METHODOLOGY REPORT OF THE 2022 NATIONAL YOUTH TOBACCO SURVEY

For questions about this report, please email Dr. Sean Hu at fik4@cdc.gov

Centers for Disease Control and Prevention

Office on Smoking and Health

Atlanta, GA

February 2023

TABLE OF CONTENTS

CHAPTER 1—NYTS SAMPLING DESIGN

1.1   OVERVIEW OF THE NATIONAL YOUTH TOBACCO SURVEY (NYTS)

Tobacco product indicators included in the NYTS are: tobacco product use (e.g., electronic cigarettes, cigarettes, cigars [including cigars, little cigars, and cigarillos], smokeless tobacco [chewing tobacco, snuff, or dip; snus, dissolvable tobacco products], hookahs, pipe tobacco, bidis, roll-your-own cigarettes, heated tobacco products, and nicotine pouches); exposure to secondhand smoke and e-cigarette aerosol; smoking cessation; minors' access to tobacco products; knowledge and attitudes about tobacco.

1.2   OVERVIEW OF THE 2022 NYTS METHODOLOGY

The 2022 NYTS was conducted using a stratified, three-stage cluster sample design to produce a nationally representative sample of middle school and high school students in the United States. Sampling procedures were probabilistic and conducted without replacement at all stages. Sampling entailed selection of (1) Primary Sampling Units (PSUs) (defined as a county, or a group of small counties, or part of a very large county) within each stratum; (2) Secondary Sampling Units (SSUs) (defined as schools or linked schools) within each selected PSU; and (3) students within each selected school.

The 2022 NYTS was administrated as a web-based survey. Students participated in the survey while at school, home, or some other location. Using a school-issued or personal internet-connected device, students logged into a secure website and watched a brief 2-minute instructional video before completing the survey.

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

Survey administration began in January and concluded in May 2022. The final NYTS sample consisted of 574 schools, of which 341 participated, yielding a school participation rate of 59.4%. A total of 28,291 student questionnaires were completed out of a sample of 37,172 students, yielding a student participation rate of 76.1%. The overall participation rate was 45.2%.

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.

CHAPTER 2—NYTS SAMPLING METHODS

2.1   SAMPLE DESIGN

The NYTS methodology was designed to produce national estimates at a 95% confidence level by school level (middle school and high school), by grade (6, 7, 8, 9, 10, 11, and 12), by sex (male and female), and by race and ethnicity (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, and non-Hispanic American Indian/Alaska Native; Appendix B). The sampling design prohibits subnational analyses.

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

The sample was a stratified, three-stage cluster sample. PSUs were stratified by racial/ethnic status and urban versus nonurban. PSUs were classified as "urban" if they were in one of the 54 largest U.S. Metropolitan Statistical Areas (MSAs); otherwise, they were classified as "nonurban." Within each stratum, PSUs were chosen without replacement. Table 2.1 presents key sampling design features.

Table 2.1   Key Sampling Design Features

Sampling Stage	Sampling Units	Stratification	Measure of Size (MOS)	Designed Sample Size
1	PSUs: Counties, portions of a county, or groups of counties	Urban vs. Nonurban (2 strata) Minority concentration (8 strata)	Aggregate school size in target grades	100 Counties, portions of a county, or groups of counties
2	Schools	Small, medium, and large High school vs middle school	Aggregate eligible enrollment	320 SSU (school) selections*: 240 large schools, 50 medium schools, and 30 small schools
3	Classes/students			2 Classes per grade in half of large schools; 1 class per grade otherwise 20,600 student participants

In this exhibit, the schools are SSUs or "virtual schools" created by combining actual, physical schools so that each virtual school unit has a complete set of grades for the level. The virtual schools are expanded to physical schools. The number of physical schools in the sample was expected to range from 345 to 375.

The first stage of sampling selected PSUs within each stratum for a total of 100 sample PSUs. At the second sampling stage, a total of 320 SSUs, or schools, were selected from the sample PSUs as

follows: two large schools were selected per sample PSU, one per level (middle or high); an additional large school for each level was selected in a subsample of 40 PSUs, for a total of 240 large SSUs. An additional 50 medium SSUs and 30 small SSUs were selected from subsample PSUs, for a total of 320 sample SSUs (320 = 240 + 30 + 50). The PSU subsamples were selected with simple random sampling, and the schools were drawn with probability proportional to the total number of eligible students enrolled in a school.

