=69	1.15% (0.16%) N=59	0.76% (0.13%) N=41	3.18% (0.78%) N=49	1.33% (0.37%) N=34
Pipe tobacco	0.45% (0.10%) N=45	0.41% (0.12%) N=19	0.48% (0.15%) N=25	0.50% (0.14%) N=24	0.58% (0.31%) N=7	0.40% (0.12%) N=14
Heated tobacco products	0.82% (0.12%) N=81	0.74% (0.18%) N=39	0.93% (0.19%) N=42	0.92% (0.18%) N=50	0.84% (0.34%) N=12	0.73% (0.24%) N=14
Roll-your-own cigarettes	0.58% (0.09%) N=68	0.67% (0.14%) N=41	0.51% (0.13%) N=27	0.59% (0.12%) N=35	0.80% (0.32%) N=14	0.57% (0.17%) N=18
Smokeless tobacco (chewing tobacco, snuff, or dip)	0.96% (0.17%) N=106	0.26% (0.10%) N=18	1.59% (0.27%) N=88	1.28% (0.23%) N=75	0.25% (0.14%) N=5	0.49% (0.19%) N=16
Snus	0.46% (0.10%) N=49	0.26% (0.09%) N=13	0.63% (0.15%) N=36	0.57% (0.15%) N=30	0.01% (0.01%) N=1	0.26% (0.10%) N=12
Nicotine pouches	1.10% (0.19%) N=101	0.55% (0.15%) N=25	1.60% (0.31%) N=76	1.41% (0.31%) N=68	0.61% (0.26%) N=8	0.66% (0.21%) N=18

Note: In the dataset, variables associated with current use of each tobacco product are as follows: electronic cigarettes (celcigt); cigars, little cigars, or cigarillos (ccigar); cigarettes (ccigt); smokeless tobacco (cslt); hookah or waterpipe (chookah); roll-your-own cigarettes (crollcigts); snus (csnus); pipe tobacco (cpipe); dissolvable tobacco products (cdissolv); bidis (cbidis); heated tobacco products (chtp); and nicotine pouches (cpouch).

[17] The estimates in tables 4.4–4.7 use the variable SCHOOLTYPE, which is coded as 1 (Middle School) if QN3 ranges from 1 to 3, and 2 (High School) if QN3 ranges from 4 – 7. Students who are missing QN3 are excluded from these tables.

Table 4.5 Current Use Estimates for Selected Tobacco Products for Middle School Students[18]

Product	Overall %(SE) N=9,763	Female %(SE) N=4,699	Male %(SE) N=5,015	White %(SE) N=4,739	Black %(SE) N=1,575	Hispanic %(SE) N=2,360
Bidis	0.17% (0.04%) N=22	0.19% (0.07%) N=10	0.16% (0.06%) N=12	0.01% (0.01%) N=2	0.49% (0.16%) N=9	0.33% (0.13%) N=10
Cigars, little cigars, or cigarillos	0.59% (0.10%) N=76	0.54% (0.13%) N=32	0.65% (0.12%) N=44	0.46% (0.11%) N=25	1.37% (0.35%) N=24	0.62% (0.18%) N=21
Cigarettes	1.04% (0.16%) N=117	1.17% (0.25%) N=63	0.89% (0.19%) N=53	1.05% (0.20%) N=59	1.09% (0.31%) N=21	1.18% (0.29%) N=26
Dissolvable tobacco products	0.13% (0.05%) N=16	0.13% (0.06%) N=8	0.14% (0.07%) N=8	0.05% (0.03%) N=5	0.28% (0.15%) N=5	0.23% (0.14%) N=5
Electronic cigarettes	2.77% (0.30%) N=310	3.18% (0.45%) N=166	2.35% (0.28%) N=141	2.61% (0.38%) N=154	2.33% (0.62%) N=39	3.93% (0.61%) N=93
Heated tobacco products	0.42% (0.09%) N=56	0.39% (0.12%) N=22	0.43% (0.12%) N=31	0.25% (0.08%) N=19	0.59% (0.21%) N=15	0.69% (0.22%) N=18
Hookah or waterpipe	0.39% (0.10%) N=41	0.53% (0.18%) N=21	0.25% (0.10%) N=19	0.15% (0.05%) N=12	0.94% (0.30%) N=12	0.57% (0.18%) N=13
Pipe tobacco	0.18% (0.05%) N=19	0.21% (0.09%) N=9	0.16% (0.07%) N=10	0.18% (0.09%) N=8	0.10% (0.07%) N=3	0.27% (0.13%) N=6
Nicotine pouches	0.30% (0.08%) N=35	0.44% (0.15%) N=20	0.17 (0.05) N=15	0.14% (0.05%) N=10	0.50% (0.23%) N=7	0.56% (0.21%) N=16
Roll-your-own cigarettes	0.36% (0.07%) N=50	0.36% (0.09%) N=26	0.33% (0.10%) N=23	0.19% (0.07%) N=16	0.67% (0.23%) N=13	0.68% (0.21%) N=20
Smokeless tobacco (chewing tobacco, snuff, or dip)	0.47% (0.09%) N=62	0.27% (0.10%) N=13	0.67% (0.14%) N=49	0.45% (0.13%) N=28	0.25% (0.13%) N=7	0.56% (0.18%) N=17
Snus	0.16% (0.04%) N=26	0.11% (0.06%) N=7	0.20% (0.06%) N=19	0.13% (0.05%) N=13	0.10% (0.07%) N=3	0.27% (0.12%) N=8

