2014 National Youth Tobacco Survey

Methodology Report

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

Prepared for Centers for Disease Control and Prevention

Prepared by ICF International, Rockville, Maryland

Centers for Disease Control and Prevention

Office on Smoking and Health

Atlanta, GA

September 2015

Recommended Citation: Office on Smoking and Health. *2014 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, 2015.

PRINT ONLY

Table of Contents

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 of tobacco-related indicators for both middle school (grades 6–8) and high school (grades 9–12) students. Tobacco-related indicators included in the NYTS include: tobacco use (e.g., bidis, cigarettes, cigars, kreteks, tobacco pipes, smokeless tobacco, snus, dissolvable tobacco products, hookahs, and electronic cigarettes); exposure to secondhand smoke; smoking cessation; school curriculum; 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 spring 2000, 2002, 2004, 2006, 2009, 2011, 2012, 2013 and 2014, the NYTS provides data that are representative of all middle school and high school students in the 50 U.S. 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 annually administer the NYTS.

1.2 Overview of the 2014 NYTS Methodology

The 2014 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. Non-Hispanic black students were oversampled. 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 created 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. Schools used either passive or active permission forms at their

[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] MacDonald, G., Starr, G., Schooley, M., Yee, S. L., Klimowski. K., Turner, K. (2001). *Introduction to program evaluation for comprehensive tobacco control programs*. Atlanta, GA: US Department of Health and Human Services, CDC.

discretion to fulfill requirements of the No Child Left Behind Act, whereby parents must be provided with a means to opt out of their child's participation.

The final sample consisted of 258 schools, of which 207 participated, yielding a school participation rate of 80.2%. A total of 22,007 student questionnaires were completed out of a sample of 24,084 students, yielding a student participation rate of 91.4%. The overall participation rate, defined as the product of the school-level and student-level participation rates, was 73.3%.

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 2014 NYTS sample selection (Chapter 2), data collection (Chapter 3), and weighting of student response data (Chapter 4).

CHAPTER 2—NYTS SAMPLING METHODS

2.1 Overview

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 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). Additional estimates were also supported for subgroups defined by grade, by sex, and by race/ethnicity, each within school level domains; however, precision levels will vary 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, 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 2014 NYTS is a continuation of the NYTS cycles that took place in 1999, 2000, 2002, 2004, 2006, 2009, 2011, 2012 and 2013. The NYTS employs a repeat cross-sectional design to develop national estimates of tobacco use behaviors and exposure to pro- and antitobacco influences among students enrolled in grades 6–12. The general sampling design framework used for the 2013 NYTS was also employed for the 2014 NYTS.

2.1.1 Oversampling of Racial/Ethnic Minorities

To facilitate accurate prevalence estimates among racial/ethnic minority groups, prior cycles of the NYTS have employed multiple strategies to increase the number of non-Hispanic black and Hispanic students included in the sample. These approaches have included over-sampling PSUs in high racial/ethnic minority strata, the use of a weighted measure of size (MOS), and double class selection in large schools that contain a sufficient proportion of minority students.

The design development process examines parameters such as thresholds for double class selection and PSU allocation to strata, to balance the dual goals of overall precision and minority group targets.

The sampling design balances increasing yields for minority students with overall precision as oversampling leads to larger variances for overall estimates. As described below, the only oversampling that remains in the more efficient design for the 2014 NYTS is double class sampling. This method has been shown to reduce design effects for survey estimates. The design effect, defined as the variance of actual survey estimates divided by the variance of a simple random sample of the same size, is a common useful measure of the precision of survey estimates.

A weighted measure of size (MOS) was previously used to increase the probability of selection of high racial/ethnic minority PSUs and schools using a Probability Proportional to Size (PPS) sampling design. The effectiveness of a weighted MOS in achieving oversampling is dependent

upon the distributions of non-Hispanic black and Hispanic students in schools. The need for a weighted MOS is predicated on a relatively low prevalence of minority students in the population; however, this premise has become less tenable with the increase of nonwhite students in the population overall, and specifically, Hispanic students. The need for oversampling Hispanic students has been gradually reduced with the increasing numbers of Hispanics among the student population. As seen below, some degree of oversampling non-Hispanic black students remains in the sampling design.

In 1990, the contactor, ICF (formerly Macro International Inc.), conducted the first in a series of simulation studies to investigate the impact of various weighting functions on the numbers and percentages of racial/ethnic minority students reached in YRBS.[3] Sampling strategies based on this work were incorporated into the NYTS, and these simulations have been updated with each cycle of the NYTS to ensure that the minimum amount of weighting in the MOS is being used, while still achieving adequate representation of non-Hispanic black and Hispanic students. When the possibility of using an unweighted measure of enrollment size was investigated for the 2012 NYTS, results demonstrated that adequate representation of non-Hispanic black and Hispanic students would be achieved through the use of an unweighted MOS. Thus, starting with the 2013 NYTS and continuing for 2014 NYTS, student enrollment was used as the unweighted MOS, leading to improvements in the statistical efficiency of the design.

The MOS used in the 2014 NYTS sampling design no longer oversampled schools with heavier minority concentrations directly. In addition, the allocation to strata was proportional so the second oversampling approach was also no longer in effect. Nevertheless, double class selection was still implemented in the 2014 NYTS sampling.

In previous NYTS cycles, schools with high racial/ethnic populations were subject to double class selection. More specifically, two classes per grade were selected in these schools, compared to one class per grade in other schools, to increase the number of racial/ethnic minority students sampled. In the 2014 NYTS, double class selection was used only in large schools that had greater than 3% non-Hispanic black student enrollment. The threshold was developed, and updated, to generate the necessary numbers of participating non-Hispanic black students to ensure estimation precision for this subgroup.

2.1.2 Frame Construction

The frame was constructed from separate sources obtained from the National Center for Education Statistics (NCES) and from a commercial vendor, Market Data Retrieval Inc. (MDR Inc). The NCES files were the Common Core of Data (CCD) for public schools and Private School Survey (PSS) for private schools. In addition, the frame incorporated data from the MDR dataset.

The reason for moving to a frame built from multiple data sources is to increase the coverage of schools nationally. This dual-source frame build method was implemented for the 2014 NYTS

[3] Errecart, M. T. (1990, October 5). *Issues in Sampling African-Americans and Hispanics in School-Based Surveys*. Atlanta, GA: Centers for Disease Control.

survey for the first time[4]. Including schools sourced from the two NCES files resulted in a coverage increase among all public and non-public high schools of 23%. There was a 15.5% increase of coverage among public schools and a 46% increase in coverage among non-public high schools.

2.1.3 Sampling Stages and Measure of Size

The three-stage cluster sample was stratified by racial/ethnic composition and urban versus nonurban status at the first (primary) stage. PSUs were defined as a county, a group of smaller counties, or a portion of a very large county. PSUs were classified as "urban" if they are in one of the 54 largest metropolitan statistical areas (MSAs) in the U.S using 2012 American Community Survey (ACS) data from the US Census Bureau. Otherwise, they were classified as "non-urban." The 16 primary strata are defined in Exhibit 2-2. Additionally, implicit stratification was imposed by geography by sorting the PSU frame by State and by 5-digit ZIP Code (within State). Within each stratum, a PSU was randomly sampled without replacement at the first stage.

In subsequent sampling stages, a probabilistic selection of schools and students was made from the sample PSUs. It may be helpful to stress one feature of the 2014 NYTS sample which, as in the 2012 NYTS, was designed to balance the student yields for the two school levels—middle schools and high schools. As high schools have more grades, 4 grades versus 3 grades for middle schools, and therefore more students, fewer sample schools were selected at the high school level than at the middle school level. This was implemented by subsampling PSUs for high schools—for large schools as well as for medium and small schools. For medium and small schools, we subsampled fewer PSUs for high school than for middle school sampling.

The sampling stages may be summarized as follows, with additional details provided in Section 2.2:

- *Selection of PSUs*: Ninety-three 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 second-stage units (SSUs) were selected from the sample PSUs. While one large middle school was selected from each of the 93 sample PSUs, one large high school was selected from each of 77 subsample PSUs (170 = 93 + 77). An additional 20 medium schools and 30 small schools were selected from subsample PSUs, for a total of 220 sample SSUs (220= 170+30+20). The PSU subsample was drawn as a simple random sample, 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 by default chosen for participation. Classes were selected from course schedules provided by each school that agreed to participate. Schedules were constructed such that all eligible students were represented one time only.

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

Schools were stratified into large, medium and small schools based on their ability to support two, one, or less than one class selection per grade. In large schools, an average of 1.46 classes were selected per grade by selecting two classes per grade in 46% of selected large schools and one class per grade in the remaining schools. The double class sampling took place in schools with non-Hispanic black enrollments over the established 3% threshold. Specifically, double class sampling took place by design in 78 of the 170 sample large schools, or 45.8% of these schools.

The sampling approach utilized PPS sampling methods. In PPS sampling, when the MOS is defined as the count of final-stage sampling units, and a fixed number of units are selected in the final stage, the result is an equal probability of selection for all members of the universe. For the NYTS, we approximate these conditions, and thus obtain a roughly self-weighting sample.

The MOS also was used to compute stratum sizes and PSU sizes. Assigning an aggregate measure of size to PSU, the sample allocates the PSU sample in proportion to the student population. Exhibit 2-1 presents a high-level summary of the key sampling design features that will be described in detail in the next sections.

EXHIBIT 2-1: KEY SAMPLING DESIGN FEATURES

Sampling Stage	Sampling Units	Sample Size (Approximate)	Stratification	Measure of Size
1	PSUs (Counties or groups of counties)	93 Counties or groups of counties	Urban vs. nonurban (2 strata); Minority concentration (8 strata)	Aggregate school size in target grades
2	SSUs (Schools)	220 SSU (school) selections: 170 large schools, 20 medium schools and 30 small schools	Small, medium and large; High school vs. middle school	Eligible enrollment
3	Classes/ students	1 or 2 classes per grade (2 per grade in large, high-minority schools)		

2.2 Stratification and Linking

This section describes the following steps that are necessary for the selection of the first- and second-stage samples of PSUs and schools: organizing counties into Primary Sampling Units; linking schools into SSUs; and the stratification and allocation methods at each of these stages.

