METHODOLOGY REPORT OF THE 2023 NATIONAL YOUTH TOBACCO SURVEY

Recommended Citation

Office on Smoking and Health. 2023 National Youth Tobacco Survey: Methodology Report. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

For questions about this report, please email Monica Cornelius at yex8@cdc.gov

Centers for Disease Control and Prevention

Office on Smoking and Health

Atlanta, GA

April 2024

Table of Contents

APPENDIX

APPENDIX A: COMMON CORE OF DATA RACE AND ETHNICITY DEFINITIONS

TABLES

CHAPTER 1—NYTS OVERVIEW

1.1. Overview of the National Youth Tobacco Survey

The National Youth Tobacco Survey (NYTS) is an annual cross-sectional survey of students in grades 6–12 in U.S. public and private schools. Randomly selected students are recruited to complete an online questionnaire focused on tobacco product use, exposure to tobacco products, smoking cessation, access to tobacco products, and knowledge and attitudes about tobacco products.

CHAPTER 2—NYTS SAMPLING METHODS

2.1. Sampling Frame

The frame used to select the 2023 NYTS sample combined data files obtained from the National Center for Education Statistics (NCES) and Market Data Retrieval Inc. (MDR). NCES data came from two sources: The 2021–2022 Common Core of Data (CCD) for public schools and the 2019–2020 Private School Universe Survey (PSS) for private schools. The 2021–2022 CCD was mandatory for public schools and the 2019–2020 PSS was optional for private schools. The response rate to the PSS was 74.5%, consequently, 25.5% of private schools were not represented in the PSS.[1] MDR data were from the summer of 2022, which were the latest data available at the time of sampling. Private schools captured by the MDR data that were not in the PSS were included in the 2023 NYTS sampling frame.

The 2023 NYTS frame included the subset of all public (including charter) and private school students enrolled in regular middle and high schools in grades 6 through 12 in the 50 US states and the District of Columbia. The following school types were excluded: alternative, special education, Department of Defense–operated, Bureau of Indian Affairs, adult education, and vocational schools.

For the purposes of NYTS sampling, schools with both high school and middle school students were considered as two separate schools. The frame had 82,394 eligible schools and 28,313,765 eligible students. The MDR data provided information for 33.2% of the private schools (4,710 out of 14,203).

After separating schools that had both high school and middle school students into middle school (grades 6–8) and high school (grades 9–12) as appropriate, any schools with fewer than 40 students were excluded from the sampling frame. Students in schools with less than 40 students had zero probability of selection into the sample; as a result, they were not covered. Overall, 0.7% of all otherwise eligible students were not covered; this figure represents 0.4% of all public school students and 5.1% of all private school students. Table 1 displays the totals used to calculate

---

[1] Broughman, S. P., Kincel, B., & Peterson, J. (2021). *Private School Universe Survey (PSS): Public-use data file user's manual for school year 2019–20* (NCES 2022021). US Department of Education: National Center for Education Statistics. Retrieved 15 August 2023 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022021

coverage error for schools and students by public and private schools, and overall totals. The table double counts the schools with both middle and high school students that were treated as two separate schools.

*Table 1: School and Student Coverage Error**

School type	Schools				Students			
	All	Schools with 40 or more students	Schools with less than 40 students	Coverage error (%)	All	Students in schools with 40 or more students	Students in schools with less than 40 students	Coverage error (%)
Public	58,557	53,542	5,015	8.6	26,103,388	26,005,063	98,325	0.4
Private	23,837	14,203	9,634	40.4	2,210,377	2,098,138	112,239	5.1
Total	**82,394**	**67,745**	**14,649**	**18.1**	**28,313,765**	**28,103,201**	**210,564**	**0.7**

*Coverage error is the set of schools, and students in those schools, that are eligible but have zero probability of selection. Coverage error comes from two sources, 1) schools that are not eligible to be sampled because they have less than 40 students (i.e., type 1), and 2) the eligible schools not included on the sample frame due to frame error (i.e., type 2). The coverage error we report is type 1. The type 2 coverage error is not reported because it is unknown.

For the remainder of this sampling section, the term "students" refers to students eligible for the 2023 NYTS—that is, 6th to 12th graders in the set of eligible schools.

2.2. SAMPLE DESIGN

2.2.1. Overview of the Design

The NYTS sampling methodology was designed to produce a nationally representative sample of students in grades 6 through 12 who attend regular US public and private schools. The sample is not designed to yield estimates for subnational geographical regions. The study is designed to produce precise national estimates by school level (middle and high school), by grade (6, 7, 8, 9, 10, 11, and 12), by sex (male and female), and by race/ethnicity (non-Hispanic [NH] White, NH Black, Hispanic, NH Asian, and NH American Indian/Alaskan Native). The sampling approach includes three stages: 1) selection of Primary Sampling Units (PSUs) within strata; (2) selection of schools within PSUs; and (3) selection of classes within each selected school (Section 2.2.5). All students are selected in each selected class (this is not considered a sampling stage). The following six steps outline the design.

1) PSUs were formed by either:
    a. Grouping schools that contain 20% or more of NH American Indian/Alaska Native (hereafter, NH AI/AN) or 20% or more NH Asian students, or, for the remaining schools,
    b. Grouping schools by counties:
        i. Counties with more than 150,000 students were broken into multiple PSUs.

ii. Counties with less than 150,000 students that contained at least five high schools and four middle schools formed a unique PSU.

iii. Counties with fewer than five high schools and four middle schools were joined with other small counties within the same state.

