2012 National Youth Tobacco Survey

METHODOLOGY REPORT

PRINT ONLY

Table of Contents

List of Exhibits and Figure

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 data sources such as the Youth Risk Behavior Surveillance System (YRBSS) that provide prevalence estimates for selected tobacco use behaviors among high school students by providing more comprehensive data about both middle school (grades 6–8) and high school (grades 9–12) students regarding tobacco use (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 magnitude of youth tobacco use. The NYTS provides multiple measures and data for the Healthy People 2020 objectiveTU-18 [including four sub-objectives: TU-18.1 (internet advertising and promotion), TU-18.2 (magazine and newspaper advertising and promotion), TU-18.3 (store advertising and promotion), and TU-18.4 (Seeing actors using tobacco on television or movies)] (USDHHS, 2010).

First conducted during fall 1999 and again during spring 2000, 2002, 2004, 2006, 2009, 2011 and 2012, the NYTS surveys provide data that are representative of all middle school and high school students in the 50 states and the District of Columbia. The current NYTS was implemented in spring 2013.

1.2 Overview of the 2012 NYTS Methodology

The 2012 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 and Hispanic students were oversampled. Sampling procedures were probabilistic and conducted without replacement at all stages, and entailed selection of Primary Sampling Units (PSUs) (a county, or a group of small counties, or part of a very large county) within each created stratum, of schools within each selected PSU, and, lastly, of students within each selected school. Participating students completed the survey via pencil and paper self-administered scannable questionnaire booklet.

Participation in the NYTS was voluntary at both the school and student level. At the student level, participation was also anonymous. Schools used either passive or active parental permission forms at their 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 284 schools, of which 228 participated, for a school participation rate of 80.3%. The survey yielded 24,658 completed student questionnaires out of a sample of 26,873 students for a student participation rate of 91.7%. The overall participation rate, the product of the school-level and student-level participation rates, was 73.6%.

---

[1] CDC. *Best Practices for comprehensive tobacco control programs*. Atlanta, GA: US Department of Health and and Human Services, Public Health Service, CDC; 1999.

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

A weighting factor was applied to each student record to adjust for non-response and for varying probabilities of selection. Weights were adjusted to ensure that the weighted proportions of students in each grade matched national population proportions. Final adjusted weights were scaled to ensure that the weighted count of students was equal to the total sample size.

The remainder of this report provides detailed information on the methodology used in the 2012 NYTS sample selection, data collection, and weighting of student response data, with the three following chapters arranged to follow the project workflow:

- Chapter 2: Sampling
- Chapter 3: Data Collection
- Chapter 4: Weighting

Chapter 2—NYTS Sampling Method

The next sections describe the sampling design (Sections 2.1 to 2.3) and then the selection methods (Section 2.4). Section 2.1 provides an overview of the sampling stages and defines the measure of size used in selecting primary sampling units and schools with probabilities proportional to size (PPS). Section 2.2 describes the stratification adopted for primary sampling units and schools. Section 2.3 discusses the sample sizes developed for the study. Section 2.4 describes the sample selection methods used at the various sampling stages.

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, Catholic, and other private school students enrolled in grades 6 through 12 in the United States. More specifically, the study was designed to produce national estimates at a 95% confidence level by school level (middle school and high school), by grade (6, 7, 8, 9, 10, 11, and 12), by sex, and by race/ethnicity for non-Hispanic white, non-Hispanic black, and Hispanic students. Additional estimates, such as cross-tabulations of grade by sex and of race/ethnicity by school level, were also supported; however, precision levels will vary considerably according to differences in sub-population sizes.

The universe for the study consisted of all public, Catholic, and other private and charter school students enrolled in regular middle schools and high schools in grades 6 through 12 in the 50 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 NYTS study is a continuation of the NYTS survey cycles that took place in 1999, 2000, 2002, 2004, 2006, 2009 and 2011. The NYTS survey system employs a repeat cross-sectional design to develop national estimates of tobacco use behaviors and exposure to pro- and anti-tobacco influences among students enrolled in grades 6–12. Unlike the 2011 survey cycle, the 2012 NYTS was not coordinated with the national Youth Risk Behavior Survey (YRBS) as no national YRBS was administered in 2012. However, while designed as a stand-alone sample, the 2012 NYTS sampling design replicated key aspects of the 2011 NYTS.

2.1.1 Sampling Stages and Measure of Size

The three-stage cluster sample was stratified by racial/ethnic composition and urban versus non-urban status at the first (primary) stage. Primary Sampling Units (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.; otherwise, they were classified as "non-urban." Additional, 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 primary sampling unit (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.

The sampling stages may be summarized as follows:

- Selection of PSUs—At the first sampling stage, 100 PSUs were selected from sixteen 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, two large schools were selected from each PSU. An additional 24 medium schools and 20 small schools were selected from subsample PSUs. 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— At the third stage, 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 both represented and represented one time only.

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, we selected an average of 1.85 classes per grade by selecting 2 classes per grade in selected large schools, and one class per grade in the remaining schools. The double class sampling took place in 80 randomly chosen schools.

The sampling approach utilized Probability Proportional to Size (PPS) sampling methods. In PPS sampling, when the measure of size (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.

Note that in 2012 we moved to the use of total eligible enrollment as a measure of size, rather than a weighted measure of size that increased the probability of selecting high minority schools. This design update was enabled by the changing demographic composition of the national student population, and increases the statistical efficiency of the sample. By using total enrollment as MOS, there was no need to use a formula to calculate the size measure – we simply took total enrollment. This is equivalent to setting the coefficients in the formula used in prior cycles of the NYTS to 1.0; meaning the formula is now 1.0 B + 1.0 H + 1.0 O, which evaluates total enrollment.

The measure of size was used also 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 summary of the sampling design features.

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	100	Urban vs. non-urban (2 strata); Minority concentration (8 strata)	Aggregate school size in target grades
2	Schools	244 school selections: 200 large schools (2 per PSU), 24 medium schools and 20 small schools	Small, medium and large; High-school vs. middle-school	Weighted enrollment (increased for black, Hispanic groups)
3	Classes/ students	1 or 2 classes per grade (2 per grade in large high-minority schools)		

2.1.2 Stratification and Linking

This section describes frame preparation steps for the selection of the first- and second-stage samples of PSUs and schools. These steps include combining counties into Primary Sampling Units, linking schools into Secondary Sampling Units, and the stratification and allocation methods at these stages.

The basis for the sampling frame is a comprehensive database of U.S. schools and school districts, obtained from a commercial vendor. The school facility data are continuously updated on a rolling basis with school contact information, facility information, and openings and closings. The frame also incorporates enrollment data, which serves as the basis for the measure of size used in the sampling. Enrollment data were obtained from the most recent Common Core of Data from the National Center for Education Statistics, which are merged on a rolling basis into the current school and school district data files of Quality Education Data, Inc. (formerly QED), acquired by MCH Strategic Data (http://www.mchdata.com/qed).

2.1.2.1 PSU

Defining a PSU

In defining PSUs, several issues are considered:

a. 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[3].
b. Each PSU should be compact geographically so that field staff can go from school to school easily.
c. There should be recent data available to characterize the PSUs.
d. PSU definitions should be consistent with secondary sampling unit (school) definitions
e. PSU should contain at least 4 middle and 5 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. County population figures were aggregated from school enrollment data for the grades of interest.

The 2012 NYTS PSU definitions were based on the 2011 definitions, and 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.

County population figures were aggregated from school enrollment data for the grades of interest.

[3] Operationally, we define “near certainty” as a selection probability of .8 or greater.