2.2.1 Primary Sampling Unit (PSU)

Defining a PSU

In general, PSUs are geographic areas defined as counties or groupings of counties. In defining a PSU, several issues are considered:

1. Each PSU should be large enough to contain the requisite numbers of schools and students by grade, yet not so large as to be selected with near certainty.
2. Each PSU should be compact geographically so that field staff can go from school to school easily.
3. Recent data should be available to characterize each PSU.
4. Each PSU should contain at least four middle and five high schools.

Generally, counties were equivalent to PSUs with two exceptions: 1) low population counties are combined to provide sufficient numbers of schools and students; and 2) counties that are very large may be split to avoid becoming certainty or near-certainty PSUs. Certainty PSUs are those whose size is large enough to ensure selection with probability one (1.0) with a PPS sampling design that selects larger PSUs with larger probabilities. As certainty PSUs lead to inefficiencies in the design, they are split so that the new smaller units are no longer selected with a probability of one. Near-certainty units are also split to build in a safety buffer in the PSU sizes. County population figures were aggregated from school enrollment data for the grades of interest.

The 2014 NYTS PSU definitions were based on the definitions developed in the coordinated 2011 and 2013 YRBS-NYTS cycles, and also used in the 2012 NYTS cycle. The exact PSUs defined in 2014 NYTS sampling frame were updated to ensure that all PSU met the criteria above. The frame had 1,268 PSUs, 529 of which were comprised of one single county.

Stratification of PSUs

The PSUs were organized into 16 strata, based on urban/nonurban location (as defined above) and racial/ethnic minority enrollment of non-Hispanic blacks and Hispanics. In the traditional stratification used by the NYTS the classification of PSUs into the two racial/ethnic minority strata, non-Hispanic black and Hispanic, is based on the predominant minority in the PSU. This classification is coupled with the density distribution of non-Hispanic blacks and Hispanics to subdivide each of the four primary strata into four substrata, indexed by 1-4 according to this density. The determination of the substratum boundaries is described below.

The approach involves the computation of optimum stratum boundaries using the cumulative square root of "f" method developed by Dalenius and Hodges.[5] The boundaries or cutoffs change as the frequency distribution ("f") for the racial groupings change from one survey cycle to the next. These rules are summarized below.

- If the PSU is within one of the 54 largest MSA in the U.S. it is classified as "urban," otherwise it is classified as "nonurban."
- If the percentage of Hispanic students in the PSU exceeded the percentage of non-Hispanic black students, then the PSU is classified as Hispanic. Otherwise it is classified as non-Hispanic black.
- Hispanic urban and Hispanic nonurban PSUs were classified into four density groupings, depending upon the percentages of Hispanics in the PSU.
    - For urban, High Hispanic PSU, the percentage cut points used to define the groups were 24, 40, and 60%.
    - For nonurban, High Hispanic PSU, the percentage cut points used to define the groups were 24, 50, and 68%.
- Non-Hispanic Black urban and non-Hispanic black nonurban PSUs also were classified into four groupings, depending upon the percentages of non-Hispanic blacks in the PSU.
    - For urban non-Hispanic black PSUs, the percentage cut points used to define the groups were 26, 36, and 54%.
    - For nonurban High non-Hispanic Black PSUs, the percentage cut points used to define the groups were 20, 36, and 58%.

[5] Dalenius, T., & Hodges, J. L. (1959). Minimum Variance Stratification. *Journal of American Statistical Association,* 54, 88−101.

Allocation of the PSU Sample

We designed and selected a sample of 93 PSUs that were allocated in proportion to student enrollment. Using simulations as in previous studies, we then made adjustments to the initial allocation to meet racial/ethnic minority targets. Specifically, the adjustments rounded fractional allocations, ensured that each stratum would have at least two sampled PSUs, and added balance to the distribution across strata.

Exhibit 2-2 presents the allocation of the PSU sample to strata. Compared to previous cycles, this allocation is closer to proportional, and therefore, more efficient statistically (i.e., it leads to smaller variances and tighter confidence intervals).

EXHIBIT 2-2: STRATUM DEFINITION AND PSU ALLOCATION TO STRATA

Predominant Minority	Urban/Nonurban	Density Group Number	Stratum Code	Student Population	Number of Sample PSUs
Non-Hispanic Black	**Urban**	1	BU1	2,720,181	9
		2	BU2	975,490	3
		3	BU3	908,299	3
		4	BU4	516,712	2
	Nonurban	1	BR1	3,937,157	12
		2	BR2	1,503,403	5
		3	BR3	1,026,612	4
		4	BR4	313,063	2
Hispanic	**Urban**	1	HU1	3,530,556	11
		2	HU2	2,429,442	7
		3	HU3	1,865,988	5
		4	HU4	2,106,242	7
	Nonurban	1	HR1	4,427,215	14
		2	HR2	1,284,402	4
		3	HR3	988,655	3
		4	HR4	523,491	2

2.2.2 Schools

Linking into Second-stage Sampling Unit (SSU)

FIGURE 2-1: LINKED SCHOOL CONSTRUCTION AND GRADE SAMPLING FOR HIGH SCHOOLS

Component School B (Whole)	Component School A (Fragment)
Grade 9	
Grade 10	Grade 10
Grade 11	Grade 11
Grade 12	Grade 12

Schools were classified as "whole" for high schools if they have all high-school grades 9 through 12, and whole for middle schools if they had all grades 6–8. Otherwise, they were considered a "fragment" school. Fragment schools formed component schools that were linked with other schools (fragment or whole) to form a linked school that has all four grades. This process is illustrated in Figure 2-1, where "Component School A" is linked with "Component School B," to form a linked school. We linked schools before sampling using an algorithm developed for use in the national YRBS that links geographically proximate schools. Linked schools were treated as second-stage sampling units (SSUs) with selection performed at the grade level, as described below.

Stratification

SSUs were stratified by school level (middle and high) and by size. Middle schools were those that contained any of grades 6 through 8, and high schools were those that contained any of grades 9 through 12. Schools that contained a mix of high and middle school grades were split into two sampling units, or one for each level.

SSUs also were stratified by school size into large, medium, and small strata on the basis of their ability to support less than one, one, or two class selections per grade. Operationally, large SSUs contained at least 56 students at each grade level, medium SSUs contained between 28 and 55 students per grade, and small SSUs contained less than 28 students at any grade level.

2.2.3 Sample Sizes

The original specifications for NYTS sample sizes were not given in terms of student yields; rather, they were specified in terms of the precision of the resulting estimates. Thus, the NYTS was designed to produce prevalence estimates for tobacco products with a margin of error (MOE) of 5% (± 5%) at a 95% level of precision. Estimates by grade, sex, and grade by sex meet this standard. The same standard is used for the estimates for racial/ethnic groups by school level.

The NYTS is designed to produce accurate estimation within a MOE of 5% at a 95% precision level for the following key subgroup estimates:

- *Middle and high school (school level):* middle school students in total (grades 6–8 combined) and high school students in total (grades 9–12 combined)
- *Grade*: individual grades 6, 7, 8, 9, 10, 11, and 12
- *Sex*: males and females in total, by school level (male middle school students, female high school students), and by individual grade (6-grade males, 6-grade females)
- *Race/Ethnicity*: in total and by school level (e.g., Hispanic middle school students)

Over the past several cycles of NYTS, we have confirmed that sample sizes, and resulting student yields, were sufficient to achieve design goals in terms of precision. The 2014 NYTS design meets the precision targets, while adjusting sampling parameters to reflect changing demographics of the in-school population of middle and high school students. Following derivations similar to previous cycles, these precision requirements necessitate a minimum sample size of 10,000 participating students per school level. This minimum sample size ensures that estimates by race/ethnicity meet the required precision levels for each school level. The target sample sizes correspond to approximately 3,000 participating students per grade—more for middle-school than high-school grades—and so they also ensure the precision of estimates by individual grade (e.g., sex by grade subgroup estimates on the basis of about 1,500 students).

Recall that the 2014 NYTS sampling design aimed at balancing student yields by school level (middle and high school) by subsampling PSUs for the high-school sample. The premise is that each high school provides student samples for four grades while each middle school provides student samples for three grades.

Across the nine previous cycles of the NYTS, the school participation has averaged 86.5%, with a low of 75.4%. Student participation has averaged 90.5% with a low of 87.6%. Historical participation rates at both school and student levels are summarized in Exhibit 2.3.

In calculating the sample sizes for the 2014 NYTS, we made our approach more robust by assuming a conservative combined rate of 77%, which was slightly lower than the historical overall response rate of 78.3%. . This combined participation rate could arise, for example, from student participation rates of 90% and 85.5%, respectively.

EXHIBIT 2-3: HISTORICAL SUMMARY OF NYTS PARTICIPATION RATES

YEAR	School Participation	Student Participation	Overall
1999	90.3%	93.2%	84.2%
2000	90.0%	93.4%	84.1%
2002	83.1%	90.6%	75.3%
2004	92.7%	87.9%	81.5%
2006	91.6%	87.6%	80.2%
2009	92.3%	91.9%	84.8%
2011	83.2%	88.0%	73.2%
2012	80.3%	91.7%	73.6%
2013	75.4%	90.7%	68.4%
Average over all cylces	86.5%	90.5%	78.3%

Next, we develop sampling parameters that will lead to a total projected sample size in excess of 21,000 participants, and more than 10,000 participants per level. The total sample size developed in Exhibit 2-4 is also translated into a total of 27,374 students selected from all sample schools when the calculations account for the response rates at the school and student levels discussed below.

The architecture described in Exhibit 2-4 balances the target sample sizes needed for the two key domains defined by middle school and high school levels. This balance is achieved by considering different target sample sizes per grade needed for high school and middle school grades, as typically, middle schools supply three grades and high schools supply four grades. That is the case in every SSU selected for the two levels, as each SSU offers a complete set of grades for the level (i.e., every SSU selected for the MS level supplies three grades and every SSU selected for the HS level supplies four grades).

Schools were further classified by size based on grade-level enrollments; the definition of size strata is provided in Section 2.2.2. This ensures that a sampled school of a given size classification is able to support the student sample sizes given in Exhibit 2-4.