2) PSUs were assigned to 17 strata. The strata were formed by the cross-classification of differing densities of NH AI/AN, NH Black, Hispanic, and NH Asian students and rural/urban classification (rural/urban refers to the location of each school).

3) Within those 17 strata, 140 PSUs were selected, with probability proportional to the number of students in the PSU.

4) From the 140 PSUs, 420 schools were selected. Three schools were selected within each PSU with probability proportional to size (PPS).[2] The size measure for high schools equals enrollment. The size measure for middle schools equals 1.46 times enrollment.[3] The oversample of middle schools was required to ensure an equal number of middle schools and high schools, which resulted in an approximately equal number of students per grade in the sample.

5) Selected schools with more than 100 students per grade were randomized to have one or two classes per grade sampled. In total, 80 middle schools and 80 high schools had two classes per grade sampled.[4] Schools with fewer than 120 students had all classes selected. The remaining schools had one class per grade selected. Schools were given the option of administering the survey to a census of all eligible classes (i.e., all classes in grades 6–12) if they preferred to do so.

6) In each selected class, a census of students was selected (i.e., all students).

2.2.2. Stratification

PSUs were organized into 17 strata, based on proportions of racial and ethnic minority student enrollment and urban/non-urban classification. The strata were formed by applying the following six steps.

1) Schools with 20% or more NH AI/AN students were grouped into two strata based on race density. One stratum contained schools with 20% to 40% NH AI/AN students. The other stratum contained schools with more than 40% NH AI/AN students.

2) Schools with 20% or more NH Asian students were grouped into two strata based on race density. One stratum contained schools with 20% to 40% NH Asian students. The other stratum contained schools with more than 40% NH Asian students.

---

[2] In the 2023 sampling methodology, the size measure was increased to oversample middle schools. This change from the 2022 methodology nullified the need to stratify by type (middle school/high school) and size before drawing the sample, while still producing a similar distribution of schools by size and type, improving precision, and making the sampling design easier to implement.

[3] The value 1.46 was determined by manipulating this number until there were equal numbers of selected high school and selected middle schools in the sample.

[4] The main sample in the 2022 design had 120 schools with two classes per grade selected, out of a total of 320 schools selected. In 2023, we selected 420 schools. We maintained the proportion, opposed to the absolute amount, of schools that had two classes per grade selected.

3) For the remaining schools, if the percentage of Hispanic students in the PSU exceeded the percentage of NH Black students, the PSU was classified as having a high density of Hispanic students. Otherwise, the PSU was classified as having a high density of NH Black students.
4) The PSU was assigned to rural/urban status based on the 2013 National Center for Health Statistics (NCHS) urban/rural classification for counties.[5]
5) PSUs in the Hispanic urban and Hispanic rural strata were classified into four density groupings depending upon the percentage of Hispanic students in the PSU. The density bounds were selected to produce approximately equal sized groups.
6) PSUs classified as NH Black urban were assigned into four groupings, depending upon the percentages of NH Black students in the PSU (Table 2). The density bounds were selected to produce approximately equal sized groups. PSUs classified as NH Black rural were grouped into one stratum because the number of students was not sufficient for more strata.

Table 2 presents the definition of the strata, counts of students (enrollment), and the stratum boundaries used in the 2023 NYTS.

[5] In the 2022 NYTS, schools were assigned based on the MDR's definition of urban/non-urban, where urban was defined as a county in a metropolitan statistical area (MSA). The MDR definition resulted in misclassification of students by rural/urban status. The 2023 sample improved upon the 2022 design by using the 2013 NCHS Urban-Rural Classification Scheme for Counties, which is an established, more accurate definition. See this web page for more information: https://www.cdc.gov/nchs/data_access/urban_rural.htm.

Table 2: 2023 NYTS Strata Definitions, Density Boundaries, and Enrollment

Stratum Code	Race/Ethnicity	Urban/ Rural	Density Group	Density Bounds	Enrollment
AI1	NH AI/AN	N/A	1	20%–40%	102,246
AI2			2	>40%–100%	104,788
AS1	NH Asian	N/A	1	20%–40%	1,196,367
AS2			2	> 40%–100%	490,121
BRx	NH Black	Rural	X	0%–100%	948,524
BU1		Urban	1	0%–26%	1,557,373
BU2			2	> 26%–40%	1,853,554
BU3			3	> 40%–54%	1,697,817
BU4			4	> 54%–100%	2,007,657
HR1	Hispanic	Rural	1	0%–24%	642,436
HR2			2	> 24%–48%	653,005
HR3			3	> 48%–68%	662,575
HR4			4	> 68%–100%	623,681
HU1		Urban	1	0%–26%	3,700,008
HU2			2	> 26%–42%	3,771,215
HU3			3	> 42%–58%	4,069,586
HU4			4	> 58%–100%	4,022,248

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic
X = PSUs classified as non-Hispanic Black rural were grouped into one stratum because the number of students was not sufficient for more strata.