Stratification of PSUs

The PSUs were organized into 16 strata, based on urban/non-urban location (as defined in section 2.1.1) and minority enrollment. The approach involves the computation of optimum stratum boundaries using the cumulative square root of "f" method developed by Dalenius and Hodges[4,5]. 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 "non-urban."
- If the percentage of Hispanic students in the PSU exceeded the percentage of black students, then the PSU is classified as Hispanic. Otherwise it is classified as black.
- Hispanic urban and Hispanic non-urban 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 56 percent
    - For non-urban, High Hispanic PSU, the percentage cut points used to define the groups were 22, 48 and 60 percent.
- Black urban and black non-urban PSUs were also classified into four groupings depending upon the percentages of blacks in the PSU.
    - For urban, High black PSU, the percentage cut points used to define the groups were 26, 38 and 56 percent.
    - For non-urban High Black PSU, the percentage cut points used to define the groups were 18, 32 and 56 percent.

Allocation of the PSU Sample

We designed and selected a sample of 100 PSUs, allocated in proportion to student enrollment. Using simulations as in previous studies, we then made adjustments to the initial allocation to meet minority targets. Specifically, the adjustments rounded fractional allocations, ensured that each strata 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.

---

[4] Potter F. Survey of Procedures to Control Extreme Sampling Weights, in Proceedings of the Section on Survey Research Methods, American Statistical Association 1988;pp 453458.

[5] Potter F. A Study of Procedures to Identify and Trim Extreme Sample Weights, Proceedings of the Section on Survey Research Methods, American Statistical Association 1990;pp. 225-230.

Exhibit 2-2: Stratum Definition and PSU Allocation to Strata

Predominant Minority	Urban/ Non-urban	Density Group Number	Stratum Code	Student Population	Percent of Student Population	Number of Sample PSUs
Black	Urban	1	BU1	2,633,051	16.29%	9
		2	BU2	2,050,303	12.69%	7
		3	BU3	825,252	5.11%	3
		4	BU4	331,731	2.05%	2
	Non-urban	1	BR1	3,854,233	23.85%	14
		2	BR2	1,817,124	11.24%	6
		3	BR3	1,374,191	8.50%	5
		4	BR4	468,702	2.90%	2
Hispanic	Urban	1	HU1	3,236,277	20.03%	14
		2	HU2	1,876,056	11.61%	4
		3	HU3	1,820,012	11.26%	3
		4	HU4	1,712,989	10.60%	2
	Non-urban	1	HR1	3,789,654	23.45%	11
		2	HR2	1,128,116	6.98%	6
		3	HR3	864,675	5.35%	6
		4	HR4	444,098	2.75%	6

2.1.2.2 Schools

Linking into SSU

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 relevant grades. This process is illustrated in Figure 2-1, where fragment school A is linked with whole school B, to form a linked school, or Secondary Sampling Unit (SSU) XXX. 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 a single school, or sampling unit, during sampling with selection performed at the grade level as described below.

Figure 2-1: Linked School Construction and Grade Sampling for High Schools

Cluster School XXX

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

Stratification

Linked schools (SSUs) were stratified by school level and by size. SSUs were split into middle schools (those that contained grades 6, 7, and 8), and high schools (those that contained grades 9, 10, 11 and 12). Schools that contained a mix of high- and middle-school grades were split into two sampling units, one for each level. SSU size was determined based on their ability to support two, one, or less than one selection per grade. Large SSUs contained at least 50 students per grade level, medium SSU contained between 25 and 49 students per grade level, small SSUs contained less than 25 students at any grade level.

2.1.3 Sample Sizes

Class selection was based on the size for each linked school. Large schools are constructed to support a draw of up to two classes (average of 25 students per class) per grade, medium schools a draw one class (average of 20 students per class), and small schools a draw of all available students (average of 12 students per class).

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 the key estimates accurate to within ± 5% at a 95% precision level. 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.

Specifically, the NYTS is designed to produce accurate estimation to within ± 5% at a 95% precision level for the following key subgroup estimates:

- Middle and high school estimates (school level)—middle school students in total (grades 6–8 combined) and high school students in total (grades 9–12 combined)
- Grade estimates—Individual grades 6, 7, 8, 9, 10, 11, and 12, separately
- Sex group estimates—males and females in total, by school level (male middle school students, female high school students, etc.), and by individual grade (6^{th} grade males, 6^{th} grade females, etc.)
- Racial group estimates (race/ethnicity)—in total and by school level (e.g., Hispanic middle school students)

Over the past several cycles of the NYTS, we have confirmed that sample sizes, and resulting student yields, were sufficient to achieve design goals in terms of precision. For the 2012 NYTS design, anticipated precision levels were developed in order to ensure that the 2012 design meets the original precision targets. The balance of this section presents this development. The number of schools selected and students per school were calculated so that, combined with anticipated response rates, we projected the study would obtain completed surveys from approximately 24,500 students.

As detailed in Section 2.2.2, linked schools (SSUs) were constructed so as to contain a full complement of grades—6 to 8 for middle schools, and 9 to 12 for high schools.

Schools were further classified by size based on grade-level enrollments; the definition of size strata is provided in Section 2.2.2. This allows us to ensure that a sampled school of a given size classification is able to support the student sample sizes given in Exhibit 2-3.

Across the eight cycles of the NYTS, the school participation has averaged 90%, with a low of 83%. Student participation has averaged 91% with a low of 88%. To be conservative, we have assumed slightly lower values in developing the sample design for the 2012 NYTS: 85% for schools and 85% for students.

Exhibit 2-3 summarizes the designed sample sizes for each SSU type. This table details the number of schools that were specified to be drawn by the sample design along with the number of participating schools and students anticipated when we developed the sample design based on the given assumptions. Section 3.4 compares these projections to the actual sample yields.

Exhibit 2-3: Sample Size Projections for Participating Students on the 2012 NYTS

School Level/ Size Class	Class Sampling	Number of Schools Selected	Anticipated Number of Responding Schools	Classes Selected Per Grade	Anticipated Number of Students Selected Per School	Anticipated Number of Selected Students	Anticipated Number of Responding Students	Per Grade
High School								
Large	Double class sampling	80	68	2	200	13,600	11,560	
	Single class sampling	20	17	1	100	1,700	1,445	
Medium	Single class sampling	12	10	1	80	800	680	
Small	Single class sampling	10	9	n/a	48	432	367	
Total Schools/SSU		**122**	**104**			**16,532**	**14,052**	**3,513**
Middle School								
Large	Double class sampling	80	68	2	150	10,200	8,670	
	Single class sampling	20	17	1	75	1,275	1,084	
Medium	Single class sampling	12	10	1	60	600	510	
Small	Single class sampling	10	9	n/a	36	324	275	
Total Schools/SSU		**122**	**104**			**12,399**	**10,539**	
Grand Total Schools/SSU		**244**	**208**			**28,931**	**24,591**	**3,513**

The design anticipated that of the 244 drawn SSUs, 208 schools would respond and participate in the study. The actual number of responding schools was higher. The 244 drawn SSUs yielded 284 physical schools, of which 228 responded. The numbers of large, medium and small schools were calibrated to generate the required numbers of students overall, by school level and by grade.

The large projected sample size permits analysis by individual grade and by sex without any special considerations in the sampling plan. Additionally, grade and sex subgroups both typically cut across schools; and sex subgroups typically cut across grades with males and females each constituting about half of the students selected. 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 than are less evenly dispersed across schools (e.g., racial/ethnic groups). Thus, the designed confidence intervals were ± 3%.

Because the design expected to yield a greater number of completed surveys from high school students than from middle school students, overall estimates were anticipated to be more precise at the high school level than those at the middle school level. Moreover, because within grade estimates by sex have slightly larger standard errors than those for estimates by grade alone, estimates of sex were expected within ± 5%.