The NYTS sample size calculations are based on the following assumptions:

- The main structure of the sampling design will be consistent with the design used to draw the sample for prior cycles of the NYTS.
- The selection of a minimum of one SSU at the high school level and one SSU at the middle school level within each PSU. Some PSUs are selected to provide up to four extra schools. A PSU is a county, a group of contiguous counties, or a section of a county if too large.
- SSUs with at least 56 students per grade are considered large, and those among the others with 28 students per grade are considered medium; otherwise, they are considered small.
- On average, each selected class includes 28 students (on the basis of historical averages).
- For SSUs classified as large, we sample double the amount of students in 46% of these schools, by sampling eight classes instead of four.
- A 77% overall response rate (based on historical averages) calculated as the product of the school and student response rate.

Based on these assumptions, 93 PSUs were selected as the basis for the sample. Next, we describe selection of schools or SSUs, recalling the definition of SSUs as "linked schools" created by combining actual schools so that each linked school unit has a complete set of grades for the level.

The linking of actual schools into SSUs means that the selected school sample size, 220 SSUs, corresponds to a larger number of schools, projected to fall between 235 and 245 actual schools using an average number of schools/SSU estimated from previous cycles.

Large middle school SSUs were selected from the 93 sample PSUs, one middle school per PSU for a total of 93 large middle schools. For the selection of large high school SSUs, 77 PSUs were subsampled from the 93 sample PSUs, with 1 large high school selected from each of the 77 subsample PSUs. Therefore, a total of 170 large schools or SSUs were selected into the sample.

Another subset of PSUs was independently subsampled to supply medium schools as follows. A subsample of 12 PSUs were selected for medium middle schools and a subsample of 8 PSUs were selected for medium high schools. For small schools, in addition, 13 PSUs were subsampled to supply high schools and 17 PSUs were subsampled to supply middle schools. In each subsample PSU, we selected one school of the specified level (MS, HS) and size (medium or small).

Exhibit 2-4 summarizes the designed sample sizes for each school type; i.e., the number of schools that were specified to be drawn along with the anticipated number of participating schools and students. Section 3.4 compares these projections to the actual sample yields.

EXHIBIT 2-4: PLANNED SAMPLE SIZES FOR THE 2014 NYTS

School Level	Size		Number of SSUs	Sample Size Calculations				Yields @ 77% Overall Response Rate
				Number of Schools Sampled	Number of Classes Per School	Number of Students Per Class	Sample Students Prior to Attrition	
Middle	Large		93	Double Classes: 43	6	28	7,224	5,562
				Single Classes: 50	3	28	4,200	3,234
	Total			**93**			**11,424**	**8,796**
	Medium		12	12	3	28	1,008	776
	Small		17	17	2.8	11.2	470	362
Total Middle			**122**	**122**			**12,902**	**9,935**
High	Large		77	Double Classes: 35	8	28	7,840	6,037
				Single Classes: 42	4	28	4,704	3,622
	Total			**77**			**12,544**	**9,659**
	Medium		8	8	4	28	896	690
	Small		13	13	3.8	18.1	1,032	794
Total High			**98**	**98**			**14,472**	**11,143**
Overall	Large		170	170			23,968	18,455
	Medium		20	20			1,904	1,466
	Small		30	30			1,502	1,157
Total Overall			**220**	**220**			**27,374**	**21,078**

Exhibit 2-4 also highlights the double class sampling that took place in a subset of large schools. As mentioned in Section 2.1.1, double classes were selected in those large schools with a sufficient proportion of non-Hispanic blacks (i.e., at least 3% black).

These schools were expected to yield approximately 21,078 students. The large projected sample size permits analysis by individual grade and by sex without any special considerations in the sampling plan. Design effects were assumed by the design to be relatively small for subgroups that cut across schools; therefore, sex group estimates will have better precision than other groups. Thus, the designed confidence intervals were ± 3%.

The next paragraphs discuss how the design was balanced to achieve precise estimates for subgroups defined by school level, grade, sex and race/ethnicity.

2.2.3.1 Middle School and High School Estimates

Estimates by school level are required to support separate analysis of students across middle school grades (6, 7, and 8) and high school grades (9, 10, 11, and 12). However, schools tend to vary in their grade structures, an inconsistency that compromises the ability to easily and efficiently link schools for sampling purposes in a manner that also uniformly divides students by grade. For example, 9th-grade students are served by both junior high schools with grades 7–9 and by high schools with grades 9–12. As a result, we have developed the school linking approach described in Section 2.2.2 that was applied independently for high schools and middle schools.

2.2.3.2 Grade Estimates

The designed sample sizes are approximately balanced for school-level and for grade-level groupings. By targeting nearly 3,000 students per grade, the sample ensures that estimates at the grade level achieve the required precision levels.

2.2.3.3 Sex Group Estimates

The large sample size permitted analysis by sex without any special considerations in the sampling plan. During the class selection process, frames of eligible classes from co-educational schools in which classrooms were segregated by sex (i.e., an all-male or all-female class) were avoided, if possible.

2.2.3.4 Race/Ethnicity Group Estimates

In order to support separate analysis of the data for non-Hispanic white, non-Hispanic black and Hispanic students, in total and by school level, adequate sample sizes were required by the design for subgroups defined by: 1) school level by racial grouping; or 2) by sex grouping. Sample sizes were not designed, however, to support detailed analyses by sex and school level within racial/ethnic subgroups (e.g., middle school Hispanic males).

2.3 Sampling Methods

This section describes the methods used in the selection of PSUs, schools, grades, and classes of students. In this process, we define the probabilities of selection associated with the various sampling stages as follows:

- Probability of selecting PSUs
- Probability of selecting schools
- Probability of selection of grades
- Probability of selecting classes and students

These probabilities provide the basis for the sampling weights discussed in Chapter 4.

The overall probability of selection for a student is the product of the probability of selection of the PSU, which contains a group of schools, multiplied by the conditional probability of selecting

the student's school and the conditional probability of selecting the student's class. These steps are detailed in the selection below.

2.3.1 Primary Sampling Unit

Selection

Within each first-stage stratum, the PSUs were sorted by five-digit ZIP Code to attain a form of implicit geographic stratification. Implicit stratification, coupled with the PPS sampling method described below, ensures geographic sample representation. With PPS sampling, the selection probability for each PSU is proportional to the PSU's measure of size.

The following systematic sampling procedures were applied to the stratified frame to select a PPS sample of PSUs.

- Select 93 PSUs with a systematic random sampling method within each stratum. The method applies within each stratum 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.
- Subsample at random 77 of the sample PSUs for the large high school sample. (Recall that there is no subsampling of large middle schools.)
- Subsample at random 12 of the sample PSUs for the medium school sample for the middle school sample and 8 of the sample PSUs for the medium high school sample.
- Subsample at random 13 of the sample PSUs for the small high school sample and 17 PSUs for the small middle school sample.

Probability

If MOS_{klm} is the measure of size for school *k* in PSU l in stratum *m* and if K_m is the number of PSUs to be selected in stratum *m*, then P^p_{lm} is the probability of selection of PSU *l* in stratum *m*:

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

2.3.2 Schools

Selection

For large schools, one high school and one middle school were selected with PPS systematic sampling within a PSU. The schools were selected into the sample with probability proportional to the measure of size. (MOS).

Small and medium schools were sampled independently from large schools; they were set in two separate strata sampled at lower rates. This approach was implemented by drawing subsample PSUs for small and medium school sampling as described earlier. One small school or medium school was then selected in each subsampled PSU with probability proportional to the MOS.

Replacement of Schools/School Systems

We did not replace refusing school districts, schools, classes, or students. We allowed for school and student nonresponse by inflating the sample sizes to account for nonresponse. With this approach, all schools can be contacted in a coordinated recruitment effort, which is not possible for methods that allow for replacing schools.

Probability

The probability of selecting large school *k* in PSU l and stratum *m*, P^{LS}_{klm}, at each level was computed as follows:

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

For large high schools, the selection probability incorporates an additional factor, 77/93, associated with the subsampling of PSUs for these schools.

For medium schools, one school was drawn from each subsampled PSU at each level, so the probability of selection of a medium school then becomes,

$$P^{MS}{}_{klm} = (12/93)\left(\frac{MOS_{klm}}{MOS_{.lm}} \right)$$

for medium middle schools and

$$P^{HS}{}_{klm} = (8/93)\left(\frac{MOS_{klm}}{MOS_{.lm}} \right)$$

for medium high schools.

For small schools, one school was drawn from subsampled PSU at each level, so the probability of selection of a small school for each level then becomes

$$P^{MS}{}_{klm} = (17/93)\left(\frac{MOS_{klm}}{MOS_{.lm}}\right)$$

for small middle schools, and

$$P^{HS}{}_{klm} = (13/93)\left(\frac{MOS_{klm}}{MOS_{.lm}}\right)$$

for small high schools.

2.3.3 Grades

Selection

Except for linked schools, all eligible grades were included in the class selection for each school.

In linked schools, grades were selected independently. One component school was selected to provide classes at each grade level, and grades within component schools were drawn with probability proportional to grade enrollment.

Probability

Most SSUs in the sample contained one component school. In these cases, all eligible grades were selected so that the probability of selecting a grade was 1.0.

In SSUs that were made up of component schools, the selection of each component school at each grade is made with PPS sampling. The school selections from each component school at each grade level were made independently.

We denote this $P^{G}{}_{jklm}$ the probability of selecting grade *j* in SSU *k*, in PSU *l*, stratum *m*. For the j^{th} grade within SSU *k*, this probability is equal to the ratio of the number of students at grade *j* in the component school to the total enrollment in grade *j* across all component schools within the SSU.

2.3.4 Classes

Selection

In large schools, an average of 1.46 classes per grade were selected by selecting 2 classes per grade in 46% of the selected large schools and one class per grade in the remaining large schools. The double class sampling took place in schools with greater than 3% non-Hispanic black enrollment and one class per grade in the remaining schools.

One class per grade was selected in medium schools.

In small schools, that is, those that could not support a full class selection at each grade, all students in all eligible grades were taken into the sample.