Depending on the average design effects, target subgroup sample sizes are between 1,200 and 1,700. Compared to previous cycles, the NYTS sampling design has had both lower effects on unequal weighting and smaller clustering effects. These factors lead to lower design effects, particularly for subgroups. Smaller design effects have led to smaller variances and improved precision.

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

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

2.2   SAMPLING FRAME

The 2022 NYTS sample was based on a sampling frame from multiple data sources to increase the coverage of schools nationally. The frame combined data files obtained from MDR Inc. (Market Data Retrieval Inc.) and from the National Center for Education Statistics (NCES). The MDR frame contained school information that included enrollments, grades, race and ethnicity distributions within the school, district and county information, and other contact information for public and nonpublic schools across the nation. The NCES frame sources included the Common Core of Data for public schools and the Private School Survey for nonpublic schools. Including schools sourced from the two NCES files resulted in substantial coverage increase among all public and nonpublic high schools. Most of the added schools were smaller schools. Each school was represented only once in the final sampling frame.

The first step was to remove schools such as Department of Defense schools, vocational schools, and adult education schools. This resulted in the exclusion of 3.9% of schools (2.8% of public schools and 8.0% of private schools) and 1.1% of students. Lastly, schools were removed that had fewer than 40 students enrolled across eligible grades, resulting in the exclusion of 20.4% of schools (12.8% public and 42.6% private) which had been eligible after the other exclusions. This exclusion of schools with fewer than 40 students led to the exclusion of only 1.03% of students of those in eligible schools. Overall, 97.8% of students in middle and high schools nationally were included in the frame. The frame contained 28,636 high schools and 42,749 middle schools for a total of 71,385 eligible schools.

2.3 SAMPLING UNITS AND MEASURE OF SIZE

2.3.1 Sample

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

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

The sampling stages may be summarized as follows:

- Selection of PSUs—One hundred (from approximately 1,257) PSUs were selected from 16 strata with probability proportional to the total number of eligible students enrolled in all eligible schools located within a PSU.

- Selection of schools—At the second sampling stage, a total of 240 large schools, or SSUs, were selected from the sample PSUs. Additionally, as described in Section 2.1, we selected 50 medium schools and 30 small schools, resulting in a total of 320 sample SSUs (320 = 240 + 50 + 30).

- Selection of students—Students were selected via whole classes whereby all students enrolled in any one selected class were chosen for participation. Classes were selected from course schedules provided by each school so that all eligible students had only a single chance of selection.

The sampling approach used probability proportional to size (PPS) sampling methods, with the measure of size (MOS) defined as the count of final-stage sampling units—students in intact classrooms. Coupled with the selection of a fixed number of units, the design resulted in an equal probability of selection for all members of the universe (i.e., a self-weighting sample). These conditions were approximated for the NYTS resulting in a roughly self-weighting sample. The MOS also was used to compute stratum sizes and PSU sizes. By assigning an aggregate measure of size to the PSU, the sample allocated to the PSU was in proportion to the student population.

The third, and final, sampling stage selected classes within each grade of a sample SSU. We selected two classes per grade in large schools and one class per grade in the remaining schools. The threshold for double class sampling was based on a simulation study to ensure that the required numbers of students in specified racial and ethnic minority groups were achieved per school level. All students in a selected class were eligible to participate in the survey.

2.4 PLANNED SAMPLE SIZES FOR THE SAMPLE

In calculating the sample sizes for the 2022 NYTS, we made our approach more robust by assuming a conservative final rate (student x school) of 60%. The student participation rate was adjusted to account both for a growing number of ineligible students and parental refusal and for the new data collection methods (i.e., 100% virtually supported fielding methodology without in-person survey administrators).

Table 2.4   Planned Sample Sizes for the 2022 NYTS, Sample

PSU	Size	# of SSUs	Number of Schools Sampled	Number of Classes per School	Number of Students per Class	Number of Sampled Students Prior to Attrition	No. of Students After 75% School RR	No. of Students After 60% Final School and Student RR
100	Large High School	120	Double classes: 60	8	24	11520	8640	6912
			Single classes: 60	4	24	5760	4320	3456
	Large Middle School	120	Double classes: 60	6	22	7920	5940	4752
			Single classes: 60	3	22	3960	2970	2376
	Large Total	**240**				**29160**	**21870**	**17496**
25 (subsample)	Medium High School	25		4	20	2000	1500	1200
	Medium Middle School	25		3	20	1500	1125	900
	Medium Total	**50**				**3500**	**2625**	**2100**
15 (subsample)	Small High School	15		4	16	960	720	576
	Small Middle School	15		3	16	720	540	432
	Small Total	**30**				**1680**	**1260**	**1008**
	Overall Total	**320**				**34340**	**25755**	**20604**

The estimated sample yield from these large schools was 29,160 students before school and student nonresponse, leading to an expected total of 17,496 participating students in large schools after accounting for nonresponse. The expected yield was 3,500 from medium schools and 1,680 students from small schools. In total, the expected number of participating students was 20,604.