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

2.1.3.1 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 grades 7–9 junior high schools and by grades 9–12 high schools. As a result, we have developed the school linking approach described in Section 2.2.2, and with this approach being applied independently for high schools and middle schools.

2.1.3.2 Grade Estimates

NYTS estimates are typically not reported by grade level but rather by school level. Still, the design balances the sample sizes for grade level by targeting 3,000 students per grade. This ensures that estimates at the grade level achieve the required precision levels. It is worth noting that this design feature resulted in a larger student allocation to the high school stratum than to the middle school stratum as high schools have four grades versus three grades for middle schools.

2.1.3.3 Sex Group Estimates

The large designed sample size permitted analysis by sex without any special considerations in the sampling plan.

2.1.3.4 Racial Group Estimates

In order to support separate analysis of the data for white, black and Hispanic students, in total and by school level, adequate sample sizes were required by the designed for subgroups defined (1) by school level by racial grouping or (2) by sex by racial 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).

2.2 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 is a group of schools, multiplied by the conditional probability of selecting the student's school, multiplied by the conditional probability of selecting the student's class. These steps are detailed in the selection below.

2.2.1 PSU

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 probability proportional to size (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 100 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 12 of the sample PSUs for the medium school sample for each school level (middle school and high school)
- Subsample at random 10 of the sample PSUs for the small school sample for each school level (middle school and high school)

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*:

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

As noted below, 10 of the 100 sample PSUs were sub-sampled for the sampling of small schools. Similarly, 12 PSUs were sub-sampled for the sampling of medium schools. In these PSU, whether sub-sampled for the medium or small school draw, two schools were selected at each level. Thus, the sub-sample PSUs have an additional factor in their selection probability for these classes of schools. This factor is incorporated into the school sampling probability below, as it is more closely associated with school selection.

2.2.2 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 weighted measure of size.

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 a sub-sample of 10 PSUs for the sampling of small schools and a subsample of 12 PSUs for medium school sampling at each level. Then, as with the large schools, two schools – one high school and one middle school – were selected from each sub-sampled PSU with probability proportional to weighted measure of size. Note that as the PSUs were sub-sampled independently, a PSU could contribute a maximum of six schools to the sample.

Replacement of Schools/School Systems

We did not replace refusing school districts, schools, classes, or students. We allowed for school and student non-response by inflating the sample sizes to account for non-response. 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 small schools, one school was drawn from sub-sampled PSU at each level, so the probability of selection of a small school, P^{SS}_{klm}, then becomes:

$$P^{SS}{}_{klm} = (10/100)\left(\frac{MOS_{klm}}{MOS_{.lm}} \right)$$

For medium schools, one school was drawn from sub-sampled PSU, so the probability of selection of a medium school at each level, P^{MS}_{klm}, then becomes:

$$P^{MS}{}_{klm} = (12/100)\left(\frac{MOS_{klm}}{MOS_{.lm}} \right)$$

Note the additional sampling factor in the probability of selection for small schools and medium schools is due to the PSU sub-sampling for these classes of schools as noted above.

2.2.3 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 more than one component school, the selection of the component school at each grade is made with PPS sampling. The school selections from the 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.

2.2.4 Classes

Selection

In large schools, we selected an average of 1.80 classes per grade by selecting 2 classes per grade in selected large schools, and one class per grade in the remaining schools. The double class sampling took place in 80 randomly chosen schools 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. Non-response 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 randomly selected for doubling[6] then 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 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 (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}.

[6] Classes were selected for doubling independently with a uniform probability of 0.80

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 2012 survey instrument included a total of 81 questions, with the first five collecting student demographic information and the remaining measuring a comprehensive set of tobacco-related topics (Appendix A). Specifically, areas covered by the survey included prevalence of tobacco use, knowledge of and attitudes towards tobacco use, pro- and anti-tobacco media and advertising, minors' access to tobacco products, nicotine dependence, cessation attempts, exposure to second-hand smoke, harm perceptions, exposure to tobacco product warnings and tobacco use prevention school curricula.

3.2 Recruitment Procedures

The schools selected to participate in the 2012 NYTS fell across 39 different states. Recruitment began in May 2011 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 while also trying 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 20, 2012, immediately after data collector training and continued until June 8, 2012. Each data collector visited an average of 3 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) Post-survey meeting with the teacher or teachers; and (6) Post-survey meeting with the principal or lead contact prior to leaving the school. Most survey administrations could be completed in one day, while at other times, due to the number of classes selected or alternating block schedules, the data collector needed to return for a second day. Procedures were designed to protect students' privacy by assuring that student participation was anonymous and voluntary. Students completed a pencil and paper self-administered scannable questionnaire booklet. Students who could not complete the survey independently were excluded from the survey.

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 listservs. Data collector training was conducted on February 16–18, 2012. Initially, data collectors observed everything they would have to say or do as "experts" performed it. Then they acquired these skills through practice, demonstrated them to each other, and finally refined each other's performance through constructive feedback. Appendix B presents the training agenda from the 2012 NYTS Data Collector Training.

3.3.2 Field Procedures

After schools had been recruited, classes selected, and a date scheduled, each school received a packet of pre-survey materials. These materials included all the information necessary to prepare the school for data collection. Teacher packets contained the parental permission forms that had to be given out to all students in the selected classes prior to data collection. The timing of these pre-survey packet mailings was determined in part by the type of permission form being used by the school. Passive parental permission forms, or forms returned only if the parents do not want their child to participate, were sent approximately one week prior to the scheduled date of data collection. 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 two weeks prior to the scheduled date of data collection. 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 to get them where they need to be, 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 (e.g., 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 and implemented. 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. In a small number of schools, however, it was extremely difficult to develop an appropriate frame using this particular approach. Therefore, in such 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. Doing so is not ideal, though, as relatively few schools hold homeroom of duration sufficient enough for conducting the survey.

3.4 Participation Rates

Across the six previous cycles of the NYTS, the school participation has averaged 89%, with a low of 83%. Student participation has averaged 90% with a low of 88%. To be conservative, we assumed slightly lower values in developing the sample design for the 2011 NYTS: 85% for schools and 85% for students. The assumed overall participation rate was therefore 72.3%

The actual response rates differed from our projections: a school participation rate of 80.3% and a student participation rate of 91.7%. These participation rates were lower for schools but higher for student participation. The overall participation rate, the product of the school-level and student-level participation rates, at 73.6% was fairly close to the value assumed for the design.

Overall, the 2012 NYTS data file contains responses from 24,658 students compared with an anticipated 24,591 responding students overall. Exhibit 3-1 shows that student yields were slightly lower than targeted for Blacks, with a shortfall of 91 students and 255 students for high and middle-schools respectively. On the other hand, yields were much higher for Hispanics, with a yield of more than a thousand over targets for both middle and high schools.

Exhibit 3-1: Sample Yields for Black and Hispanic Students by School Level

Subgroup	Projected Participants	Actual Participants
Middle School Blacks	1,705	1,450
Middle School Hispanics	1,553	2,614
High School Blacks	1,946	1,855
High School Hispanics	1,724	3,098

Chapter 4—Weighting of NYTS Response Data

4.1 Overview

This section describes the procedures used to weight the data collected in NYTS 2012. The process involved the steps outlined below:

- Sampling weights
- Non-response adjustments
- Weight trimming
- Post-stratification 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 was weighted to reflect the initial probabilities of selection and non-response patterns, to mitigate large variations in sampling weights, and to post-stratify the data to known sampling frame characteristics.

4.2 Sampling Weights

The sampling weight attached to each student response is the inverse of the probability of selection for that student. This basic weight can be adjusted to compensate for non-response, to alleviate excess weight variation, and to match the weighted data to known control totals. A convenient way of computing the basic weight is by inverting the probabilities of selection at each stage, to derive a partial weight or stage weight. The stage weights are then multiplied together to form the overall weight.