All students in a selected class who could complete the survey without special assistance were considered eligible and offered the opportunity to participate in the survey. Refusing students were not replaced. Nonresponse at the student level was accounted for in the sample size using an average per class yield that assumed student response rates derived from historical experience with the NYTS.

A set of classes was identified for each school at each grade level such that every student in a given grade level was enrolled in exactly one of the classes in the set. For example, a required English course might be used. If the school's estimated non-Hispanic black enrollment exceeded 3%, two classes were randomly selected, without replacement, from the list. Otherwise, one class was randomly selected. Selections were made at all eligible grade levels in the school.

Probability

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 one or two 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 (i.e., $1/C_{jklm}$ or $2/C_{jklm}$).

Note that the probability of student selection within a class does not vary by race, ethnicity, or sex. We denote this probability as P^C_{ijklm} as the probability of selecting class *i* in grade *j*, school *k*, PSU *l*, stratum *m*. Since every student in a selected class is also selected, the probability of selecting any student in class *i*, grade *j*, school *k*, PSU *l*, stratum *k*, is also equal to P^C_{ijklm}.

CHAPTER 3—NYTS DATA COLLECTION

3.1 Survey Instrument

The NYTS collects data on key short-term, intermediate, and long-term tobacco prevention and control outcome indicators. The 2014 survey instrument included a total of 81 questions, with the first 5 collecting student demographic information and the remaining measuring 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 antitobacco 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.

3.2 Recruitment Procedures

The schools selected to participate in the 2014 NYTS were located in 36 different States. Recruitment began in May 2013 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. 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 an electronic calendar on a secure shared drive to facilitate communication and to avoid scheduling two schools for the same data collector on the same day.

3.3 Survey Administration

Survey administration in the schools began on February 10, 2014, immediately after data collector training, and continued until June 13, 2014. 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) precontact 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) postsurvey meeting with the teacher or teachers; and 6) postsurvey meeting with the principal or lead contact prior to leaving the school. Most survey administrations could be completed in 1 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.

3.3.1 Data Collection Staffing

Data collectors were recruited from a pool of previously trained data collectors, as well as retired teachers associations, school health networks, and a variety of health education organizations. Data collector training was conducted on February 5–6, 2014. Key components of the training included the following:

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

3.3.2 Field Procedures

After schools had been recruited, classes selected, and a date scheduled, each school received a packet of presurvey materials. These materials included all the information necessary to prepare the school for data collection. Teacher packets contained the parental permission forms that had to be given out to all students in the selected classes prior to data collection. The timing of these presurvey 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 1 week prior to the scheduled date of data collection in the majority of schools. Active parental permission forms, forms that must be returned with the parent's signature in order for the child to participate, were sent out at least 2 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, and their must-read weekly bulletin. Weekly bulletins underlined key performance issues, corrected misconceptions, provided consistent direction on any procedural changes, and kept everyone abreast of the latest must-have 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, data 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.

3.3.3 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 such 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 (i.e., 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, school homerooms were used as the frame for class selection.

3.4 Participation Rates

Across the nine previous cycles of the NYTS, the school participation has averaged 86%, with a low of 75%. Student participation has averaged 91% with a low of 88%, and the overall response rate has averaged 78%. To be conservative, we assumed slightly lower values in developing the sample design for the 2014 NYTS: an assumed overall participation rate of 77%.

The actual response rates in 2014 differed from our projections: an actual school participation rate of 80.2% and a student participation rate of 91.4%. The overall participation rate, the product of the school-level and student-level participation rates was 73.3%. While the participation rate is slightly lower than the levels assumed in the projections, the shortfall has no meaningful impact on the estimation precision. As seen below, however, some design modifications may be recommended to increase the yields achieved for black students.

The 2014 NYTS data file contains responses from 22,007 students compared to the 21,078 responding students anticipated by the design. Exhibit 3-1 shows that student yields were lower than targeted for non-Hispanic Blacks, with a shortfall of 76 and 255 students for high and middle schools, respectively. Among Hispanics, yields far exceeded the targets at both levels.

EXHIBIT 3-1: SAMPLE YIELDS FOR NON-HISPANIC BLACK AND HISPANIC STUDENTS BY SCHOOL LEVEL

Subgroup	Projected Participants	Actual Participants
Middle School non-Hispanic Blacks	1,775	1,520
Middle School Hispanics	1,775	3,006
High School non-Hispanic Blacks	1,975	1,899
High School Hispanics	1,975	2,933

CHAPTER 4—WEIGHTING OF NYTS RESPONSE DATA

4.1 Overview

This section describes the procedures used to weight the data collected in the NYTS 2014. The process involved the steps outlined below:

- Sampling weights
- Nonresponse adjustments
- Weight trimming
- Poststratification to national estimates of racial totals by grade, sex and school type

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 poststratify the data to known sampling frame characteristics.

4.2 Sampling Weights

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

The NYTS computation of sampling weights begins at the student sampling stage, and then moves to the school and PSU sampling stages. This sequence allows the student sampling weights to incorporate adjustments for student nonresponse. These adjustments, described next, use enrollment data by sex and by grade collected for each participating school. Because the process begins with the student weights within a given grade, school and PSU, we refer to these weights as conditional weights.

4.2.1 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 (see Appendix B) 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.

We denote the student selection weight 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 is 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* is defined by a sequence of rules that depends on the number of responding students. This is done to avoid large weights for classes with low numbers of respondents. This process operates entirely within schools.

Initially, the weighting class is defined by grade and sex within each school. We then combine weighting classes if the weight for the class exceeds a maximum value, C. This cap *C* is computed using the following equation:

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

The combination sequence first groups males and females within grade. Both the cap and the weight are then recomputed. If the weight still exceeds the cap, grades are combined. The process is repeated, and if the student weight still exceeds the cap, the school is taken as the weighting class.

This has the effect, within 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 is collapsed to the school level.

4.2.2 School Sampling Weights

For large schools, the partial school weight is 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 is:

$$W^{SS}{}_{klm} = (93/15)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{SS}{}_{klm}}$$

For medium schools, the partial school weight is:

$$W^{MS}{}_{klm} = (93/12)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{MS}{}_{klm}}$$

$$W^{HS}{}_{klm} = (93/8)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{HS}{}_{klm}}$$

4.2.3 Grade Sampling Weights

Grade selection occurs within linked schools where the grade is 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, is simply the inverse of the probability of selection described in Section 2.4. In a non-linked school, the weight is 1.0. We denote the grade weight as $W^{G}{}_{jklm}$.

4.2.4 PSU Sampling Weights

The weight of the PSU is 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. (Subsampling of PSUs also was conducted for the large high schools.) This PSU subsampling component of the PSU weight is accounted for in the school selection probability and corresponding weight.

4.2.5 Overall Sampling Weight

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

4.3 Nonresponse Adjustments

This section describes how weights are adjusted for nonparticipation by entire schools, using strata as weighting classes.

The adjustment process is different in small schools than in medium and large schools, as represented by the following equations for the adjustment factor.

$$\begin{cases} A_m^{MS,LS} = \dfrac{\sum_{k,l\ \varepsilon\ \text{medium and large schools sampled}} W_{lm}^{P} W_{klm}^{MS,LS} MOS_{klm}^{MS,LS}}{\sum_{k,l\ \varepsilon\ \text{medium and large schools responding}} W_{lm}^{P} W_{klm}^{MS,LS} MOS_{klm}^{MS,LS}} \\ A^{SS} = \dfrac{\sum_{k,l\ \varepsilon\ \text{small schools sampled}} W_{lm}^{P} W_{klm}^{SS} MOS_{klm}^{SS}}{\sum_{k,l\ \varepsilon\ \text{small schools with respondents}} W_{lm}^{P} W_{klm}^{SS} MOS_{klm}^{SS}} \end{cases}$$

The first equation applies to medium and large schools combined, and the second applies to small schools. Note that this adjustment is made within stratum for medium and large schools and across the whole sample for small schools. The student weight, adjusted for nonresponse, is $A^{SS}{}_{lm}\ W^{T1}{}_{hijklm}$ for small schools and $A^{MS,LS}{}_{lm}\ W^{T1}{}_{hijklm}$ for medium and large schools.

To avoid very large weight adjustment factors, which may lead to variance increases, weighting classes combined the top two sampling strata in terms of racial/ethnic minority concentrations. These weighting cells were created for computing nonresponse adjustments only—the collapsed strata not kept on the analytic file. Specifically, weighting cells combined the following pairs of strata: BU4 and BU3; BR3 and BR4; HU3 and HU4; and HR3 and HR4. School response rates by weighting class, and the resulting nonresponse adjustment factors, are detailed in Exhibit 4-1. Note that the weighting classes are defined using collapsed sampling strata.

EXHIBIT 4-1: MEDIUM AND LARGE SCHOOL NONRESPONSE

School Level	Stratum (Nonresponse)	Sampled Schools	Responding Schools	Percent Responding	Nonresponse Adjustment
HS	BR1	13	10	76.9%	1.3
HS	BR2	5	5	100.0%	1.0
HS	BR3 and BR4	6	6	100.0%	1.0
HS	BU1	10	8	80.0%	1.2
HS	BU2	4	3	75.0%	1.2
HS	BU3 and BU4	6	3	50.0%	2.0
HS	HR1	10	9	90.0%	1.2
HS	HR2	4	3	75.0%	1.4
HS	HR3 and HR4	5	4	80.0%	1.2
HS	HU1	13	9	69.2%	1.4
HS	HU2	4	4	100.0%	1.0
HS	HU3 and HU4	12	11	91.7%	1.1
HS	Total	92	75	81.5%	
MS	BR1	16	12	75.0%	1.3
MS	BR2	6	6	100.0%	1.0
MS	BR3 and BR4	9	6	66.7%	1.6
MS	BU1	16	11	68.8%	1.3
MS	BU2	3	2	66.7%	1.4
MS	BU3 and BU4	6	6	100.0%	1.0
MS	HR1	21	18	85.7%	1.1
MS	HR2	5	3	60.0%	1.5
MS	HR3 and HR4	6	6	100.0%	1.0
MS	HU1	19	13	68.4%	1.4
MS	HU2	13	10	76.9%	1.3
MS	HU3 and HU4	16	15	93.8%	1.1
MS	Total	136	108	79.4%	
	Gross Total	228	183	80.0%	

4.4 Weight Trimming

Extreme variation in sampling weights can cause inflated sampling variances, and offset the precision gained from a well-designed sampling plan. One strategy to compensate for this is to trim extreme weights and distribute the trimmed weight among the untrimmed weights. The method we used[6] is based on a similar procedure employed for the National Assessment of Educational Progress (NAEP).