Note: In the dataset, variables associated with current use of each tobacco product are as follows: electronic cigarettes (celcigt); cigars, little cigars, or cigarillos (ccigar); cigarettes (ccigt); smokeless tobacco (cslt); hookah or waterpipe (chookah); roll-your-own cigarettes (crollcigts); snus (csnus); pipe tobacco (cpipe); dissolvable tobacco products (cdissolv); bidis (cbidis); heated tobacco products (chtp); and nicotine pouches (cpouch).

[18] The estimates in tables 4.4–4.7 use the variable SCHOOLTYPE, which is coded as 1 (Middle School) if QN3 ranges from 1 to 3, and 2 (High School) if QN3 ranges from 4 – 7. Students who are missing QN3 are excluded from these tables.

Table 4.6 Ever Use Estimates for Selected Tobacco Products for High School Students[19]

Product	Overall %(SE) N=10,515	Female %(SE) N=5,172	Male %(SE) N=5,312	White %(SE) N=5,334	Black %(SE) N=1,857	Hispanic %(SE) N=2,684
Bidis	0.96% (0.13%) N=108	1.06% (0.19%) N=58	0.88% (0.20%) N=50	0.78% (0.15%) N=44	1.45% (0.46%) N=22	1.06% (0.24%) N=35
Cigars, little cigars, or cigarillos	7.63% (0.50%) N=757	4.93% (0.38%) N=273	10.03% (0.75%) N=482	8.72% (0.78%) N=437	8.67% (0.90%) N=163	5.37% (0.60%) N=126
Cigarettes	11.16% (0.63%) N=1095	10.81% (0.77%) N=517	11.51% (0.82%) N=576	13.01% (0.89%) N=655	6.36% (0.94%) N=126	9.53% (1.01%) N=255
Dissolvable tobacco products	0.52% (0.10%) N=53	0.54% (0.12%) N=30	0.51% (0.14%) N=23	0.62% (0.14%) N=29	0.31% (0.21%) N=5	0.44% (0.15%) N=15
Electronic cigarettes	28.86% (1.23%) N=2887	30.16% (1.43%) N=1502	27.70% (1.39%) N=1375	33.85% (1.59%) N=1747	16.93% (1.31%) N=306	25.04% (1.39%) N=694
Heated tobacco products	2.27% (0.16%) N=191	2.45% (0.29%) N=97	2.11% (0.20%) N=93	2.30% (0.22%) N=99	2.13% (0.59%) N=28	2.36% (0.43%) N=52
Hookah or waterpipe	3.85% (0.34%) N=419	4.22% (0.47%) N=237	3.45% (0.36%) N=180	2.93% (0.36%) N=149	6.81% (0.84%) N=126	4.00% (0.51%) N=116
Pipe tobacco	1.45% (0.20%) N=136	1.14% (0.25%) N=54	1.73% (0.28%) N=81	1.79% (0.26%) N=88	0.80% (0.34%) N=12	1.14% (0.25%) N=31
Nicotine pouches	2.97% (0.30%) N=269	1.74% (0.28%) N=75	4.09% (0.46%) N=194	4.13% (0.44%) N=205	1.11% (0.36%) N=13	1.52% (0.28%) N=40
Roll-your-own cigarettes	2.49% (0.23%) N=258	2.47% (0.27%) N=129	2.51% (0.31%) N=128	2.56% (0.33%) N=145	2.05% (0.45%) N=36	2.47% (0.34%) N=59
Smokeless tobacco (chewing tobacco, snuff, or dip)	4.08% (0.48%) N=419	1.70% (0.35%) N=89	6.23% (0.68%) N=330	5.73% (0.64%) N=314	1.41% (0.43%) N=25	1.68% (0.34%) N=52
Snus	1.92% (0.26%) N=183	0.81% (0.18%) N=38	2.93% (0.41%) N=145	2.88% (0.36%) N=145	0.01% (0.01%) N=1	0.74% (0.21%) N=25

Note: In the dataset, variables associated with ever use of each tobacco product are as follows: electronic cigarettes (eelcigt); cigars, little cigars, or cigarillos (ecigar); cigarettes (ecigt); smokeless tobacco (eslt); hookah or waterpipe (ehookah); roll-your-own cigarettes (erollcigts); snus (esnus); pipe tobacco (epipe); dissolvable tobacco products (edissolv); bidis (ebidis); heated tobacco products (ehtp); and nicotine pouches (epouch).