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, a non-response adjustment and a ratio adjustment to control to known school enrollment totals.

This three step process is simplified algebraically[7] and computed directly as the ratio of the number of enrolled students to the number of responding students in a given weight class within a school. The weighting class definition is set dynamically so as to avoid extreme weights.

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 weight computation class, described below. This weight is computed as below, where N is the number of enrolled students[8] and R is the number of responding students in weight class *c* within a given school:

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

[7] The details of this are given in Appendix E – Student Weight Detail

[8] The student enrollment for each school used in this calculation is obtained from the school during data collection. These counts are obtained by grade and gender.

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 school.

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

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

The combination sequence first groups males and females within grade. Then both the cap and the weight are re-computed. 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 weight class.

This has the effect, within school, of setting an upper limit on the weight in class C of 2 in weight classes with an enrollment of less than 10, and 20% of the enrollment in weight classes with an enrollment of more than 10[9].

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:

$$W^{SS}{}_{klm} = (100/10)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{SS}{}_{klm}}$$

For medium schools the partial school weight is:

$$W^{MS}{}_{klm} = (100/12)\left(\frac{MOS_{.lm}}{MOS_{klm}}\right) = \frac{1}{P^{MS}{}_{klm}}$$

Appendix C gives, for each sampled PSU, the PSU size measure (over large schools) and the PSU-level sampling weight after removal of certainty PSUs. The overall weights for a given PSU, school and grade combination were the product of the adjusted PSU, school and grade level weights.

Appendix D gives the school measure of size and the school component of the sampling weight. Note that for schools identified as small schools, only one school was selected per sub-sampled PSU.

4.2.3 Grade Sampling Weights

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 $W^{G}{}_{jklm}$.

[9] The cap could be exceeded in cases where the weight class is collapsed to the school level.

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 schools and medium school selections, the enclosing PSU were drawn as a subsample. 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 non-response, trimmed, and post-stratified 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 schools, medium schools, and small schools respectively, where the weights in the right hand side of the equations are defined in the preceding sections.

4.3 Non-Response 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.

$$\begin{cases} A^{LS}_{m} = \dfrac{\displaystyle\sum_{k,l\ \varepsilon\ \text{large and medium, schools sampled}} W^{P}_{lm} W^{LS}_{klm} MOS^{LS}_{klm}}{\displaystyle\sum_{k,l\ \varepsilon\ \text{large and medium, schools } responding} W^{P}_{lm} W^{LS}_{klm} MOS^{LS}_{klm}} \\ A^{SS} = \dfrac{\displaystyle\sum_{k,l\ \varepsilon\ \text{small schools sampled}} W^{P}_{lm} W^{LS}_{klm} MOS^{LS}_{klm}}{\displaystyle\sum_{k,l\ \varepsilon\ \text{small schools with respondents}} W^{P}_{lm} W^{LS}_{klm} MOS^{LS}_{klm}} \end{cases}$$

The first equation applies to large and medium schools combined, and the second applies to small schools. Note that this adjustment is made within stratum for large and medium schools and across the whole sample for small schools. The student weight, adjusted for non-response, is $A^{SS}{}_{lm}\ W^{T1}{}_{hijklm}$ for small schools and $A^{LS}{}_{lm}$ $W^{T1}{}_{hijklm}$ for large and medium 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 minority concentrations. These weighting cells were created for computing non-response 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 non-response adjustment factors, are detailed in Exhibit 4-1. Note that the weighting classes are defined using collapsed sampling strata.

Exhibit 4-1: Large and Medium School Non-Response

School Level	Stratum (Non-Response)	Sampled Schools	Responding Schools	Percent Responding	Non-response Adjustment A^{LS}
HS	BR1	15	12	80.00%	1.22
HS	BR2	6	5	83.33%	1.21
HS	BR3	8	6	75.00%	1.31
HS	BU1	11	8	72.73%	1.0
HS	BU2	8	7	87.50%	1.15
HS	BU3	5	4	80.00%	1.25
HS	HR1	17	11	64.71%	1.52
HS	HR2	4	3	75.00%	1.34
HS	HR3	5	5	100.00%	1.00
HS	HU1	13	9	69.23%	1.46
HS	HU2	7	5	71.43%	1.40
HS	HU3	12	12	100.00%	1.00
HS TOTAL		111	87	78.38%	
MS	BR1	16	13	81.3%	1.26
MS	BR2	6	4	66.67%	1.53
MS	BR3	10	9	90.00%	1.09
MS	BU1	12	11	91.67%	1.08
MS	BU2	8	5	62.50%	1.60
MS	BU3	6	6	100.00%	1.00
MS	HR1	15	15	100.00%	1.00
MS	HR2	5	4	80.00%	1.13
MS	HR3	7	7	100.00%	1.00
MS	HU1	16	13	81.25%	1.23
MS	HU2	7	7	100.00%	1.00
MS	HU3	15	11	73.33%	1.35
MS TOTAL		113	105	92.92%	
GROSS TOTAL		234	192	82.05%	

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[10] is based on a similar procedure done for the National Assessment of Educational Progress (NAEP).

The trimming is an iterative procedure. During each iteration, an optimal weight, W_o[11] is calculated from the sum of the squared weights in the sample. Then, each weight W_i is 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; increasing it will reduce the level of trimming. For the current study, c has been set so that approximately 5% of the weight will be trimmed in the first iteration of the trimming algorithm.

The results of the first iteration of the trimming operation are summarized in Exhibit 4-2.

Exhibit 4-2: Results of First Trimming Iteration

					Coefficient of Variation (CVs in %)		Design Effects	
Trimming Class	Number Cases	Trimming Factor	Total Weight	Percent Trimmed	After Trimming	Before Trimming	After Trimming	Before Trimming
BR106	462	5.7	503697.8	4.92677	62.109	80.094	1.38492	1.64011
BR107	550	7.6	617116.2	4.94372	75.984	98.652	1.57631	1.97145
BR108	576	6.4	634453.2	4.96405	79.278	95.69	1.6274	1.91406
BR109	421	2.4	543295.3	4.82888	32.827	46.315	1.1075	1.21399
BR110	361	2.6	525183.9	4.97582	41.134	65.507	1.16874	1.42792
BR111	377	3.9	458311.4	4.98172	46.348	63.44	1.21424	1.4014
BR112	416	2.4	509942.5	4.84803	42.241	61.161	1.178	1.37317
BR206	145	1.9	248455.8	4.85364	26.818	38.816	1.07142	1.14963
BR207	157	1.8	206218.2	4.98079	23.117	36.182	1.0531	1.13008
BR208	169	1.8	198530.7	4.53634	50.303	56.334	1.25154	1.31547
BR209	150	3.7	238585.3	4.92402	39.32	56.431	1.15358	1.31633
BR210	153	3.2	228663.3	4.71459	42.249	55.734	1.17733	1.3086
BR211	215	1.3	207371.1	4.87046	15.318	26.457	1.02336	1.06967
BR212	170	8.7	199296	4.92372	66.74	89.851	1.4428	1.80257
BR306	224	2.2	207854.9	4.98635	51.873	66.13	1.26787	1.43536
BR307	257	11.4	233584	4.96529	116.162	135.567	2.3441	2.8307

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

[11] 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 non-response adjusted sampling weight described in the preceeding section. The subscripts *k* and *n* represent the number of iterations and the number of cases/weights respectively.