2.2.3. Formation of PSUs

For schools with over 20% NH AI/AN students, PSUs were formed by sorting the schools by NH AI/AN density and forming groups of consecutive schools that had 5,000 or more students and four or more middle schools and five or more high schools. For schools with over 20% NH Asian students, the same procedure was applied, except the number of students per PSU was constrained to 6,000 or more students.[6]

Fourteen counties had more than 150,000 eligible students. The schools in these counties were sorted by school district and broken into groups of schools with less than 100,000 students in each group.

There were 1,990 counties that did not have four or more middle schools and five or more high schools and therefore needed to be combined with other counties to achieve that criterion. Within

[6] The 2023 sampling design used a slightly different technique to oversample NH AI/AN and NH Asian students than the 2022 design. In 2022, oversampling occurred at the PSU level and then at the school level. The 2023 sample oversamples at the school level alone and is therefore more efficient than the 2022 design.

strata, the small counties were combined with other small counties in the same state so that the combined counties had four or more middle schools and five or more high schools. There were 140 PSUs selected out of the 2,140 PSUs on the frame.

2.2.4. Sample Size Goals and Response Rate Assumptions

Our goal was to collect data from 285 responding schools. Table 3 displays the NYTS school response rate by year. School response rates declined in 2020 in the context of COVID-19 and had not yet recovered to pre-pandemic levels in 2022. Our goal was to achieve an overall response rate (school response rate * student response rate) of at least 60%. However, in combination with our assumption that 4% of the selected schools would be ineligible, we assumed a 2023 school response rate of 70.6% and 96% school eligibility rate. Consequently, to obtain 285 responding schools, 420 schools were selected (285 / (0.706 * 0.96)) in 140 PSUs. We selected three schools per PSU.

Table 3: School Response Rate by NYTS Data Collection Year

NYTS Data Collection Year	School Response Rate (%)
2018	76.8
2019	77.2
2020	49.9
2021	54.9
2022	59.4

2.2.5. Sample Selection

The NYTS 2023 sample selection had three sampling stages:

1) 140 PSUs were selected within 17 strata via a systematic selection with probability proportional to the number of students in the PSU. Within strata, the PSUs were sorted by state Federal Information Processing Standards (FIPS) code and county FIPS code, ensuring geographic diversity in the sample.
2) Using a systematic selection with PPS, three schools were selected within each PSU. The size measure for high schools was equal to the school enrollment. The size measure for middle schools was equal to 1.46 times the school enrollment. The oversample of middle schools was required to produce an approximately equal number of students in each grade in the 2023 NYTS sample.

3) Selection of classes:

   a. In schools with more than 100 students per grade, either one or two classes per grade were selected in each school.[7] A random process was used to assign 80 middle schools and 80 high schools to have two classes per grade selected.

   b. In schools with 120 or fewer students across all eligible grades (grades 6–12), all classes were selected.[8]

   c. In all other schools, one class per grade was selected in each school.

   d. During the recruitment of schools, some schools either refused to provide a class list to enable classroom selection or insisted that all classes participate in the survey. In these situations, a census of classes was taken.

The following formula describes the probability of selection for each student. Each factor in the product on the right size of the formula corresponds to a sampling stage.

$$Prob_{i,j,k,l,m} = \frac{\alpha_i * MOS_{i,j}}{\sum_{\forall j} MOS_{i,j}} * \frac{3 * \left(1 + 0.46 * 1_{MS}(school_{i,j,k})\right) * MOS_{i,j,k}}{MOS_{i,j}} * \frac{\beta_{i,j,k} * students\ per\ class_{i,j,k}}{MOS_{i,j,k}}$$

Where, i = stratum, j = PSU, k = school, l = class, m = student.

$Prob_{i,j,k,l,m}$ = Probability of selecting the *m*th student, in the *l*th class, in the *k*th school, in the *j*th PSU in the *i*th stratum.

α_i = The number of PSUs selected in the *i*th stratum.

$MOS_{i,j}$ = number of eligible students in the *i*th stratum and *j*th PSU; MOS—measure of size.

The notation $\left(1 + 0.46 * 1_{MS}(school_{i,j,k})\right)$ is equal to 1.46 if the $school_{i,j,k}$ is a middle school and 1 if the $school_{i,j,k}$ is a high school.[9]

$MOS_{i,j,k}$ = This is the number of eligible students in the *i*th stratum and *j*th PSU and *k*th school.

$\beta_{i,j,k}$ = classes selected in the *k*th school, in the *j*th PSU, in the *i*th stratum.

The following is a simplification of the above formula describing the probability of selection for each student.

[7] The selection of schools with one or two classes per grade differs from the 2022 NYTS methodology. In 2022, the PSUs were stratified into school sizes. The additional stratification applied in 2022 increased the unequal weighting effect.
[8] This differs from the 2022 NYTS, where there was no census applied to small schools. Without the census, the weights of the students in small schools would be higher than the weights of students in larger schools, thereby, lowering the precision of the estimates due to an increase in unequal weighting effect. Also, including more students increases the precision of the estimates due to the increase in responding students.
[9] This notation uses an indicator function. See https://dk81.github.io/dkmathstats_site/prob-indicator.html for a description.

$$Prob_{i,j,k,l,m} = \frac{3 * \alpha_i * \beta_{i,j,k} * \left(1 + 0.46 * 1_{MS}(school_{i,j,k})\right) * students\ per\ class_{i,j,k}}{\sum_{\forall j} MOS_{i,j}}$$

Table 4 presents, by stratum and overall, the counts of PSUs, schools, and students in the sampling frame and selected for participation in the study. Table 2 contains the stratum code definitions that correspond to the first column of Table 4. The column labeled "mean student weight" is the student's weight before adjusting for the mismatch between the school's enrollment between the frame and the observed enrollment, school nonresponse, class nonresponse, and student nonresponse.