The trimming is an iterative procedure. During each iteration, an optimal weight (W_o)[7], is calculated from the sum of the squared weights in the sample. Each weight W_i is then marked and trimmed if it exceeds that optimal weight. The trimmed weight is summed within grade and spread out proportionally over the unmarked cases in the grade. This process is repeated for 20 iterations or until no weight is being trimmed.

W_{ok} is determined by the following:

$$W_{ok} = \left(c \sum_{k=1}^{n} \frac{w_k^2}{n} \right)^{\frac{1}{2}}$$

The constant, "c," is arbitrary. Setting it to a low level will generate high levels of trimming, while increasing it will reduce the level of trimming. For the current study, "c" has been set so that approximately 5% of the weight is trimmed in the first iteration of the trimming algorithm.

Let W_{ik} and W_{ok} be the weight for the *i*th case and the optimum weight for the *k*th iteration, respectively, and define t_{ik} as 1 if W_{ik} is greater than or equal to W_{ok}, and 0 otherwise.

Then the trimmed weight for the *k + 1* iteration is defined as follows:

$$W_{i,k+1} = \begin{cases} W_{ok} & \text{if } W_{ik} \geq W_{ok} \\ \dfrac{\sum_{i=1}^{n} W_{ik} \left(1 - \dfrac{t_{ik} \times W_{ok}}{W_{ik}} \right)}{\sum_{i=1}^{n} W_{ik} (1 - t_{ik})} & \text{if } W_{ik} < W_{ok} \end{cases}$$

Trimming has the effect of reducing the variation of the weights substantially in each trimming class and overall.

---

[6] Potter, F. (1988). Survey of Procedures to Control Extreme Sampling Weights. *American Statistical Association 1988 Proceedings: Survey Research Methods Section*, pp. 225–230.

[7] In the following discussion, the subscripts are used to indicate the iterative process used in the trimming algorithm. To avoid overly cumbersome notation, we have omitted the subscripts indexing the sampling stages. W, the initial weight, is taken as the nonresponse adjusted sampling weight described in the preceding section. The subscripts *k* and *n* represent the number of iterations and the number of cases/weights, respectively.

4.5 Poststratification to National Student Population Estimates

To obtain accurate counts of students in schools considered eligible for the NYTS by sexgrade, and race/ethnicity for use in poststratification, we turned to two school universe surveys conducted by NCES. Raw school-level data files were downloaded and processed to mirror eligibility requirements imposed on the sampling frame.

National estimates of racial/ethnic percentages were obtained from two sources: 1) private school enrollments by grade and five racial/ethnic groups were obtained from the Private School Universe Survey (PSS); and 2) public school enrollments by grade, sex, and five racial/ethnic categories were obtained from CCD, both produced by NCES (Appendix C). These databases were combined to produce the enrollments for all schools, and to develop population percentages to use as controls in the poststratification step.

Specifically, population control totals for public school enrollments were taken from the most recent NCES CCD Public Elementary/Secondary School Universe Survey (2010–11).[8] Records for special education, vocational, and other/alternative schools were deleted prior to computing control totals. Control totals for private school enrollments were taken from the NCES PSS, School Year 2009−10 (most recent PSS data); this file also was restricted to "regular" schools.

Given a national estimate of student counts R_a and a weighted response total of P_a for poststratification adjustment class "a," the poststratification factor was the ratio of R_a to P_a. Exhibit 4-2 gives the population control totals used in poststratification adjustments alongside the sum of the weights in each poststratum cell, as well as the adjustment factors calculated as the ratio of these two totals. More specifically, the adjustments in column G in this exhibit are computed as E/F, control total for the cell divided by the weight sum in the cell.

Poststratification adjustment cells were defined by school type, grade, sex and race/ethnicity. Because estimates are typically reported separately for middle schools and high schools, the weights were adjusted separately for both subpopulations. Within the private school adjustment cells, sex was omitted, as enrollments by sex were not available for these schools. This is indicated by a "Combined" sex in Exhibit 4-2. Also within private schools, the racial/ethnic groups were collapsed to preclude small numbers of students in the adjustment classes. For the public schools, five racial/ethnic categories were used: non-Hispanic white; non-Hispanic black; Hispanic; non-Hispanic Asian/ Pacific Islander; and non-Hispanic Native American/Alaska Native (Appendix C).

[8] Common Core of Data, National Center for Education Statistics http://nces.ed.gov/ccd/. School Year 2010–11.

EXHIBIT 4-2: POSTSTRATIFICATION ADJUSTMENTS

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Poststratification Adjustment
Private	6	Combined	Combined	306,460	347,636	178	0.88156
Private	7	Combined	Combined	305,170	222,678	167	1.37046
Private	8	Combined	Combined	305,489	326,797	178	0.93480
Private	9	Combined	Combined	281,606	205,659	107	1.36929
Private	10	Combined	Combined	279,592	221,519	117	1.26216
Private	11	Combined	Combined	275,027	187,267	102	1.46864
Private	12	Combined	Combined	271,152	162,703	106	1.66654
Public	6	Non-Hispanic Asian and Pacific Islander	Female	83,993	103,811	103	0.80910
Public	6	Non-Hispanic Black	Female	293,043	239,227	276	1.22496
Public	6	Hispanic	Female	437,910	490,908	494	0.89204
Public	6	Non-Hispanic Native American	Female	22,815	91,203	71	0.25015
Public	6	Non-Hispanic White	Female	961,220	749,440	656	1.28258
Public	6	Non-Hispanic Asian and Pacific Islander	Male	86,306	87,415	98	0.98732
Public	6	Non-Hispanic Black	Male	305,483	269,934	300	1.13170
Public	6	Hispanic	Male	456,836	572,043	561	0.79860
Public	6	Non-Hispanic Native American	Male	23,708	113,762	95	0.20840
Public	6	Non-Hispanic White	Male	1,021,460	748,834	652	1.36407
Public	7	Non-Hispanic Asian and Pacific Islander	Female	82,961	94,515	67	0.87775
Public	7	Non-Hispanic Black	Female	292,644	260,596	255	1.12298
Public	7	Hispanic	Female	430,449	584,601	528	0.73631
Public	7	Non-Hispanic Native American	Female	22,838	98,331	66	0.23226
Public	7	Non-Hispanic White	Female	968,116	865,517	704	1.11854
Public	7	Non-Hispanic Asian and Pacific Islander	Male	86,022	83,863	85	1.02575
Public	7	Non-Hispanic Black	Male	303,320	291,794	280	1.03950
Public	7	Hispanic	Male	449,797	657,112	555	0.68451
Public	7	Non-Hispanic Native American	Male	23,776	80,989	69	0.29357
Public	7	Non-Hispanic White	Male	1,024,015	937,254	741	1.09257

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Poststratification Adjustment
Public	8	Non-Hispanic Asian and Pacific Islander	Female	85,580	95,227	83	0.89870
Public	8	Non-Hispanic Black	Female	288,628	295,210	277	0.97770
Public	8	Hispanic	Female	422,768	639,572	554	0.66102
Public	8	Non-Hispanic Native American	Female	22,312	49,971	47	0.44651
Public	8	Non-Hispanic White	Female	968,394	869,283	686	1.11402
Public	8	Non-Hispanic Asian and Pacific Islander	Male	89,334	90,511	79	0.98700
Public	8	Non-Hispanic Black	Male	295,815	303,686	286	0.97408

EXHIBIT 4-2: POSTSTRATIFICATION ADJUSTMENTS (CONTINUED)

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Poststratification Adjustment
Public	8	Hispanic	Male	437,034	597,939	506	0.73090
Public	8	Non-Hispanic Native American	Male	22,800	86,784	67	0.26272
Public	8	Non-Hispanic White	Male	1,022,119	906,212	718	1.12790
Public	9	Non-Hispanic Asian and Pacific Islander	Female	87,415	151,401	91	0.57738
Public	9	Non-Hispanic Black	Female	321,224	378,173	237	0.84941
Public	9	Hispanic	Female	441,642	636,775	388	0.69356
Public	9	Non-Hispanic Native American	Female	23,819	71,285	43	0.33413
Public	9	Non-Hispanic White	Female	1,015,797	954,198	596	1.06455
Public	9	Non-Hispanic Asian and Pacific Islander	Male	91,952	142,060	90	0.64728
Public	9	Non-Hispanic Black	Male	344,666	399,004	228	0.86382
Public	9	Hispanic	Male	472,389	727,924	414	0.64895
Public	9	Non-Hispanic Native American	Male	25,536	67,946	36	0.37583
Public	9	Non-Hispanic White	Male	1,082,621	1,079,589	691	1.00281
Public	10	Asian and Pacific Islander	Female	87,245	137,173	98	0.63602
Public	10	Non-Hispanic Black	Female	298,317	437,994	266	0.68110
Public	10	Hispanic	Female	399,950	617,521	421	0.64767
Public	10	Non-Hispanic Native American	Female	22,167	67,441	32	0.32869
Public	10	Non-Hispanic White	Female	996,754	907,303	602	1.09859
Public	10	Non-Hispanic Asian and Pacific Islander	Male	91,740	144,639	96	0.63427
Public	10	Non-Hispanic Black	Male	302,961	327,576	230	0.92486
Public	10	Hispanic	Male	414,302	583,150	408	0.71045
Public	10	Non-Hispanic Native American	Male	23,140	41,956	28	0.55152
Public	10	Non-Hispanic White	Male	1,042,852	1,103,116	701	0.94537
Public	11	Non-Hispanic Asian and Pacific Islander	Female	83,060	144,618	100	0.57434
Public	11	Non-Hispanic Black	Female	270,278	296,051	248	0.91294
Public	11	Hispanic	Female	351,600	517,164	387	0.67986