					Coefficient of Variation (CVs in %)		Design Effects	
Trimming Class	Number Cases	Trimming Factor	Total Weight	Percent Trimmed	After Trimming	Before Trimming	After Trimming	Before Trimming
BR308	279	6.9	236091.3	4.85471	89.524	107.137	1.79857	2.14372
BR309	147	6.3	275883.8	4.8954	73.075	89.639	1.53037	1.79805
BR310	214	4	229707.1	4.98715	50.353	67.039	1.25235	1.44732
BR311	217	3.1	235284.1	4.88506	57.01	69.309	1.32351	1.47817
BR312	215	1.6	170803.5	4.85203	31.013	40.491	1.09573	1.16319
BR406	58	1	16231.13	1.92715	0.088	3.997	1	1.00157
BR407	52	1	28855.34	4.92917	0.457	10.601	1.00002	1.01102
BR408	36	1.3	24217.88	3.89211	10.819	19.388	1.01138	1.03655
BR409	19	2.1	57268.5	4.04271	35.828	45.26	1.12161	1.19406
BR410	32	1.3	38724.41	4.15765	13.074	21.799	1.01656	1.04604
BR411	31	1	44178.56	1.7243	0.057	3.579	1	1.00124
BR412	29	1.2	52359.77	3.9821	9.585	17.813	1.00887	1.03064
BU106	413	1.7	403275.6	4.90102	38.596	46.135	1.1486	1.21233
BU107	433	2.1	364638.6	4.65029	43.359	51.297	1.18757	1.26253
BU108	422	2.4	350991.6	4.87755	47.186	57.435	1.22213	1.3291
BU109	280	2.1	413714.9	4.6411	40.618	49.12	1.16439	1.24042
BU110	256	2.1	457904.5	4.63113	43.012	52.423	1.18428	1.27375
BU111	322	2.1	407137.7	4.59435	45.712	56.576	1.20831	1.31909
BU112	283	2.5	363897.5	4.81333	42.464	55.474	1.17968	1.30665
BU206	203	2.5	402765.4	4.80507	60.685	68.832	1.36645	1.47146
BU207	243	5.7	520165.4	4.99358	109.212	120.467	2.18782	2.44525
BU208	226	15.6	531231.1	4.99463	107.147	133.678	2.14296	2.77907
BU209	263	1.6	267314.2	4.35642	38.621	47.147	1.14859	1.22144
BU210	285	1.5	260373.4	4.59539	29.443	39.285	1.08639	1.15379
BU211	283	7.1	244269.6	4.97214	62.218	83.508	1.38574	1.69489
BU212	258	1.8	221995	4.72904	36.872	46.286	1.13543	1.21341
BU306	135	3.4	176441.9	4.5478	85.631	92.667	1.72784	1.85236
BU307	146	2.6	180796.6	4.39031	78.934	83.801	1.61879	1.69745
BU308	114	1.5	170223.8	3.09225	41.438	43.84	1.1702	1.19051
BU309	92	1	92890.4	4.36393	1.116	11.565	1.00012	1.01323
BU310	77	1.9	91490.53	4.51353	19.128	31.775	1.03611	1.09965
BU311	65	1.2	84135.48	4.25414	11.652	19.381	1.01337	1.03698
BU312	64	2.8	78723.46	4.44223	65.358	74.642	1.4205	1.54844
BU406	21	1	759.74	3.68273	0.263	7.888	1.00001	1.00593
BU407	128	1.1	33501.4	4.03778	17.13	21.906	1.02911	1.04761
BU408	67	1	31335.81	4.64547	7.055	12.201	1.0049	1.01467
BU409	70	1.5	63681.79	3.68339	29.319	35.177	1.08473	1.12197
BU410	66	1.7	63555.04	4.76378	26.652	35.971	1.06996	1.12743
BU411	103	1.5	67918.03	3.22176	37.471	40.439	1.13904	1.16194
BU412	75	1.8	50221.65	4.9229	18.059	31.499	1.03218	1.09789
HR106	662	4.8	592697.2	4.93611	72.929	87.882	1.53106	1.77116
HR107	707	10.3	844934	4.9423	131.007	148.472	2.71386	3.20127
HR108	625	13.2	861104.6	4.90193	143.646	163.146	3.06013	3.65741
HR109	384	2.3	528893	4.54242	46.112	56.429	1.21207	1.31759
HR110	430	3.1	564194.7	4.86845	36.065	52.82	1.12977	1.27834
HR111	423	3.1	542369.5	4.96416	44.434	61.394	1.19697	1.37603
HR112	432	1.8	543630.1	4.5066	26.536	44.206	1.07025	1.19497
HR206	118	1.8	124743.6	4.89928	42.684	49.174	1.18064	1.23976
HR207	110	3.9	154582.3	4.9784	84.958	94.125	1.71523	1.8779
HR208	180	2.4	117008.5	4.57046	50.053	57.799	1.24914	1.33221
HR209	79	10.4	103493.5	4.97718	61.907	95.518	1.37839	1.90082
HR210	90	10.9	142615.7	4.86044	140.321	159.749	2.94712	3.5236
HR211	107	10.9	130614.5	4.95576	134.851	154.956	2.80148	3.3787
HR212	113	8.4	145874.3	4.84628	112.345	129.803	2.25096	2.66997
HR306	127	2.6	145920.8	4.84527	45.161	56.742	1.20235	1.31943
HR307	134	2	90330.27	4.72991	33.553	43.81	1.11174	1.1905

					Coefficient of Variation (CVs in %)		Design Effects	
Trimming Class	**Number Cases**	**Trimming Factor**	**Total Weight**	**Percent Trimmed**	**After Trimming**	**Before Trimming**	**After Trimming**	**Before Trimming**
HR308	143	2.2	112336	4.87864	37.396	47.325	1.13887	1.2224
HR309	149	1.5	109778.1	4.84794	17.027	28.749	1.0288	1.08209
HR310	131	1.1	99305.68	4.53661	9.709	17.88	1.00935	1.03172
HR311	164	1.3	98428.89	3.35482	15.972	21.865	1.02536	1.04752
HR312	158	1	96842.6	4.05278	3.625	10.299	1.00131	1.01054
HR406	56	1	42678.32	4.9489	11.307	16.374	1.01256	1.02633
HR407	58	1.8	58317.8	4.70435	43.571	49.63	1.18657	1.24206
HR408	76	2	59472.5	4.81871	18.106	33.594	1.03235	1.11137
HR409	78	1.6	103570.2	4.87661	34.593	44.145	1.11813	1.19238
HR410	67	2.7	36332.12	4.84478	54.577	63.846	1.29342	1.40155
HR411	64	4	29400.05	4.97424	68.83	80.911	1.46636	1.64442
HR412	41	2.9	32477.41	4.9347	82.796	93.381	1.6688	1.85073
HU106	433	3.3	362008.5	4.80882	67.208	76.742	1.45064	1.58757
HU107	521	10.9	672376.7	4.95037	134.416	153.932	2.80331	3.36497
HU108	403	7.1	669441.3	4.96663	101.557	115.886	2.02883	2.33962
HU109	372	1.9	366115	4.32104	43.228	49.412	1.18636	1.2435
HU110	269	2.5	365520	4.96237	54.225	63.586	1.29294	1.40282
HU111	364	3.8	405708.8	4.97717	55.033	69.68	1.30203	1.48419
HU112	351	1.6	373703.5	4.30803	38.294	43.801	1.14622	1.19131
HU206	302	1.6	219909.5	4.24698	21.078	30.635	1.04428	1.09354
HU207	273	2.1	202872.6	4.86398	35.741	48.866	1.12727	1.23791
HU208	275	2.4	195101.4	4.80557	36.029	49.95	1.12934	1.2486
HU209	283	2	335880.1	4.24756	36.59	44.32	1.13341	1.19573
HU210	125	3.7	235037.8	4.97368	46.633	62.055	1.21573	1.382
HU211	173	2.5	220355.2	4.99082	49.301	59.09	1.24166	1.34714
HU212	162	1.1	155395.8	3.87578	9.601	16.048	1.00916	1.0256
HU306	124	2.8	175403.4	4.69368	37.101	49.807	1.13654	1.24607
HU307	128	3.3	164790.4	4.62901	58.893	71.124	1.34413	1.50192
HU308	171	3.4	164009.1	4.94427	39.14	54.841	1.1523	1.299
HU309	263	2.5	245450.1	4.79569	46.365	57.121	1.21416	1.32503
HU310	259	4.5	250217	4.96694	54.915	72.677	1.3004	1.52615
HU311	268	2.3	230629.7	4.60523	44.369	53.438	1.19613	1.2845
HU312	264	3.4	257510.7	4.92166	69.247	80.227	1.4777	1.64119
HU406	180	2.1	162517.2	4.56364	61.636	66.561	1.37779	1.44057
HU407	177	3.2	223130.9	4.71036	65.747	75.79	1.42982	1.57116
HU408	193	2.4	219845.6	4.82259	42.522	54.586	1.17988	1.29642
HU409	224	2.2	232281.7	4.92646	67.563	74.045	1.45444	1.54582
HU410	319	3.6	202345.8	4.95649	63.114	74.312	1.39709	1.5505
HU411	186	4.8	191236.3	4.93318	83.43	95.517	1.69232	1.90744
HU412	165	5.9	119144.8	4.95203	47.473	72.106	1.22401	1.51677