Table 4: Counts of PSUs, Schools, and Students in the Sampling Frame and Selected for Participation in the Study by Stratum with Mean Student Weight

Stratum Code	Frame Counts			Selection Counts			Mean Student Weight
	PSUs	Schools	Students	PSUs	Schools	Students	
AI1	19	490	102,246	4	12	1,204	80.6
AI2	20	758	104,788	7	21	1,800	49.7
AS1	94	1,602	1,196,367	7	21	1,992	541.6
AS2	33	616	490,121	5	15	1,872	259.7
BRx	249	3,519	948,524	4	12	1,204	748.7
BU1	125	3,642	1,557,373	8	24	2,808	528.2
BU2	79	3,905	1,853,554	8	24	2,612	624.0
BU3	53	3,822	1,697,817	8	24	2,796	539.6
BU4	91	4,924	2,007,657	9	27	2,840	658.8
HR1	210	3,110	642,436	4	12	1,236	429.6
HR2	195	2,813	653,005	4	12	1,236	476.1
HR3	166	2,599	662,575	4	12	1,252	484.7
HR4	168	2,318	623,681	4	12	1,408	425.2
HU1	315	9,021	3,700,008	16	48	6,156	575.1
HU2	125	7,754	3,771,215	16	48	5,704	670.5
HU3	92	8,477	4,069,586	16	48	5,912	630.2
HU4	106	8,375	4,022,248	16	48	5,656	685.1
All Strata	**2,140**	**67,745**	**28,103,201**	**140**	**420**	**47,688**	

Abbreviation: PSU = Primary Sampling Unit

The 2022 NYTS design, which we followed for 2023, resulted in an unequal weighting effect due to oversampling of specific groups. The unequal weighting effect of the 2023 design for estimates to the entire population is 1.23.[10] The 2023 design deviated from an equal probability of sampling method (EPSOM) design[11] for all sample members for the following reasons:

1) The number of PSUs selected in a stratum (α_i) varied. This aspect of the design was implemented:
   a. To increase the proportion of NH AI/AN, NH Asian, NH Black, Hispanic, and/or rural students.
   b. Because an integer number of PSUs were selected within each stratum. Consequently, an allocation exactly proportional to size across strata was not possible.
2) During the selection of schools, middle schools were oversampled to achieve three-sevenths of the respondents attending middle school.
3) The number of classes selected in each school differed across schools.
   a. Middle schools with grades 6–8 could have either three or six classes selected. High schools with grades 9–12 could have either four or eight classes selected.
   b. Some middle schools and some high schools did not have a full complement of grades, reducing the number of classes selected.
4) The average number of students per class varied in the different schools.
   a. There was a natural variation across schools.
   b. Due to the pandemic, schools have organized classes differently. We observed classes with more than 100 students and classes with as few as 4 students.

Tables 5a, 5b, and 5c present the estimate of the number of respondents by school type and overall, for grade, race/ethnicity category, and rural status, assuming overall response rates of 40%, 50%, and 60%, respectively. These results assume that response propensity is not correlated with the various characteristics (e.g., race/ethnicity category, high school/middle school, and rural status).[12] Tables 10 and 11 display the number of students we observed in each of these categories.

Table 5a: Estimate of Respondents by School Type, Race/Ethnicity Category and Rural Status with 40% Overall Response Rate

School Type	Total Respondents	Each Grade	NH White	NH Black	Hispanic	NH Asian	NH AI/AN	Rural
High school	10,963	2,741	4,544	1,393	2,809	681	435	1,997
Middle school	8,112	2,704	3,457	993	1,960	493	353	1,358
Total	**19,075**	**N/A**	**8,001**	**2,386**	**4,769**	**1,173**	**788**	**3,355**

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic

[10] This reported design effect applied to the student weights without nonresponse adjustment and without calibration.
[11] The documentation for the 2022 design suggests that the design was EPSOM. We believe this is not the case for the enumerated reasons on this page. We estimate that the unequal weighting effect for student selected in the 2023 design was less than the unequal weighting effect of the 2022 design.
[12] We found that response propensity was correlated with grade, high school/middle school, and race.

Table 5b: Estimate of Respondents by School Type, Race/Ethnicity Category and Rural Status with 50% Overall Response Rate

School Type	Total Respondents	Each Grade	NH White	NH Black	Hispanic	NH Asian	NH AI/AN	Rural
High school	13,704	3,426	5,680	1,741	3,511	851	544	2,496
Middle school	10,140	3,380	4,321	1,242	2,450	616	441	1,698
Total	**23,844**	**N/A**	**10,001**	**2,983**	**5,961**	**1,467**	**985**	**4,194**

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic

Table 5c: Estimate of Respondents by School Type, Race/Ethnicity Category and Rural Status with 60% Overall Response Rate

School Type	Total Respondents	Each Grade	NH White	NH Black	Hispanic	NH Asian	NH AI/AN	Rural
High school	16,445	4,111	6,816	2,089	4,213	1,021	653	2,995
Middle school	12,168	4,056	5,185	1,490	2,940	739	529	2,038
Total	**28,613**	**N/A**	**12,001**	**3,579**	**7,153**	**1,760**	**1,182**	**5,033**

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic

2.2.6. Replacement Schools

Data collection failed to meet school respondent goals due to data collection starting later than anticipated. On April 28, 2023, for schools that were hard refusals, we selected and began recruitment of replacement schools. In the original school selection, within strata, the list of schools was first sorted in order of descending school size and then a systematic selection was applied. The replacement for the original school that was selected was the next school on the ordered list after the original school, provided that the replacement school existed and was not in the sample. In the rare case in which two consecutive schools were selected or the last school on the list was selected, the replacement school was the previous school in the list.