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Poststratification Adjustment
Public	11	Non-Hispanic Native American	Female	20,462	25,316	18	0.80826
Public	11	Non-Hispanic White	Female	955,058	845,960	579	1.12896
Public	11	Non-Hispanic Asian and Pacific Islander	Male	87,816	125,553	106	0.69943
Public	11	Non-Hispanic Black	Male	260,589	315,934	249	0.82482
Public	11	Hispanic	Male	354,460	475,928	371	0.74477
Public	11	Non-Hispanic Native American	Male	20,769	20,236	17	1.02633
Public	11	Non-Hispanic White	Male	988,539	923,478	672	1.07045
Public	12	Non-Hispanic Asian and Pacific Islander	Female	80,949	137,666	92	0.58801

EXHIBIT 4-2: POSTSTRATIFICATION ADJUSTMENTS (CONTINUED)

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Poststratification Adjustment
Public	12	Non-Hispanic Black	Female	259,801	296,168	245	0.87721
Public	12	Hispanic	Female	324,717	503,785	361	0.64455
Public	12	Non-Hispanic Native American	Female	19,339	35,662	23	0.54227
Public	12	Non-Hispanic White	Female	941,752	793,320	554	1.18710
Public	12	Non-Hispanic Asian and Pacific Islander	Male	84,706	141,448	110	0.59885
Public	12	Non-Hispanic Black	Male	240,128	275,861	239	0.87047
Public	12	Hispanic	Male	318,489	462,472	363	0.68867
Public	12	Non-Hispanic Native American	Male	19,505	34,231	22	0.56980
Public	12	Non-Hispanic White	Male	969,959	895,150	641	1.08357

Following poststratification, the adjusted weights sum to the population control totals.

4.6 Analysis Strata and Variance Estimation

Sampling variances for complex sampling designs can be estimated using one of several methods, including linearized estimators and balanced repeated replication. These methods are implemented with a variety of software packages, including SUDAAN, WesVar, Stata and SAS using special sample survey procedures (such as Proc SurveyMeans in SAS Version 9). The 2014 NYTS data were prepared for estimating variances using the method of linearized estimators.

Because estimates are typically reported separately for middle schools and high schools, analysis strata need to ensure that each stratum has two or more PSUs for variance estimation within each subpopulation (middle schools and high schools separately).

As noted earlier, the allocation ensured that every stratum had at least two PSUs in the sample. This does not necessarily translate to two PSUs with valid student data for each school level (middle schools and high schools) in every stratum due to the effects of nonresponse at the school level. In particular, nonparticipating schools may lead to PSUs without student data for a given school level..

Exhibit 4-3 displays the correspondence between the sampling strata and the analysis strata, which are represented by two variables on the analysis file. All strata/level combinations but one had at least two PSUs. As a result, strata BU3 and BU4 were collapsed (into analysis stratum 113) because BU4 had only one PSU for one school level. Thus, the analytic file contains 15 values in the analysis strata variable and 16 values in the design strata variable.

In addition, stratum codes used in sampling and weighting were converted to a numeric “analysis stratum” code for use in SUDAAN, which requires numeric variables.

EXHIBIT 4-3: SAMPLING AND ANALYSIS STRATUM CODING SCHEMES

High non-Hispanic Black		High Hispanic	
Sampling Stratum Code	Analysis Stratum Code	Sampling Stratum Code	Analysis Stratum Code
BR1	101	HR1	201
BR2	102	HR2	202
BR3	103	HR3	203
BR4	103	HR4	203
BU1	111	HU1	211
BU2	112	HU2	212
BU3 and BU4	113	HU3	213
		HU4	214

Exhibit 4-4 presents selected key survey estimates and their sampling errors estimated using Taylor series linearization methods, which are usually employed by NYTS data analysts, and implemented with SUDAAN or similar software (e.g., SAS Proc SurveyMeans). Specifically, the Exhibit presents the percent and standard error of the percent for estimates of current use of selected tobacco products separately for high schools (Exhibit 4-4a) and middle schools (Exhibit 4-4b).

EXHIBIT 4-4A: CURRENT USE ESTIMATES FOR SELECTED TOBACCO PRODUCTS FOR HIGH SCHOOL STUDENTS (STANDARD ERRORS IN PARENTHESES)

Current Use

Product	Overall	Female	Male	Non-Hispanic White	Non-Hispanic Black	Hispanic
Cigarettes	9.2% (0.6)	7.9% (0.6)	10.6% (0.9)	10.8% (0.8)	4.5% (0.6)	8.8% (0.9)
Cigar	8.2% (0.5)	5.5% (0.5)	10.8% (0.7)	8.3% (0.6)	8.8% (1.2)	8.0% (0.8)
Smokeless Tobacco	5.5% (0.5)	1.2% (0.2)	9.9% (1.0)	7.8% (0.8)	1.1% (0.3)	3.1% (0.5)
Hookah	9.4% (0.6)	9.8% (0.8)	8.9% (0.7)	9.4% (0.7)	5.6% (0.7)	13.0% (1.4)
Electronic Cigarettes	13.4% (1.2)	11.9% (1.2)	15.0% (1.4)	15.3% (1.6)	5.6% (1.2)	15.3% (1.9)

EXHIBIT 4-4B: CURRENT USE ESTIMATES FOR SELECTED TOBACCO PRODUCTS FOR MIDDLE SCHOOL STUDENTS (STANDARD ERRORS IN PARENTHESES)

Current Use

Product	Overall	Female	Male	Non-Hispanic White	Non-Hispanic Black	Hispanic
Cigarettes	2.5% (0.2)	2.0% (0.3)	3.0% (0.4)	2.2% (0.3)	1.7% (0.4)	3.7% (0.6)
Cigar	1.9% (0.2)	1.4% (0.3)	2.4% (0.4)	1.4% (0.4)	2.0% (0.4)	2.9% (0.4)
Smokeless Tobacco	1.6% (0.2)	1.1% (0.3)	2.1% (0.4)	1.7% (0.4)	.9% (0.3)	1.3% (.3)
Hookah	2.5% (0.3)	2.6% (0.4)	2.4% (0.3)	1.4% (0.2)	1.5% (0.5)	5.6% (0.7)
Electronic Cigarettes	3.9% (0.5)	3.3% (0.5)	4.5% (0.6)	3.1% (0.5)	3.8% (0.7)	6.2% (0.8)

Example specifications for applying the method with both SAS and SUDAAN are provided below for computing prevalence.

Example: Estimates, Current Use by School Type

SAS:

SAS:

```
Proc Surveymeans Data=nyts2014 mean;
Var ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Class ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Stratum stratum2;
Cluster psu2;
Weight wt;
Domain Schooltype Schooltype*Sex Schooltype*Race_S;
Title “NYTS 2014, 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=nyts2014 Filetype= SAS Design=WR;
Var ccigt_r ccigar_r cslt_r chookah_r celcigt_r;
Catlevel 1   1    1    1    1;
Nest Stratum2 PSU2 / Missunit;
Weight wt;
Subgroup School Sex  Race_S;
Levels  2    2    3;
Tables School School*Sex School*Race_S;
Title “NYTS 2014, Estimates by School Type, by School Type and Sex Cross-Classified, and by School Type and
Race Cross-Classified”;
Print Percent Sepercent / Style=NCHS;
```

Appendix A

Questionnaire

Form Approved
OMB No.: 0920-0621
Expiration Date: 01/31/2015

National Youth Tobacco Survey (NYTS) 2014 Questionnaire

This survey is about tobacco. We would like to know about you and things you do that may affect your health. Your answers will be used for programs for young people like yourself.

DO NOT write your name on this survey. The answers you give will be kept private.

NO one will know what you write. Answer the questions based on what you really do and know.

Completing the survey is voluntary. Whether or not you answer the questions will not affect your grade in this class. Try to answer all the questions. If you do not want to answer a question, just leave it blank. There are no wrong answers.

The questions that ask about your background will only be used to describe the types of students completing this survey. The information will not be used to find out your name. No names will ever be reported.

Please read every question. Try to answer all the questions. Fill in the circles in the booklet completely. When you are finished, follow the instructions of the person giving you the survey.

Thank You Very Much For Your Help.

The first five questions ask for some background information about you.

1. How old are you?
   a. 9 years old
   b. 10 years old
   c. 11 years old
   d. 12 years old
   e. 13 years old
   f. 14 years old
   g. 15 years old
   h. 16 years old
   i. 17 years old
   j. 18 years old
   k. 19 years old or older

2. What is your sex?
   a. Male
   b. Female

3. What grade are you in?
   a. 6th
   b. 7th
   c. 8th
   d. 9th
   e. 10th
   f. 11th
   g. 12th
   h. Ungraded or other grade

4. Are you Hispanic, Latino/a, or Spanish origin (One or more categories may be selected)?
   a. No, not of Hispanic, Latino/a, or Spanish origin
   b. Yes, Mexican, Mexican American, Chicano or Chicana
   c. Yes, Puerto Rican
   d. Yes, Cuban
   e. Yes, Another Hispanic, Latino/a, or Spanish origin

5. What race or races do you consider yourself to be? (**You can CHOOSE ONE ANSWER or MORE THAN ONE ANSWER**)
   a. American Indian or Alaska Native
   b. Asian
   c. Black or African American
   d. Native Hawaiian or Other Pacific Islander
   e. White

The next five sections of questions ask about your use of particular kinds of tobacco products.

The first 13 questions are about smoking cigarettes.