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 zero otherwise.

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

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

4.5 Post-Stratification to National Student Population Estimates

To obtain accurate counts of high school students in schools considered eligible for the NYTS by grade, sex, and race for use in post-stratification, we turned to two school universe surveys conducted by the National Center for Education Statistics (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 the two sources. Private schools enrollments by grade and five racial/ethnic groups were obtained from the Private School Universe Survey (PSS), and public school enrollments by grade, sex, and five racial/ethnic categories were obtained from the Common Core of Data (CCD), both produced by the NCES. These databases were combined to produce the enrollments for all schools, and to develop population percentages to use as controls in the post-stratification 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).[12] 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 Private School Universe Survey (PSS), School Year 2008-09 (most recent PSS data); this file was also restricted to "regular" schools.

Exhibit 4-3 gives counts of schools and students by grade for public and non-public schools based on both the raw file and the resulting set of eligible schools. The latter set of student totals was used as control totals for the post-stratification adjustments described next.

Exhibit 4-3: Counts by Schools and Students by Type and Eligibility Status

		Raw (all)		Eligible	
Type	**Grade**	**Schools**	**Students**	**Schools**	**Students**
Public	6	38,697	3,718,686	35,103	3,692,442
	7	31,149	3,715,411	27,597	3,683,607
	8	32,728	3,695,081	27,659	3,654,361
	9	27,625	4,029,092	20,752	3,906,412
	10	26,423	3,817,457	19,529	3,678,799
	11	26,627	3,551,936	19,237	3,392,000
	12	25,899	3,475,170	19,058	3,258,714
	Total	209,148	26,002,833	168,935	25,266,335
Private	6	17,533	326,509	48,594	296,239
	7	16,607	324,403	46,231	294,840
	8	16,459	326,006	45,596	295,191
	9	8,719	304,056	22,957	272,669
	10	8,317	302,609	22,065	270,774
	11	7,959	297,630	21,331	266,308
	12	7,766	293,614	21,003	262,611
	Total	83,360	2,174,827	227,777	1,958,631
Total	6	56,230	4,045,195	83,697	3,988,681
	7	47,756	4,039,814	73,828	3,978,447

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

	8	49,187	4,021,087	73,255	3,949,552
	9	36,344	4,333,148	43,709	4,179,081
	10	34,740	4,120,066	41,594	3,949,573
	11	34,586	3,849,566	40,568	3,658,308
	12	33,665	3,768,784	40,061	3,521,325
	Total	292,508	28,177,660	396,712	27,224,966

Given a national estimate of student counts R_a and a weighted response total of P_a for post-stratification adjustment class "a", the post-stratification factor was the ratio of R_a to P_a. Exhibit 4-4 gives the population control totals used in post-stratification adjustments side by side with the sum of the weights in each post-stratum cell, as well as the adjustment factors calculated as the ratio of these two totals.

In other words, 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.

Post-stratification 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 4-4. Also within private schools, the racial groups were collapsed to preclude small numbers of students in the adjustment classes. For the public schools, five racial/ethnic categories were used: white, black, Hispanic, Asian/ Pacific Islander, and Native American.

Following post-stratification, the adjusted weights sum to the control population totals.

Exhibit 4-4: Post-Stratification Adjustments

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Post-Stratification Adjustment
Private	6	Combined	Combined	296,239	276,826	226	1.07013
Private	7	Combined	Combined	294,840	474,338	343	0.62158
Private	8	Combined	Combined	295,191	437,687	317	0.67443
Private	9	Combined	Combined	272,669	129,312	148	2.10862
Private	10	Combined	Combined	270,774	190,832	172	1.41891
Private	11	Combined	Combined	266,308	164,237	144	1.62148
Private	12	Combined	Combined	262,611	91,938	83	2.85640
Public	6	Asian and Pacific Islander	Female	83,721	105,883	112	0.79069
Public	6	Black	Female	292,946	242,781	255	1.20662
Public	6	Hispanic	Female	437,935	385,073	409	1.13728
Public	6	Native American	Female	22,491	53,546	50	0.42003
Public	6	White	Female	961,683	925,527	906	1.03907
Public	6	Asian and Pacific Islander	Male	86,036	93,100	96	0.92413
Public	6	Black	Male	305,397	240,613	243	1.26924
Public	6	Hispanic	Male	456,876	377,543	389	1.21013
Public	6	Native American	Male	23,385	62,481	62	0.37428
Public	6	White	Male	1,021,973	1,021,988	915	0.99999
Public	7	Asian and Pacific Islander	Female	82,688	115,294	109	0.71719
Public	7	Black	Female	292,547	281,909	273	1.03773
Public	7	Hispanic	Female	430,467	498,449	465	0.86361
Public	7	Native American	Female	22,515	61,749	58	0.36462
Public	7	White	Female	968,586	1,075,641	957	0.90047
Public	7	Asian and Pacific Islander	Male	85,752	167,708	124	0.51132