CHAPTER 3—NYTS DATA COLLECTION AND PROCESSING

3.1. Survey Instrument

The NYTS collects data on key short-term, intermediate, and long-term tobacco product prevention and control outcome indicators. The 2023 survey instrument included 149 questions.

Several updates were made to the 2023 instrument based on results of cognitive testing of the 2022 instrument. To reduce survey burden, the 2023 instrument combined the question asking about use of any flavor with the question asking about specific flavor types for each tobacco product type used. This change allowed a more comprehensive assessment of flavor use by including additional flavor categories (e.g., tobacco-flavored, unflavored, non-alcoholic drinks). The 2023 instrument also updated tobacco product brands and added questions about brands of cigars (cigars, cigarillos, or little cigars), smokeless tobacco (chewing tobacco, snuff, or dip), and nicotine pouches, to accommodate changes in the popularity and availability of different brands over time. Similarly, the survey added new questions about e-cigarette, cigar (cigar, cigarillo, and little cigar), and nicotine pouch flavors to accommodate changes in popularity and tobacco industry marketing techniques over time—specifically to capture use of "ice" or "iced" flavors and concept flavors.

The 2023 survey replaced questions about dissolvable tobacco products with questions about oral nicotine products (defined as lozenges, discs, tablets, gums, dissolvable tobacco products, and other products) and updated sexual orientation and gender identity (SOGI) questions to accommodate changing terminology. The 2023 survey removed questions about visits with health professionals, some questions about the relative harmfulness of tobacco products, questions about exposure to tobacco advertising except for social media, and the Family Affluence Scale. Added items included a question about seeing e-cigarette use at school, the Adolescent Discrimination Distress Index, and the Neighborhood Environment Scale.

The 2023 NYTS represented the fifth cycle that the study was conducted with electronic data collection methods rather than traditional paper-and-pencil interviewing and the third cycle that was conducted 100% online. The web survey was created using Qualtrics XM and all data were stored in the FedRAMP-approved platform. To take the web survey, students navigated to a dedicated URL. Students logged in using a class ID that was tied to school ID. Neither identifier contained any identifying information about the class or school. School and class IDs were appended to all responses to facilitate data collection tracking and to calculate class and school response rates. Individual-level IDs were assigned and appended to each record but not displayed to participants.

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

The length of interview (LOI) was captured for each record and was calculated as the time lapse between the date/time of the first response and the date/time of the last response given. LOI ranged from 1 minute, 58 seconds to 105 minutes, with a median time of 18 minutes, 50 seconds.

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 product information on social media, 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 needed as they answered related questions. The 2023 NYTS also included questions about experiences of discrimination, neighborhood characteristics, depression and anxiety, and SOGI.

3.2. External Review and Approvals

Three bodies reviewed and approved the survey instrument, processes, privacy and security elements, and sampling design of the 2023 NYTS: the Office of Management and Budget (OMB), RTI International's Institutional Review Board (IRB), and CDC's IRB.

With the transition to an electronic, web-based format for the 2021 NYTS, the Security Assessment and Authorization (SA&A) approval and Enterprise Performance Life Cycle (EPLC) review remained valid and were updated for the 2023 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 requirements mandated by the Federal Information Security Management Act. EPLC is a framework to enhance the Department of Health and Human Services IT governance through rigorous application of sound investment and project management principals, in conjunction with industry best practices.

3.3. Study Liaison Staffing

The role of the Technical Assistance Provider 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 2023 NYTS cycle, and the role title was updated to Study Liaison (SL). SLs 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 of the questions that school contacts and/or teachers had 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, SLs were assigned by their geographical location. Every time zone that contained one or more sampled schools had at least one SL local to that part of the country. Ten SLs were recruited from a pool of previously trained data collectors. Three virtual trainings were held for SLs on October 19, October 26, and November 9, 2022, that focused on recruitment and gaining cooperation.

Key components of the training included the following:

- Pre- and post-survey communications with the schools and teachers
- Orientation to school contacting system
- IT troubleshooting
- Liaison roles and responsibilities

An additional two-day virtual training was held January 11, 2023, to instruct SLs on the logistics for classroom sampling, consent form protocols, and providing technical support to participating schools.

3.4. Recruitment Procedures

The 420 schools selected to participate in the 2023 NYTS were located in 42 states. Recruitment began in October 2022, a month later than previous NYTS cycles, with calls to state departments of education and health to inform them of the survey effort and sampled schools in their state. After notification at the state level, district- and school-level recruitment began. Due to the condensed recruitment cycle, districts and dioceses received a notification letter but were not contacted for approval. After district notification, schools were approached directly unless a district refusal was received. A date for survey implementation was selected that was convenient to the school. SLs used a secure web-based contacting system to facilitate communication and schedule survey dates upon request by the school. During the course of recruitment, 4 of the 420 schools (1%) were found to be ineligible to participate in the NYTS, resulting in 416 eligible schools.