6. Have you ever been curious about smoking a cigarette?
   a. Definitely yes
   b. Probably yes
   c. Probably not
   d. Definitely not

7. Have you **ever tried** cigarette smoking, even one or two puffs?
   a. Yes
   b. No

8. Do you think you will smoke a cigarette in the next year?
   a. Definitely yes
   b. Probably yes
   c. Probably not
   d. Definitely not

9. Do you think that you will try a cigarette soon?
   a. Definitely yes
   b. Probably yes
   c. Probably not
   d. Definitely not

10. If one of your best friends were to offer you a cigarette, would you smoke it?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

11. How old were you when you **first tried** cigarette smoking, even one or two puffs?
    a. I have never smoked cigarettes, not even one or two puffs
    b. 8 years old or younger
    c. 9 years old
    d. 10 years old
    e. 11 years old
    f. 12 years old
    g. 13 years old
    h. 14 years old
    i. 15 years old
    j. 16 years old
    k. 17 years old
    l. 18 years old
    m. 19 years old or older

12. About how many cigarettes have you smoked in your **entire life**?
    a. I have never smoked cigarettes, not even one or two puffs
    b. 1 or more puffs but never a whole cigarette
    c. 1 cigarette
    d. 2 to 5 cigarettes
    e. 6 to 15 cigarettes (about 1/2 a pack total)
    f. 16 to 25 cigarettes (about 1 pack total)
    g. 26 to 99 cigarettes (more than 1 pack, but less than 5 packs)
    h. 100 or more cigarettes (5 or more packs)

13. During the **past 30 days**, on how many days did you smoke cigarettes?
    a. 0 days
    b. 1 or 2 days
    c. 3 to 5 days
    d. 6 to 9 days
    e. 10 to 19 days
    f. 20 to 29 days
    g. All 30 days

14. During the past 30 days, **on the days you smoked**, about how many cigarettes did you smoke per day?
    a. I did not smoke cigarettes during the past 30 days
    b. Less than 1 cigarette per day
    c. 1 cigarette per day
    d. 2 to 5 cigarettes per day
    e. 6 to 10 cigarettes per day
    f. 11 to 20 cigarettes per day
    g. More than 20 cigarettes per day

15. When was the last time you smoked a cigarette, even one or two puffs? (**PLEASE CHOOSE THE FIRST ANSWER THAT FITS**)
    a. I have never smoked cigarettes, not even one or two puffs
    b. Earlier today
    c. Not today but sometime during the past 7 days
    d. Not during the past 7 days but sometime during the past 30 days
    e. Not during the past 30 days but sometime during the past 6 months
    f. Not during the past 6 months but sometime during the past year
    g. 1 to 4 years ago
    h. 5 or more years ago

16. During the past 30 days, what brand of cigarettes did you usually smoke? (**CHOOSE ONLY ONE ANSWER**)
    a. I did not smoke cigarettes during the past 30 days
    b. I did not smoke a usual brand
    c. American Spirit
    d. Camel
    e. GPC, Basic, or Doral
    f. Kool
    g. Lucky Strike
    h. Marlboro
    i. Newport
    j. Parliament
    k. Virginia Slims
    l. Some other brand not listed here
    m. Not sure

17. Menthol cigarettes are cigarettes that taste like mint. During the past 30 days, were the cigarettes that you usually smoked menthol?
    a. I did not smoke cigarettes during the past 30 days
    b. Yes
    c. No
    d. Not sure

18. How likely is it that you will try to purchase cigarettes within the **next 30 days**?
    a. I do not smoke cigarettes
    b. Very likely
    c. Somewhat likely
    d. Somewhat unlikely
    e. Very unlikely

<u>The next seven questions are about the use of cigars, cigarillos or little cigars such as Black and Mild, Swisher Sweets, Dutch Masters, White Owl, or Phillies Blunts.</u>

19. Have you ever been curious about smoking a cigar, cigarillo, or little cigar such as Black and Mild, Swisher Sweets, Dutch Masters, White Owl, or Phillies Blunts?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

20. Have you **ever tried** smoking cigars, cigarillos, or little cigars, such as Black and Mild, Swisher Sweets, Dutch Masters, White Owl, or Phillies Blunts, even one or two puffs?
    a. Yes
    b. No

21. Do you think that you will try a cigar, cigarillo or little cigar soon?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

22. If one of your best friends were to offer you a cigar, cigarillo or little cigar, would you smoke it?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

23. How old were you when you **first tried** smoking a cigar, cigarillo, or little cigar, even one or two puffs?
    a. I have never smoked cigars, cigarillos, or little cigars, not even one or two puffs
    b. 8 years old or younger
    c. 9 years old
    d. 10 years old
    e. 11 years old
    f. 12 years old
    g. 13 years old
    h. 14 years old
    i. 15 years old
    j. 16 years old
    k. 17 years old
    l. 18 years old
    m. 19 years old or older

24. During the **past 30 days**, on how many days did you smoke cigars, cigarillos, or little cigars?
    a. 0 days
    b. 1 or 2 days
    c. 3 to 5 days
    d. 6 to 9 days
    e. 10 to 19 days
    f. 20 to 29 days
    g. All 30 days

25. During the past 30 days, **on the days that you smoked**, about how many cigars, cigarillos, or little cigars did you smoke per day?
    a. I did not smoke cigars, cigarillos, or little cigars during the past 30 days
    b. Less than 1 cigar, cigarillo or little cigar per day
    c. 1 per day
    d. 2 to 5 per day
    e. 6 to 10 per day
    f. 11 to 20 per day
    g. More than 20 per day

<u>The next four questions are about the use of chewing tobacco, snuff or dip. Do not think about using snus when you answer these questions.</u>

26. Have you ever been curious about using chewing tobacco, snuff, or dip, **such as** Redman, Levi Garrett, Beechnut, Skoal, Skoal Bandits, or Copenhagen?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

27. Have you **ever used** chewing tobacco, snuff, or dip, such as Redman, Levi Garrett, Beechnut, Skoal, Skoal Bandits, or Copenhagen, even just a small amount?
    a. Yes
    b. No

28. How old were you when you **used** chewing tobacco, snuff, or dip for the first time?
    a. I have never used chewing tobacco, snuff, or dip
    b. 8 years old or younger
    c. 9 years old
    d. 10 years old
    e. 11 years old
    f. 12 years old
    g. 13 years old
    h. 14 years old
    i. 15 years old
    j. 16 years old
    k. 17 years old
    l. 18 years old
    m. 19 years old or older

29. During the **past 30 days**, on how many days did you use chewing tobacco, snuff, or dip?
    a. 0 days
    b. 1 or 2 days
    c. 3 to 5 days
    d. 6 to 9 days
    e. 10 to 19 days
    f. 20 to 29 days
    g. All 30 days

<u>The next six questions are about the use of electronic cigarettes or e-cigarettes such as blu, 21st Century Smoke or NJOY.</u>

30. Have you ever been curious about using an electronic cigarette or e-cigarette such as blu, 21st Century Smoke or NJOY?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

31. Have you **ever tried** an electronic cigarette or e-cigarette such as blu, 21st Century Smoke or NJOY?
    a. Yes
    b. No

32. Do you think that you will try an electronic cigarette or e-cigarette soon?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

33. If one of your best friends were to offer you an electronic cigarette or e-cigarette, would you use it?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

34. How old were you when you **first tried** using an electronic cigarette or e-cigarette?
    a. I have never used electronic cigarettes or e-cigarettes
    b. 8 years old or younger
    c. 9 years old
    d. 10 years old
    e. 11 years old
    f. 12 years old
    g. 13 years old
    h. 14 years old
    i. 15 years old
    j. 16 years old
    k. 17 years old
    l. 18 years old
    m. 19 years old or older

35. During the **past 30 days**, on how many days did you use electronic cigarettes or e-cigarettes such as blu, 21st Century Smoke, or NJOY?
    a. 0 days
    b. 1 or 2 days
    c. 3 to 5 days
    d. 6 to 9 days
    e. 10 to 19 days
    f. 20 to 29 days
    g. All 30 days

The next two questions are about the use of other tobacco products, not described in the previous sections.

36. Which of the following tobacco products have you **ever tried**, even just one time? (**CHOOSE ALL THAT APPLY)**
    a. Smoking tobacco from a hookah or waterpipe
    b. Pipe filled with tobacco (not waterpipe)
    c. Snus, such as Camel or Marlboro Snus
    d. Dissolvable tobacco products such as Ariva, Stonewall, Camel orbs, Camel sticks, Marlboro sticks or Camel strips
    e. Bidis (small brown cigarettes wrapped in a leaf)
    f. I have never tried any of the products listed above

37. In the **past 30 days**, which of the following products have you used on **at least one day**? (**CHOOSE ALL THAT APPLY**)
    a. Smoking tobacco from a hookah or waterpipe
    b. Pipe filled with tobacco (not waterpipe)
    c. Snus, such as Camel or Marlboro
    d. Dissolvable tobacco products, such as Ariva, Stonewall, Camel orbs, Camel sticks, or Camel strips
    e. Bidis (small brown cigarettes wrapped in a leaf)
    f. I have not used any of the products listed above in the past 30 days

The next question asks about flavors in tobacco products.

38. Which of the following tobacco products that you used in the past 30 days were flavored to taste like menthol (mint), alcohol (wine, cognac), candy, fruit, chocolate or other sweets (**CHOOSE ALL THAT APPLY**)?
    a. Cigars, cigarillos, or little cigars
    b. Chewing tobacco, snuff, or dip
    c. Electronic cigarettes or e-cigarettes
    d. Smoking tobacco out of a hookah or waterpipe
    e. Pipe filled with tobacco (not waterpipe)
    f. Snus
    g. Dissolvable tobacco products
    h. I did not use flavored tobacco products in the past 30 days

The next question asks about the first tobacco product ever tried.

39. Which of the following tobacco products did you try first (CHOOSE ONLY ONE ANSWER)?
    a. Cigarettes
    b. Cigars, cigarillos, or little cigars
    c. Chewing tobacco, snuff, or dip
    d. Electronic cigarettes or e-cigarettes
    e. Some other tobacco product
    f. Not sure about the product I tried first
    g. I have never tried any tobacco products

The next four questions are about getting tobacco products.

40. During the past 30 days, how did you get your own tobacco products? (You can CHOOSE ONE ANSWER or MORE THAN ONE ANSWER)
    a. I did not get any tobacco products during the past 30 days
    b. I bought them myself
    c. I had someone else buy them for me
    d. I asked someone to give me some
    e. Someone offered them to me
    f. I bought them from another person
    g. I took them from a store or another person
    h. I got them some other way

41. During the past 30 days, where did you buy your own tobacco products? (You can CHOOSE ONE ANSWER or MORE THAN ONE ANSWER)
    a. I did not buy tobacco products during the past 30 days
    b. A gas station or convenience store
    c. A grocery store
    d. A drugstore
    e. A vending machine
    f. Over the Internet
    g. Through the mail
    h. Some other place not listed here

42. During the **past 30 days**, did anyone **refuse** to sell you any tobacco products because of your age?
    a. I did not try to buy any tobacco products during the past 30 days
    b. Yes
    c. No

43. How easy do you think it is for kids your age to buy tobacco products in a store?
    a. Easy
    b. Somewhat easy
    c. Not easy at all

The next four questions are about issues related to urges or needs to use tobacco products.