School Type	Grade	Race/Hispanic Origin	Sex	(E) Control Total	(F) Weighted Estimate	No. of Cases	(G) Post-Stratification Adjustment
Public	7	Black	Male	303,232	345,201	248	0.87842
Public	7	Hispanic	Male	449,832	491,194	488	0.91579
Public	7	Native American	Male	23,454	83,493	71	0.28091
Public	7	White	Male	1,024,534	1,001,234	938	1.02327
Public	8	Asian and Pacific Islander	Female	85,237	125,081	90	0.68146
Public	8	Black	Female	288,503	301,101	257	0.95816
Public	8	Hispanic	Female	422,787	462,040	468	0.91504
Public	8	Native American	Female	21,902	82,326	58	0.26604
Public	8	White	Female	968,999	993,115	930	0.97572
Public	8	Asian and Pacific Islander	Male	88,995	164,565	123	0.54079
Public	8	Black	Male	295,698	341,199	269	0.86664
Public	8	Hispanic	Male	437,069	512,195	467	0.85332
Public	8	Native American	Male	22,390	59,083	51	0.37895
Public	8	White	Male	1,022,781	1,097,002	925	0.93234
Public	9	Asian and Pacific Islander	Female	86,884	124,785	105	0.69627
Public	9	Black	Female	321,052	294,116	206	1.09158
Public	9	Hispanic	Female	441,655	511,929	496	0.86273
Public	9	Native American	Female	23,190	49,259	44	0.47077
Public	9	White	Female	1,016,693	937,680	766	1.08427
Public	9	Asian and Pacific Islander	Male	91,427	118,991	92	0.76836
Public	9	Black	Male	344,531	307,685	231	1.11975
Public	9	Hispanic	Male	472,449	450,865	397	1.04787
Public	9	Native American	Male	24,910	72,798	57	0.34218
Public	9	White	Male	1,083,620	980,677	732	1.10497
Public	10	Asian and Pacific Islander	Female	86,737	131,108	101	0.66157
Public	10	Black	Female	298,143	299,969	236	0.99392
Public	10	Hispanic	Female	399,937	388,364	429	1.02980
Public	10	Native American	Female	21,556	58,998	49	0.36537
Public	10	White	Female	997,689	929,670	742	1.07316
Public	10	Asian and Pacific Islander	Male	91,239	136,552	98	0.66816
Public	10	Black	Male	302,795	277,955	224	1.08937
Public	10	Hispanic	Male	414,312	394,775	363	1.04949
Public	10	Native American	Male	22,530	68,621	54	0.32833
Public	10	White	Male	1,043,860	914,328	666	1.14167
Public	11	Asian and Pacific Islander	Female	82,558	95,371	91	0.86565
Public	11	Black	Female	270,096	296,983	243	0.90947
Public	11	Hispanic	Female	351,558	361,172	391	0.97338
Public	11	Native American	Female	19,852	39,374	35	0.50419
Public	11	White	Female	956,051	924,499	836	1.03413
Public	11	Asian and Pacific Islander	Male	87,322	128,417	114	0.67998
Public	11	Black	Male	260,391	287,947	277	0.90430
Public	11	Hispanic	Male	354,423	349,350	365	1.01452
Public	11	Native American	Male	20,159	52,652	51	0.38288
Public	11	White	Male	989,590	897,347	815	1.10280
Public	12	Asian and Pacific Islander	Female	80,449	101,279	105	0.79433
Public	12	Black	Female	259,621	275,721	257	0.94161
Public	12	Hispanic	Female	324,651	360,944	370	0.89945
Public	12	Native American	Female	18,729	42,823	36	0.43736
Public	12	White	Female	942,787	861,573	723	1.09426
Public	12	Asian and Pacific Islander	Male	84,213	109,462	113	0.76933
Public	12	Black	Male	239,912	229,075	229	1.04731
Public	12	Hispanic	Male	318,412	328,587	353	0.96904
Public	12	Native American	Male	18,896	41,278	40	0.45777
Public	12	White	Male	971,044	929,139	887	1.04510

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 2012 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 non-response at the school level. In particular, non-participating schools may lead to PSUs without student data for a given school level.

Due to nonresponse an entire PSU dropped out; therefore, all strata but one had at least two PSUs. Stratum BR3 ended up with only one PSU, so it needed to be collapsed with another stratum for analytic purposes (variance estimation). Specifically, stratum BR3 was combined with BR4 to result in an analysis stratum with at least two PSUs (analysis stratum coded 103).

Exhibit 4-5 displays the correspondence between the sampling strata and the analysis strata, which are represented by two variables on the analysis file. 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 4-5: Sampling and Analysis Stratum Coding Schemes

High 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	204
BU1	111	HU1	211
BU2	112	HU2	212
BU3	113	HU3	213
BU4	114	HU4	214

Exhibit 4-6 presents 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 standard error for prevalence estimates of current use and ever use of tobacco products separately for high schools (4-7a) and middle schools (4-7b).

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

Example: Estimates, Ever Use by School Type

SAS:

```
Proc Surveymeans Data=NYTS2012 mean;
Var etob ecigt ecigar eslt epipe ebidis ekreteks;
Class etob ecigt ecigar eslt epipe ebidis ekreteks;
Stratum v_stratum;
Cluster psu;
Weight finwt;
Domain Schooltype  Schooltype*Sex  Schooltype*Race;
Title "NYTS 2012, Estimates by School Type, by School Type and Sex Cross-Classified, and by School Type and Race Cross-Classified ";
run;
```

SUDAAN:

```
Proc Descript Data="C:\NYTS2012.ssd" Filetype= SAS Design=WR;

Var etob ecigt ecigar eslt epipe ebidis ekreteks;
Catlevel      1      1      1      1      1      1      1;
Nest Stratum2 PSU2/Missunit;
Weight finwt;
Subgroup Schoolty  Sex   Race ;
Levels       2       2     4
Tables Schoolty  Schoolty*Sex  Schoolty*Race;
Title "NYTS 2012, Estimates by School Type, by School Type and Sex Cross-Classified, and by School Type and Race Cross-Classified ";
Print Percent Sepercent / Style=NCHS;
```

Exhibit 4-6a: Current and Ever Use Estimates for High School Students

Ever Use—High School			Sex		Race/ethnicity		
Product	**Estimate**	**Overall**	**Female**	**Male**	**Whites**	**Blacks**	**Hispanics**
Any tobacco	Prevalence	45.7%	40.8%	50.4%	44.7%	50.5%	47.9%
	Standard Error	1.39%	1.76%	1.31%	1.76%	2.35%	2.08%
Cigarettes	Prevalence	36.1%	33.3%	38.9%	35.7%	35.4%	39.8%
	Standard Error	1.43%	1.67%	1.40%	1.91%	2.27%	2.06%
Cigar	Prevalence	31.3%	25.2%	37.2%	30.4%	39.2%	32.0%
	Standard Error	1.10%	1.37%	1.16%	1.37%	2.42%	1.57%
SLT	Prevalence	13.6%	5.4%	21.6%	17.6%	4.8%	10.0%
	Standard Error	0.96%	0.59%	1.44%	1.28%	0.85%	1.12%
Pipe	Prevalence	10.3%	7.5%	13.1%	11.1%	5.5%	12.4%
	Standard Error	0.56%	0.55%	0.75%	0.73%	0.76%	0.86%
Bidis	Prevalence	2.5%	1.6%	3.4%	2.6%	1.5%	3.0%
	Standard Error	0.21%	0.18%	0.36%	0.30%	0.33%	0.41%
Kreteks	Prevalence	3.1%	2.0%	4.2%	3.9%	0.8%	2.8%
	Standard Error	0.28%	0.25%	0.41%	0.41%	0.23%	0.41%

Current Use – High School			Sex		Race/ethnicity		
Product	**Estimate**	**Overall**	**Female**	**Male**	**Whites**	**Blacks**	**Hispanics**
Any tobacco	Prevalence	22.68	16.95	28.24	23.76	22.64	21.50
	Standard Error	0.94	1.01	1.15	1.22	1.64	1.45
Cigarettes	Prevalence	14.02	11.68	16.28	15.37	9.60	14.26
	Standard Error	0.82	0.82	0.96	1.16	1.11	1.25
Cigar	Prevalence	12.62	8.40	16.69	12.20	16.69	12.39
	Standard Error	0.63	0.66	0.87	0.77	1.27	0.97
SLT	Prevalence	6.42	1.52	11.16	8.10	2.17	5.10
	Standard Error	0.52	0.25	0.89	0.68	0.42	0.74
Pipe	Prevalence	4.55	3.24	5.82	4.53	2.89	6.21
	Standard Error	0.30	0.31	0.43	0.39	0.66	0.56

Bidis	Prevalence	0.90	0.49	1.30	0.74	0.83	1.39
	Standard Error	0.10	0.11	0.20	0.11	0.30	0.32
Kreteks	Prevalence	0.96	0.46	1.45	1.09	0.60	0.93
	Standard Error	0.12	0.10	0.20	0.16	0.18	0.21