3.5. Survey Administration

Data were collected during March 9, 2023 to June 16, 2023. Although the details of each data collection varied, there were six core steps followed for every school:

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

Procedures were designed to protect students' privacy by assuring that student participation was anonymous and voluntary. Teachers read a script at the start of the student session that introduced the study and provided informed consent. Using a school-issued or personal internet-connected device, students logged into a secure website, entered a unique class ID, and responded to a question regarding their location (e.g., classroom, home, other location) before beginning the survey. All surveys were submitted directly to the Qualtrics GovCloud, an environment that is specific only to federal customers and is FedRAMP-approved at moderate impact level.

3.5.1. Field Procedures

After schools were recruited, classes selected, and a date for survey administration scheduled, the school coordinator received an email with pre-survey materials containing instructions for survey administration, the survey URL to confirm access, and the parental permission forms to be distributed to all students in the selected classes prior to data collection. Pre-survey materials were sent to schools about three weeks prior to survey administration. SLs 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 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 high school students in a few states—physical education and/or health also were considered core courses. However, in a small number of schools, it was difficult to develop an appropriate frame using this approach. Therefore, in these schools, classes were selected by using a time of day (e.g., second period) when all eligible students were scheduled to be attending a class of one kind or another, and randomly selecting from all classes held at this time. Last, in some schools, homerooms or advisory periods were used as the frame for class selection.

3.6. Web-Based Data Collection Management Application

Project management staff and SLs used a web-based data collection management system to facilitate information exchange with project staff and allow all members of the project management, recruitment, supervisory teams, and remote staff access to information necessary to implement the study. The system is designed with differing levels of access depending on the user's role on the study. The system's main functions include tracking recruitment progress, scheduling data collection, sampling classes, and tracking school and student response rates.

3.7. Data Recording

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

3.8. PARTICIPATION RATES

See Sections 4.3.1 and 4.3.2 for this information.

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 definitions of race and ethnicity used in previous rounds of the NYTS. 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 dataset 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 compiled during frame construction, sample selection, recruitment, and data collection were merged correctly using a common school identifier.

CHAPTER 4—WEIGHTING OF NYTS RESPONDENT DATA

4.1. OVERVIEW

This chapter describes the weighting methodology applied to create the analytic weights. The analytic weights are used when estimating population parameters with the 2023 NYTS data. The 2023 weighting methodology is consistent with the 2022 NYTS methodology. Slight differences between the 2022 and 2023 weighting methodology are noted in the text.

Sampling weights are the link between the sample and the population. In lay terms, the sampling weights ensure that estimates calculated from the sample represent the population. To construct the weights used to make population estimates, known as the analysis weights, we applied the following six steps.

Six steps in the weighting:

1) Calculate the base weight for every school sampled—the inverse of the probability of selecting each school.
2) Calculate a school-level nonresponse adjustment to the base weight.
3) Calculate the class weight—the inverse of the probability of selecting each class within the school.

4) Calculate two class-level nonresponse adjustments.
   a. Adjusts for nonresponse of an entire class.
   b. Adjusts for nonresponse of students within a responding class.
5) Calculate the nonresponse adjusted student weights—the product of the school base weight, the school nonresponse adjustment, the inverse of the class selection probability, and the two-class nonresponse adjustments.
6) Calibrate the nonresponse adjusted student weights weight to population totals.

No weight trimming was applied during the creation of the 2023 NYTS weights, in contrast to previous iterations of the NYTS.

4.2. Calculate the Weights

4.2.1. *Weighting Step 1—Calculate the Base Weight for Each School*

The base weight of each school is the inverse of the probability of selecting each school. The probability of selecting each school is the product of the probability of selecting the PSU, and the probability of selecting the school.

4.2.1.1. Probability of Selecting the PSU

The probability of selecting the PSU is $\frac{\alpha_i * MOS_{i,j}}{\sum_{\forall j} MOS_{i,j}}$.

Where, α_i = The number of PSUs selected in the *i*th stratum.

$MOS_{i,j}$ = Number of eligible students in the *i*th stratum and *j*th PSU;

$\sum_{\forall j} MOS_{i,j}$ is the sum of the enrollment over all PSUs in stratum i.

The symbol ∀ is read "for all."

4.2.1.2. Probability of Selecting the School

The probability of selecting the school is $\frac{3*\left(1+0.46*1_{MS}\left(school_{i,j,k}\right)\right)*MOS_{i,j,k}}{MOS_{i,j}}$.

where, $\left(1 + 0.46 * 1_{MS}\left(school_{i,j,k}\right)\right)$ is equal to 1.46 if the $school_{i,j,k}$ is a middle school and 1 if the $school_{i,j,k}$ is a high school.

$MOS_{i,j}$ is the number of eligible students in the *i*th stratum and *j*th PSU.