44. During the past 30 days, have you had a strong craving or felt like you really needed to use a tobacco product of any kind (such as smoking a cigarette or cigar, or using chewing tobacco)?
    a. Yes
    b. No

45. During the past 30 days, was there a time when you wanted to use a tobacco product so much that you found it difficult to think of anything else?
    a. Yes
    b. No

46. How soon after you wake up do you want to use a tobacco product?
    a. I do not want to use tobacco
    b. Within 5 minutes
    c. From 6 to 30 minutes
    d. From more than 30 minutes to 1 hour
    e. After more than 1 hour but less than 24 hours
    f. I rarely want to use tobacco

47. How true is this statement for you? I feel restless and irritable when I don't use tobacco for a while.
    a. I do not use tobacco
    b. Not at all true
    c. Sometimes true
    d. Often true
    e. Always true

The next four questions are about quitting tobacco products.

48. Are you seriously thinking about quitting cigarettes? (PLEASE CHOOSE THE FIRST ANSWER THAT FITS)
    a. I do not smoke cigarettes
    b. Yes, within the next 30 days
    c. Yes, within the next 6 months
    d. Yes, within the next 12 months
    e. Yes, but not within the next 12 months
    f. No, I am not thinking about quitting cigarettes

49. During the **past 12 months**, how many times have you stopped smoking **cigarettes** for **one day or longer** because you were trying to quit smoking cigarettes **for good**?
    a. I did not smoke cigarettes during the past 12 months
    b. I did not try to quit during the past 12 months
    c. 1 time
    d. 2 times
    e. 3 to 5 times
    f. 6 to 9 times
    g. 10 or more times

50. Are you seriously thinking about quitting the use of all tobacco products? (PLEASE CHOOSE THE FIRST ANSWER THAT FITS)
    a. I do not use tobacco products
    b. Yes, within the next 30 days
    c. Yes, within the next 6 months
    d. Yes, within the next 12 months
    e. Yes, but not within the next 12 months
    f. No, I am not thinking about quitting the use of all tobacco products

51. During the **past 12 months**, how many times have you stopped using **all tobacco products** for **one day or longer** because you were trying to quit all tobacco products **for good**?
    a. I did not use tobacco products during the past 12 months
    b. I did not try to quit all tobacco products during the past 12 months
    c. 1 time
    d. 2 times
    e. 3 to 5 times
    f. 6 to 9 times
    g. 10 or more times

The next questions ask about your thoughts on tobacco products.

52. How much do you think people harm themselves when they smoke cigarettes some days but not every day?
    a. No harm
    b. Little harm
    c. Some harm
    d. A lot of harm

53. How much do you think people harm themselves when they smoke **cigars, cigarillos or little cigars** some days but not every day?
    a. No harm
    b. Little harm
    c. Some harm
    d. A lot of harm

54. Do you believe that **cigars, cigarillos, or little cigars** are (LESS HARMFUL, EQUALLY HARMFUL, or MORE HARMFUL) than cigarettes?
    a. Less harmful
    b. Equally harmful
    c. More harmful
    d. I have never heard of cigars, little cigars or cigarillos
    e. I don't know enough about these products

55. Do you believe that **cigars, cigarillos or little cigars** are (LESS ADDICTIVE, EQUALLY ADDICTIVE, or MORE ADDICTIVE) than cigarettes?
    a. Less addictive
    b. Equally addictive
    c. More addictive
    d. I have never heard of cigars, cigarillos, or little cigars
    e. I don't know enough about these products

56. How much do you think people harm themselves when they use **chewing tobacco, snuff, dip, or snus**, some days but not every day?
    a. No harm
    b. Little harm
    c. Some harm
    d. A lot of harm

57. Do you believe that **chewing tobacco, snuff, dip, or snus** is (LESS HARMFUL, EQUALLY HARMFUL, or MORE HARMFUL) than cigarettes?
    a. Less harmful
    b. Equally harmful
    c. More harmful
    d. I have never heard of chewing tobacco, snuff, dip, or snus
    e. I don't know enough about these products

58. Do you believe that **chewing tobacco, snuff, dip or snus** is (LESS ADDICTIVE, EQUALLY ADDICTIVE, or MORE ADDICTIVE) than cigarettes?
    a. Less addictive
    b. Equally addictive
    c. More addictive
    d. I have never heard of chewing tobacco, snuff, dip or snus
    e. I don't know enough about these products

59. How much do you think people harm themselves when they use **e-cigarettes** some days but not every day?
    a. No harm
    b. Little harm
    c. Some harm
    d. A lot of harm

60. Do you believe that **e-cigarettes** are (LESS HARMFUL, EQUALLY HARMFUL, or MORE HARMFUL) than regular cigarettes?
    a. Less harmful
    b. Equally harmful
    c. More harmful
    d. I have never heard of e-cigarettes
    e. I don't know enough about these products

61. Do you believe that **e-cigarettes** are (LESS ADDICTIVE, EQUALLY ADDICTIVE, or MORE ADDICTIVE) than cigarettes?
    a. Less addictive
    b. Equally addictive
    c. More addictive
    d. I have never heard of e-cigarettes
    e. I don't know enough about these products

62. How strongly do you agree with the statement 'All tobacco products are dangerous'?
    a. Strongly agree
    b. Agree
    c. Disagree
    d. Strongly disagree

63. Do you think that breathing smoke from other people's cigarettes or other tobacco products causes…
    a. No harm
    b. Little harm
    c. Some harm
    d. A lot of harm

The next ten questions ask about different issues related to tobacco.

64. When you are using the Internet, how often do you see ads or promotions for cigarettes or other tobacco products?
    a. I do not use the Internet
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

65. When you read newspapers or magazines, how often do you see ads or promotions for cigarettes or other tobacco products?
    a. I do not use read newspapers or magazines
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

66. When you go to a convenience store, supermarket, or gas station, how often do you see ads or promotions for cigarettes or other tobacco products?
    a. I never go to a convenience store, supermarket, or gas station
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

67. When you watch TV or go to the movies, how often do you see actors and actresses using cigarettes or other tobacco products?
    a. I do not watch TV or go to the movies
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

68. When you are using the Internet, how often do you see ads or promotions for electronic cigarettes or e-cigarettes?
    a. I do not use the Internet
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

69. When you read newspapers or magazines, how often do you see ads or promotions for electronic cigarettes or e-cigarettes?
    a. I do not use read newspapers or magazines
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

70. When you go to a convenience store, supermarket, or gas station, how often do you see ads or promotions for electronic cigarettes or e-cigarettes?
    a. I never go to a convenience store, supermarket, or gas station
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

71. When you watch TV or go to the movies, how often do you see ads or promotions for electronic cigarettes or e-cigarettes?
    a. I do not watch TV or go to the movies
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

72. In the past 30 days, how often have you thought about the harmful chemicals in tobacco products?
    a. Never
    b. Rarely
    c. Sometimes
    d. Often
    e. Very often

73. During the past 30 days, how often did you see a warning label on a smokeless tobacco product such as chewing tobacco, snuff, dip, or snus?
    a. I did not see a smokeless tobacco product during the past 30 days
    b. Never
    c. Rarely
    d. Sometimes
    e. Most of the time
    f. Always

The nex six questions ask about your thoughts on people your age who use cigarettes, cigars, smokeless tobacco and e-cigarettes

74. Do you think **smoking cigarettes** makes young people look cool or fit in?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

75. Do you think young people who **smoke cigarettes** have more friends?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

76. Do you think **smoking cigars, cigarillos or little cigars** makes young people look cool or fit in?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

77. Do you think young people who **smoke cigars, cigarillos, or little cigars** have more friends?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

78. Do you think young people who **use chewing tobacco, snuff, or dip** have more friends?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

79. Do you think young people who **use electronic cigarettes or e-cigarettes** have more friends?
    a. Definitely yes
    b. Probably yes
    c. Probably not
    d. Definitely not

The next question asks about your experiences at home.

80. Does anyone who lives with you now…? (**CHECK ALL THAT APPLY**).
    a. Smoke cigarettes
    b. Smoke cigars, cigarillos, or little cigars
    c. Use chewing tobacco, snuff, or dip
    d. Use electronic cigarettes or e-cigarettes
    e. Smoke tobacco from a hookah or waterpipe
    f. Smoke pipes filled with tobacco (not waterpipes)
    g. Use snus
    h. Use dissolvable tobacco products
    i. Smoke bidis (small brown cigarettes wrapped in a leaf)
    j. No one who lives with me now uses any form of tobacco

Some cigarette or other tobacco companies make items like sports gear, T-shirts, hats, jackets, sunglasses or other items that people can buy or receive for free.

81. How likely is it that you would ever use or wear something--such as a t-shirt, hat, or sunglasses--that has a tobacco brand name, logo, or picture on it?
    a. Very likely
    b. Somewhat likely
    c. Somewhat unlikely
    d. Very unlikely

Appendix B

Student Weight Detail

APPENDIX B. STUDENT WEIGHT DETAIL

Students are selected from schools via the selection of intact class sections as described in Section 2.2.4. The student sampling weight is computed based on a ratio of enrolling to responding students described in Section 4.2.1. 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—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, we omit the subscripts denoting the sampling stages and weight class. The unsubscripted quantities presented are assumed to be within weight class *c,* as defined in section 4.2.1.

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 student weighting class. The initial selection probability is taken to be the inverse of this sampling probability.

In our simplified notation, letting K represent the number of sampled class sections, we have:

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

Nonresponse Adjustment

The nonresponse adjustment inflates the weight of the responding students to equal that of the sampled students. The adjustment is 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:

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

Simplifying, we get

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

Appendix C

Common Core of Data Race/Ethnicity Definitions

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.