Exhibit 4-6b: Current and Ever Use Estimates for Middle School Students

Ever Use—Middle School			Sex		Race/ethnicity		
Product	**Estimate**	**Overall**	**Female**	**Male**	**Whites**	**Blacks**	**Hispanics**
Any tobacco	Prevalence	18.2%	15.9%	20.5%	15.1%	23.6%	25.2%
	Standard Error	1.07%	1.29%	1.14%	1.16%	2.29%	1.64%
Cigarettes	Prevalence	13.6%	12.7%	14.4%	11.3%	16.2%	19.5%
	Standard Error	0.89%	1.02%	0.98%	0.99%	1.49%	1.54%
Cigar	Prevalence	8.1%	6.6%	9.5%	5.8%	12.3%	12.3%
	Standard Error	0.60%	0.82%	0.66%	0.52%	1.86%	1.08%
SLT	Prevalence	4.4%	2.9%	5.8%	4.7%	3.2%	4.4%
	Standard Error	0.44%	0.38%	0.67%	0.62%	0.70%	0.63%
Pipe	Prevalence	3.8%	3.1%	4.5%	2.8%	2.4%	7.3%
	Standard Error	0.31%	0.37%	0.42%	0.28%	0.48%	0.81%
Bidis	Prevalence	1.1%	1.0%	1.3%	0.9%	0.9%	1.9%
	Standard Error	0.13%	0.18%	0.19%	0.15%	0.24%	0.33%
Kreteks	Prevalence	0.9%	0.8%	1.1%	0.6%	1.7%	1.3%
	Standard Error	0.14%	0.23%	0.13%	0.09%	0.68%	0.27%

Current Use – Middle School			Sex		Race/ethnicity		
Product	**Estimate**	**Overall**	**Female**	**Male**	**Whites**	**Blacks**	**Hispanics**
Any tobacco	Prevalence	6.46	5.38	7.52	4.93	7.67	10.07
	Standard Error	0.49	0.55	0.54	0.54	0.97	0.99
Cigarettes	Prevalence	3.48	3.18	3.78	3.06	2.59	5.44
	Standard Error	0.36	0.37	0.42	0.40	0.57	0.74
Cigar	Prevalence	2.84	2.44	3.22	1.59	4.99	4.93
	Standard Error	0.25	0.33	0.29	0.20	0.70	0.65
SLT	Prevalence	1.68	1.15	2.20	1.57	0.60	2.39
	Standard Error	0.20	0.19	0.30	0.26	0.23	0.45
Pipe	Prevalence	1.79	1.72	1.86	1.20	1.19	3.74
	Standard Error	0.22	0.27	0.25	0.21	0.38	0.60
Bidis	Prevalence	0.58	0.41	0.75	0.34	0.61	1.21
	Standard Error	0.07	0.10	0.12	0.08	0.16	0.25
Kreteks	Prevalence	0.50	0.42	0.58	0.31	0.23	1.04
	Standard Error	0.07	0.10	0.12	0.08	0.13	0.25

Appendix A
Questionnaire

Insert 2-column, 11-page questionnaire here at PDF Stage.

Appendix B
Data Collector Training Agenda

2012 National Youth Tobacco Survey (NYTS) Data Collector Training Agenda

Calverton, Maryland—February 16-18, 2012

TRAINING LEADERS

René Arrazola
Office of Smoking and Health (OSH)
Centers for Disease Control and Prevention (CDC)

Kate Flint, Rená Agee, Amy Hughes, Sophia Stringfellow, Kevin Lacy
ICF International

Thursday, February 16, 2012

Module 1 (8:30–9:00)	Introduction and Orientation
Module 2 (9:00–9:15)	Overview of Training
Module 3 (9:15–9:45)	Office on Smoking and Health and the Youth Tobacco Survey
Module 4 (9:45–10:00)	Parental Permission
Module 5 (10:00–10:15)	Roles and Responsibilities
Module 6 (10:15–10:30)	Steps Leading to Data Collection
Module 7 (10:30–10:40)	Break
Module 8 (10:40–10:55)	Introduction to the Case Management System
Module 9 (10:55–11:15)	Overview of the Data Collection Process
Module 10 (11:15–11:35)	Receiving Assignments
Module 11 (11:35–12:00)	Starting the Data Collection Process—Advance Call to Schools
Module 12 (12:00–12:30)	Preparing Your Materials
Module 13 (12:30–1: 30)	Lunch
Module 14 (1:30–1:50)	Entry Meeting with Principal/Contact Person
Module 15 (1:50–2:15)	Teacher Meeting Before Survey
Module 16 (2:15–3:00)	Survey Administration
Module 17 (3:00–3:45)	Classroom Forms/Essential Paperwork
Module 18 (3:45–4:05)	Exit Meeting with Teachers
Module 19 (4:05–4:30)	Exit Meeting with Principal/Contact Person
Module 20 (4:30–5:00)	Priority Areas Across the Data Collection Process
Module 21 (5:00–5:15)	Day 1 Wrap Up and Quick Look at Day 2

Friday, February 17, 2012

Module 22 (9:00–9:30)	Review of Day 1—Q&A
Module 23 (9:30–10:30)	Dealing with Difficult Situations Before/During/After the Data Collection Process
Module 24 (10:30–10:45)	Distribution and Review of Materials Needed for Calls to Week 1 and 2 Schools
Module 25 (10:45–11:15)	Telephone Calls to Week 1 and 2 Schools
Module 26 (11:15–11:45)	Debriefing on Calls to Schools
Module 27 (11:45–12:15)	Unpacking and Checking Your Data
Module 28 (12:15–1:15)	Lunch
Module 29 (1:15–2:00)	Student Participation Rates and the CMS
Module 30 (2:00–2:30)	Employment Forms, Administrative Procedures, and Expense Reimbursement Procedures
Module 31 (2:30–3:00)	Expense Report Problems and Procedures
Module 32 (3:00–4:00)	Day 2 Wrap up, Role Play/Simulation Assignments, and Practice
OPTIONAL: (4:00–4:30)	Open Skill Building Session

Saturday, February 18, 2012

Module 33 (9:00–9:15)	Review of Day 2—Q&A
Module 34 (9:15–11:00)	Role Plays/Simulations—Advance Call, Entry Meeting with Principal/Contact Person, and Entry Meeting with Teacher
Module 35 (11:00–11:15)	Break
Module 36 (11:15–12:30)	Continued Role Plays/Simulations—Survey Administration, Exit Meeting with Teachers, and Exit Meeting with Principal/Contact Person
Module 37 (12:30–1:30)	Lunch
Module 38 (1:30–2:00)	Wrap Up Role Plays/Simulations
Module 39 (2:00–2:30)	Distribution and Return of Project Materials
Module 40 (2:30–2:45)	Break
Module 41 (2:45–3:15)	Areas for Priority Attention
Module 42 (3:15–3:30)	Day 3 Wrap Up, Q & A, and Closure

Appendix C
Student Weight Detail

Student Weight Development Details

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 – as are the other stage sampling weights – followed by two adjustments, one for non-response, and another post-stratifying to known enrollment totals.

Note that in this section, for the sake of clarity, we omit the subscripts denoting the sampling stages and weight class. The un-subscripted quantities given are taken 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, and is 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:

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

Non-Response Adjustment

The non-response 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;

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

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 non-response 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

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

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:

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

Thus confirming the use of the simplified formula given in Section 4.2.1.

Appendix D Common Core of Data Race/Ethnicity Definitions

American Indian or Alaska Native—A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

Asian—A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Native Hawaiian or Other Pacific Islander— A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

Black or African American— A person having origins in any of the black racial groups of Africa.

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

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

PRINT ONLY