4.2.1.3. School Base Weight

Each school's base weight is the inverse of the product of the probabilities of selection of the first two sampling stages. The following is the formula for $Prob_{i,j,k}$, the probability of selecting for each school sampled.

$$Prob_{i,j,k} = \frac{\alpha_i * MOS_{i,j}}{\sum_{\forall j} MOS_{i,j}} * \frac{3 * \left(1 + 0.46 * 1_{MS}\left(school_{i,j,k}\right)\right) * MOS_{i,j,k}}{MOS_{i,j}}$$

The base weight assigned to each selected school is the inverse of $Prob_{i,j,k}$.

$$W_{i,j,k}^{School} = \frac{1}{Prob_{i,j,k}}$$

where $W_{i,j,k}^{School}$ is the base weight of the k^{th} school in the j^{th} PSU, in the i^{th} stratum.

4.2.2. Weighting Step 2—Calculate a School Nonresponse Adjustment

A logistic regression model was used to predict each school's probability of response based on the characteristics of the school. The weights of the responding schools were increased by dividing the school's weight by the predicted probability of response. The data from the 416 eligible schools were fit to a logistic regression model where school response was the dependent variable, and the following independent variables were considered:

- Stratum
- Census region (4 levels)[13]
- Census division (9 levels)[14]
- Private school/public school
- Middle school/high school
- Enrollment
- Rural/urban
- Percentage of NH White students categorized into quintiles
- Percentage of NH Black students categorized into quintiles
- Percentage of Hispanic students categorized into quintiles
- Percentage of NH Asian students categorized into quintiles
- Percentage of NH AI/AN students categorized into quintiles

A backward stepwise model fitting procedure was implemented. A logistic regression model was fit with the school response as the dependent variable and all the independent variables listed above. The variable with the largest p-value greater than 0.1 was removed from the model. This process was iterated until all variables left in the model had p-values less than 0.1. The final model contained the following independent variables: stratum, division, private/public, middle school/high school, NH Asian quintile, and enrollment. The area under the ROC curve was 0.76, indicating that the independent variables explain much of the variance in school response.

The fitted logistic regression model has the form:

$$\text{logit}(p_{i,j,k}) = \beta_0 + \beta_1 X_{1,i,j,k} + \cdots + \beta_n X_{n,i,j,k}$$

[13] The definition of Census region and Census division can be found here: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

[14] See the previous footnote.

We calculated the probability of response for each school sampled:

$$p_{\mathrm{i,\,j,\,k}} = \frac{1}{1 + e^{-\mathrm{logit}(p_{i,j,k})}}$$

The index i corresponds to a stratum i.
The index j corresponds to a PSU j.
The index k corresponds to school k.

The school nonresponse adjustment was calculated as

$$Adj_{i,j,k}^{School\ NR} = \frac{1}{P_{\mathrm{i,j,k}}},$$

where,

$Adj_{i,j,k}^{School\ NR}$ = School nonresponse adjustment for the k^{th} responding school, in the j^{th} PSU, in the i^{th} stratum.

$P_{\mathrm{i,\,j,\,k}}$ = the predicted probability of response from the logistic regression model for the k^{th} responding school, in the j^{th} PSU, in the i^{th} stratum.

Use of a logistic regression model to estimate the response propensity for a school and applying a nonresponse adjustment to each school equal to the inverse of the school's modeled response propensity, differs from the nonresponse adjustment methodology applied in 2022. In 2022, the school nonresponse adjustment was based exclusively on the stratum; the adjustment was equal to the ratio of the number of students in the schools sampled in a stratum divided by the number of students in the responding schools in that stratum. In the approach applied in 2023, we incorporated stratum as well as five other school characteristics that were shown to be correlated with school response propensity. Following the 2022 methodology of creating a logistic regression model that uses stratum as the only independent variable results in an area under the ROC curve of 0.67. The 2023 model improves upon the area under the ROC, increasing it to 0.76, by including stratum, division, private/public, middle school/high school, NH Asian quintile, and enrollment in the model. This is evidence that the model used in 2023 explains more of the variance than just using stratum exclusively. We believe the difference in the school-level nonresponse adjustment between 2022 and 2023 data will not adversely affect the comparison of outcomes across the two data collection years. The effect on the estimates of including the variables, other than stratum, in the nonresponse will be small (e.g., it will potentially only affect the tenths place of the estimate, at most).

4.2.3. Weighting Step 3—Adjust for the Probability of Selecting each Class Within the School

The probability of selecting a class within each selected school is:

$$Prob_{i,j,k,l} = \frac{\beta_{i,j,k} * mean\ students\ per\ class_{i,j,k}}{MOS_{i,j,k}}$$

Where $Prob_{i,j,k,l}$ = the probability of selecting the l^{th} class, in the kth school, in the jth PSU, in the ith stratum.

$\beta_{i,j,k}$ = Number of classes selected in the kth school, in the jth PSU, in the ith stratum.

$MOS_{i,j,k}$ = The number of eligible students in the ith stratum and jth PSU and kth school.

l = Class.

The class weight is:

$$W_{i,j,k,l}^{Class} = \frac{1}{Prob_{i,j,k,l}}$$

Where $W_{i,j,k,l}^{Class}$ is the contribution to the weight from selecting the class within a school for the l^{th} class, in the kth school, in the jth PSU, in the ith stratum.

4.2.3.1. Weighting Step 4a—Class-Level Nonresponse Adjustment Within School

If an entire class fails to respond, the weight of the responding classes in that school is adjusted by multiplying by the ratio of the sum of the selected classes to the sum of the responding classes.