Within each school, one class was selected from each grade to participate in the survey except in large schools with high racial and ethnic minority populations, where two classes per grade were selected. Note that the set of schools with high racial and ethnic minority populations defined for double class sampling is a subset of the large schools that can support such double class sampling. For the 2022 NYTS, we implemented double class selection for half of large schools in the primary sample (randomly selected) to ensure sufficient student yields.

2.5 SAMPLING UNITS

2.5.1 Sampling Units (PSUs)

In defining PSUs, several issues were considered:

- Each PSU should be large enough to contain the requisite numbers of schools and students by grade, and small enough so as not to be selected with near certainty.
- Each PSU should be compact geographically to control the number of school districts contacted and recruited.
- Recent data should be available to characterize each PSU.
- PSUs are defined as containing at least four middle and five high schools.

Generally, counties were equivalent to PSUs, with two exceptions:
- Low population counties were grouped to provide sufficient numbers of schools and students.
- High population counties were divided into multiple PSUs so that the resulting PSUs would not be selected with certainty.

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

2.5.2 Forming Secondary Sampling Units (SSUs)

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

2.6 STRATIFICATION

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

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

- Non-Hispanic Black urban and non-Hispanic Black nonurban PSUs were also classified into four groupings depending upon the percentages of Black students in the PSU.

The density grouping bounds were computed using an optimization algorithm[1] that was refreshed each cycle to reflect changes in the racial/ethnic distribution of the student population. The boundaries or cutoffs changed as the frequency distribution ("*f*") for the racial and ethnic groupings changed from one survey cycle to the next. Table 2.6 presents the stratum boundaries used in the 2022 NYTS.

Table 2.6   Stratum Boundaries: Minority Percentage Cutoffs

Minority Concentration	Density Group	Bounds	
		Urban	Nonurban
Black	1	0%–26%	0%–20%
	2	>26%–40%	>20%–34%
	3	>40%–54%	>34%–54%
	4	>54%–100%	>54%–100%
Hispanic	1	0%–26%	0%–24%
	2	>26%–42%	>24%–48%
	3	>42%–58%	>48%–68%
	4	>58%–100%	>68%–100%

As described earlier, SSUs were stratified into three sizes for small, medium, and large schools for the primary sample only. For the supplement samples, the frame was restricted to large schools which support double class sampling at every grade. The two supplement samples were not stratified explicitly but only implicitly by region and by state. Specifically, the frame was sorted by region and by state for PPS selection of PSUs. Implicit stratification helps improve the geographic representation of the supplement samples.

---

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

CHAPTER 3—NYTS DATA COLLECTION AND PROCESSING

3.1   SURVEY INSTRUMENT

The NYTS collects data on key tobacco product prevention and control outcome indicators. The 2022 survey instrument included 166 questions. The web survey was created using ColdFusion and all data were stored in a MS SQL Server. To take the web survey, students navigated to a dedicated URL, nyts.cdc.gov, and entered a randomly generated, five-digit access code.

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

The first five questions on the survey collected student demographic information, and the rest measured a comprehensive set of tobacco-related topics. Specific areas covered by the survey included: prevalence of tobacco product use; knowledge of and attitudes toward tobacco product use; exposure to tobacco media and advertising; minors' access to tobacco products; nicotine dependence; cessation attempts; exposure to secondhand smoke; harm perceptions; and exposure to tobacco product warnings. At the beginning of each tobacco product section, a description of the product (with example brands) and generic images of specific tobacco products were provided to assist with product recognition and increase the accuracy of student data. Students could refer to this description and the images as they answered related questions. The NYTS also included sociodemographic questions about family affluence, depression and anxiety, and sexual orientation and gender identity (SOGI).