[19] The estimates in tables 4.4–4.7 use the variable SCHOOLTYPE, which is coded as 1 (Middle School) if QN3 ranges from 1 to 3, and 2 (High School) if QN3 ranges from 4 – 7. Students who are missing QN3 are excluded from these tables.

Table 4.7 Ever Use Estimates for Selected Tobacco Products for Middle School Students[20]

Product	Overall %(SE) N=9,763	Female %(SE) N=4,699	Male %(SE) N=5,015	White %(SE) N=4,739	Black %(SE) N=1,575	Hispanic %(SE) N=2,360
Bidis	0.40% (0.09%) N=45	0.50% (0.14%) N=23	0.30% (0.08%) N=22	0.19% (0.10%) N=10	0.71% (0.20%) N=13	0.72% (0.23%) N=20
Cigars, little cigars, or cigarillos	2.09% (0.23%) N=227	2.03% (0.31%) N=99	2.17% (0.26%) N=128	1.66% (0.27%) N=88	4.30% (0.66%) N=72	2.47% (0.54%) N=53
Cigarettes	4.08% (0.46%) N=451	4.24% (0.60%) N=221	3.92% (0.47%) N=229	3.84% (0.53%) N=209	5.67% (1.13%) N=94	4.06% (0.75%) N=102
Dissolvable tobacco products	0.40% (0.10%) N=40	0.43% (0.15%) N=17	0.37% (0.10%) N=23	0.39% (0.14%) N=19	0.43% (0.17%) N=9	0.55% (0.24%) N=11
Electronic cigarettes	7.28% (0.60%) N=765	7.50% (0.73%) N=384	7.02% (0.68%) N=373	7.51% (0.68%) N=385	6.97% (1.21%) N=110	8.36% (1.08%) N=205
Heated tobacco products	1.12% (0.15%) N=111	0.98% (0.19%) N=51	1.25% (0.21%) N=57	0.83% (0.19%) N=41	1.02% (0.38%) N=17	1.75% (0.48%) N=39
Hookah or waterpipe	1.31% (0.23%) N=138	1.53% (0.33%) N=79	1.07% (0.19%) N=57	0.63% (0.15%) N=49	2.67% (0.50%) N=36	2.14% (0.60%) N=37
Pipe tobacco	0.49% (0.09%) N=60	0.45% (0.14%) N=24	0.53% (0.13%) N=36	0.37% (0.10%) N=24	0.43% (0.21%) N=10	0.69% (0.19%) N=18
Nicotine pouches	0.57% (0.11%) N=69	0.60% (0.18%) N=30	0.54% (0.12%) N=39	0.41% (0.10%) N=29	0.18% (0.10) N=5	0.99% (0.30%) N=26
Roll-your-own cigarettes	1.12% (0.17%) N=131	1.18% (0.23%) N=67	1.04% (0.21%) N=63	0.84% (0.18%) N=55	1.10% (0.35%) N=23	1.87% (0.44%) N=44
Smokeless tobacco (chewing tobacco, snuff, or dip)	1.87% (0.29%) N=204	1.60% (0.31%) N=73	2.15% (0.37%) N=131	1.82% (0.27%) N=107	1.54% (0.51%) N=27	2.03% (0.46%) N=45
Snus	0.35% (0.07%) N=52	0.29% (0.09%) N=16	0.40% (0.10%) N=36	0.30% (0.08%) N=26	0.10% (0.07%) N=3	0.44% (0.16%) N=15

Note: In the dataset, variables associated with ever use of each tobacco product are as follows: electronic cigarettes (eelcigt); cigars, little cigars, or cigarillos (ecigar); cigarettes (ecigt); smokeless tobacco (eslt); hookah or waterpipe (ehookah); roll-your-own cigarettes (erollcigts); snus (esnus); pipe tobacco (epipe); dissolvable tobacco products (edissolv); bidis (ebidis); heated tobacco products (ehtp); and nicotine pouches (epouch).

[20] The estimates in tables 4.4–4.7 use the variable SCHOOLTYPE, which is coded as 1 (Middle School) if QN3 ranges from 1 to 3, and 2 (High School) if QN3 ranges from 4 – 7. Students who are missing QN3 are excluded from these tables.

APPENDIX A. QUESTIONNAIRE

Questionnaire only included in PDF version of this document.

APPENDIX B. COMMON CORE OF DATA RACE AND 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.

APPENDIX C. 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:

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