3.2   EXTERNAL REVIEW AND APPROVALS

Three bodies reviewed and approved the instrumentation, processes, privacy and security elements, and sampling design of the 2022 NYTS: the Office of Management and Budget (OMB), ICF's Institutional Review Board (IRB), and CDC's Institutional Review Board (IRB).

With the transition to an electronic data collection format for the 2019 NYTS, the Security Assessment and Authorization (SA&A) approval and Enterprise Performance Life Cycle (EPLC) review was valid for the 2022 NYTS cycle. The SA&A is a formal methodology for testing and evaluating the security controls of the system to ensure that it is configured properly to meet the security mandated by the Federal Information Security Management Act (FISMA). EPLC is a framework to enhance the Department of Health and Human Services (HHS) IT governance through rigorous application of sound investment and project management principals, in conjunction with industry's best practices.

3.3   TECHNICAL ASSISTANCE PROVIDER (TAPs) STAFFING

The role of the Technical Assistance Provider (TAP) was developed for the 2021 NYTS in response to anticipated complications due to COVID-19 that prohibited data collectors from conducting in-person survey administration. This role continued for the 2022 NYTS cycle, as

well. TAPs provided 100% virtual support to schools and teachers before, during, and after survey administration to (1) ensure teachers had received all the necessary materials to administer the survey, (2) answer any questions schools contacts and/or teachers may have prior to, during, or after survey administration, (3) ascertain that parental consent was properly obtained prior to the scheduled survey administration date, and (4) provide remote IT support, if needed. To ensure schools in various time zones would be adequately supported during school hours, TAPs were hired geographically across the country so that every time zone with sampled schools had at least one TAP in that part of the country. TAPs were recruited from a pool of previously trained data collectors. An in-person, 2-day training for TAPs was conducted December 6–7, 2021.

Key components of the training included the following:

- Pre- and postsurvey communications with the schools and teachers
- Orientation to student and teacher portals
- IT troubleshooting
- Communication with headquarters staff

3.4   RECRUITMENT PROCEDURES

Recruitment began in September 2021 with calls to state departments of education and health, informing them of the survey effort and sampled schools in their state. After notification at the state level, district- and school-level recruitment began. For public or diocesan schools, verbal or written agreement was first obtained by their district or diocese, respectively, before contact was made with the school. However, private schools were approached directly. A date for survey implementation was selected that was convenient to the school. Recruiters and TAPs used a secure web-based calendar to facilitate communication and adjust survey dates upon request by the school.

3.5   SURVEY ADMINISTRATION

Survey administration began in January and continued through May 2022. While the details of each data collection varied, there were six core steps followed for every school:

1) Conduct precontact call with the principal or lead contact to confirm survey arrangements and to answer any questions.
2) Send tailored communications and survey materials to selected teachers.
3) Reach out directly to teachers to confirm receipt of materials, verify intentions to administer the survey on the scheduled date, confirm parental consent procedures were followed, and provide additional survey instructions.
4) Virtually monitor survey activities and respond to requests for technical support, as needed.
5) Follow up with teachers regarding student response rates and class enrollment.
6) Report final progress to school contact and thank them for their school's participation.

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

3.5.1 Field Procedures

After schools had been recruited, classes selected, and a date for survey administration scheduled, each school received a mailing with presurvey materials containing instructions for the school contact and packets for the teacher of each selected class. Teacher packets contained the parental permission forms to be distributed to all students in the selected classes prior to data collection. The timing of these presurvey materials was determined in part by the type of permission form being used by the school; this decision was made by the school district or individual school. Opt-out parental permission forms (i.e., forms returned only if the parents do not want their child to participate) were sent approximately 2 weeks prior to the scheduled date of data collection in the majority of schools. Active parental permission forms (i.e., forms that must be returned with the parent's signature for the child to participate) were sent out 4 weeks prior to the scheduled date of data collection for schools that require active consent. TAPs conducted follow-up calls and sent emails to the selected schools to answer any questions and to make sure materials were received and distributed to selected classes and students.

3.5.2 Classroom Selection

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

3.6 WEB-BASED DATA COLLECTION MANAGEMENT APPLICATION (DCMA)

For multiple cycles of the NYTS, a web-based data collection management application (DCMA) has been used to help: centralize the management of the study; facilitate information exchange with project staff; and allow all members of the project management teams, recruitment teams, supervisory teams, and remote staff access to information necessary to implement the study. The system is designed with differing levels of access depending on the user's role on the study. The system's primary functions include generating invitation letters, tracking recruitment progress,

scheduling data collection, registering student records submitted to the central repository, and tracking school and student response rates.

3.7 DATA RECORDING

Preliminary student participation rates were calculated based on (1) class enrollment numbers from teachers of selected classes and (2) the number of surveys received in the central repository. If teachers reported a different number of expected completes than what was received in the central repository, a TAP followed up to resolve discrepancies and determine ways to maximize student participation. As additional surveys were received after the initial survey administration date, the DCMA automatically updated the number of records received; participation reporting was revised accordingly.

3.8 PARTICIPATION RATES

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

3.8.1 School-Level Participation Rates

The sample includes 574 schools that were selected across 243 districts in 41 states and the District of Columbia. During sample validation, 34 schools were deemed ineligible and were replaced. In total, 341 schools (59.4%) participated in the study. Of refusals, 145 of them were due to district-level refusals to allow contact with schools to discuss participation, and 88 were school-level refusals.

3.8.2 Student-Level Participation Rates

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

The final student participation rate for the 2022 NYTS sample was 76.1%. Overall, 37,172 eligible students from the 341 participating schools were invited to participate in the survey, and 28,291 did so.

Table 3.1   Overall NYTS 2022 Student Participation Rate

	# Eligible	# Completed	Participation %
Final Sample	37,172	28,291	76.1%

The 2022 NYTS final sample attained an actual school participation rate of 59.4% and a student participation rate of 76.1%. The overall participation rate was 45.2% for the final sample.

3.9   DATA MANAGEMENT

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

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

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

CHAPTER 4—WEIGHTING OF NYTS RESPONSE DATA

4.1   ESTIMATORS AND VARIANCE ESTIMATION

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

These estimates are accompanied by measures of sampling variability, or sampling error, such as variances and standard errors, that account for the complex sampling design. These measures support the construction of confidence intervals and other statistical inference such as statistical testing (e.g., subgroup comparisons or trends over successive NYTS cycles). Sampling variances can be estimated using the method of general linearized estimators[2] as implemented in SAS survey procedures. These software packages must be used because they permit estimation of sampling variances for multistage stratified sampling designs. They also account for unequal weighting and for sample clustering and stratification.

The final weight files also include PSU and strata variables which support the analysis of clustered survey data and accurate variance estimation. As in previous cycles, a variable for "variance strata" was added, which may differ from the design strata, to ensure that all variance strata had at least two PSUs.[3]

---

[2] Skinner CJ, Holt D, Smith TMF. *Analysis of Complex Surveys*. John Wiley & Sons; 1989: 50.

[3] Specifically, two strata were combined into one variance stratum because the original stratum had only one PSU when analyzed at both the middle and high school level.

Exhibit 4.1   Example SAS and SUDAAN Code for Generating Weighted Tobacco Product Use Estimates (Ever Use, Current Use)* and Standard Errors

```
SAS:
Proc Surveymeans Data=nyts2022 mean;

Var eelcigt ecigt ecigar eslt ehookah celcigt ccigt ccigar cslt chookah;

Class eelcigt ecigt ecigar eslt ehookah celcigt ccigt ccigar cslt chookah;

Stratum v_stratum2;

Cluster psu2;

Weight finwgt;

Domain SCHOOLTYPE SCHOOLTYPE*Sex SCHOOLTYPE*Race_S;

Title "NYTS 2022, Tobacco Product Use 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=nyts2022 Filetype= SAS Design=WR;

Var eelcigt ecigt ecigar eslt ehookah celcigt ccigt ccigar cslt chookah;

Catlevel 1 1 1 1 1 1 1 1 1 1;

Nest v_stratum2 PSU2 / Missunit;
```

APPENDIX A. QUESTIONNAIRE

QUESTIONNAIRE ONLY INCLUDED IN PDF VERSION OF THIS DOCUMENT.

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 attachment.

Non-Hispanic Asian—A person having origins in any of the original peoples of the East Asia, Southeast Asia, or the Indian subcontinent.

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

Non-Hispanic Pacific Islander[4]—A person having origins in any of the original peoples in the Pacific Islands. This area includes, for example, Guam, Hawaii, Samoa, and other Pacific Islands.

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.

---

[4] Our design and estimation processes separate out the two subgroups, Asian and Pacific Islander, as per Final Standards, Office of Minority Health (https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=53).