The class nonresponse adjustment is:

$$Adj_{i,j,k,l}^{Class\ NR} = \frac{Classes\ selected\ in\ school_{i,j,k}}{Classes\ responding\ in\ school_{i,j,k}}$$

4.2.3.2. Weighting Step 4b—Student-Level Nonresponse Adjustment Within Class

All students are selected within a class. The probability of selecting a student in a selected class is 1, and there is no adjustment to weights for the student selection with a class. There is an adjustment for student nonresponse. The weight of the responding students is adjusted, within class, by ratio of the sum of the students enrolled in that classes to the sum of the responding students.

The student nonresponse adjustment is:

$$Adj_{i,j,k,l,m}^{Student\ NR} = \frac{Students\ selected\ in\ class_{i,j,k,l}}{Students\ responding\ in\ class_{i,j,k,l}}$$

4.2.3.3. Weighting Step 5—Calculate the Nonresponse Adjusted Student Weight

The nonresponse adjusted student weight, is the product of the school weight, the school nonresponse adjustment, the class weight, the class nonresponse adjustment, and the student nonresponse adjustment.

$$W_{i,j,k,l,m}^{NR\ adj\ student} = W_{i,j,k}^{School} * Adj_{i,j,k}^{School\ NR} * W_{i,j,k,l}^{Class} * Adj_{i,j,k,l}^{Class\ NR} * Adj_{i,j,k,l,m}^{Student\ NR}$$

Where, $W_{i,j,k,l,m}^{NR\ adj\ student}$ is the nonresponse adjusted student weight, for the m^{th} student, in the l^{th} class, in the kth school, in the jth PSU, in the ith stratum.

The nonresponse adjusted student weight is the input weight used in the calibration procedure.

4.2.4. Weighting Step 6—Calibration

In calibration, for various demographic distributions, the sums of the weights are constrained to equal population totals calculated from the sampling frame. Calibration can reduce coverage bias and nonresponse bias. Calibration was performed using SUDAAN PROC WTADJUST. We used the same calibration categories as the 2022 NYTS:

- For public schools:
  - Combinations of grade and sex
  - Combinations of grade and race/ethnicity[15]
    - The race categories are NH AI/AN, NH Asian, NH Black, Hispanic, NH White, and NH Other (people who identify with some other race not in this list and/ or identify with more than one race)
- For private schools:
  - Grade

Grade was missing for 197 (0.9%) of the students, sex was missing for 215 (1.0%) of the students, and race was missing for 375 (1.7%) of the students. These missing values were imputed using two different methods.[16] Grade was imputed with the mode of the reported grade in the class of the missing data item. Sex and race were imputed using a hot-deck imputation procedure. In the hot-deck procedure, students are sorted by school and class, then randomized within class. The missing values are replaced with the data value from the preceding student based on the ordering of the file. The hot-deck imputation method preserves the distribution of the imputed variables, in expectation.

Tables 6, 7, and 8 contain the distributions and population totals used in the calibration as well as the number and percentage of respondents in each calibration cell.

[15] In the 2022 NYTS documentation, NH Asian and NH AI/AN students were presented separately. In the 2023 NYTS documentation, NH Asian and NH AI/AN are considered another category of race. The 2022 and 2023 methodologies are equivalent, but the presentation is different.
[16] Imputation is only used during the calibration. The imputed values will not be used when reporting estimates. It is necessary to impute prior to calibration because a weight cannot be calculated for cases where the respondent is missing any of the variables used in the calibration.

Table 6: Calibration Distribution Public/Private Schools by Sex and Grade

Public/ Private	Sex	Grade	Population		Respondents	
			n	%	n	%
Public	Female	6th grade	1,808,970	6.4	1,988	9.0
		7th grade	1,870,517	6.6	1,791	8.1
		8th grade	1,881,602	6.7	1,684	7.6
		9th grade	1,924,163	6.8	1,692	7.7
		10th grade	1,857,990	6.6	1,401	6.3
		11th grade	1,749,682	6.2	1,303	5.9
		12th grade	1,688,829	6.0	831	3.8
	Male	6th grade	1,896,643	6.7	2,000	9.1
		7th grade	1,960,036	7.0	1,849	8.4
		8th grade	1,971,925	7.0	1,664	7.5
		9th grade	1,997,516	7.1	1,851	8.4
		10th grade	1,928,988	6.8	1,501	6.8
		11th grade	1,814,579	6.4	1,287	5.8
		12th grade	1,751,948	6.2	924	4.2
Private	Male and Female	All grades	2,098,138	7.4	303	1.4
Total			**28,201,526***	**100.0**	**22,069**	**100.0**

*This number is slightly different than the total eligible students on the frame because there were inconsistencies on the frame between the total school enrollment and enrollment in each grade.

Table 7: Calibration Distribution among Public/ Private Schools by Race/Ethnicity and Grade

Public/ Private	Race/ Ethnicity	Grade	Population		Respondents	
			n	%	n	%
Public	NH White	6th grade	1,666,623	5.9	1,623	7.4
		7th grade	1,736,841	6.2	1,430	6.5
		8th grade	1,756,870	6.2	1,304	5.9
		9th grade	1,829,240	6.5	1,345	6.1
		10th grade	1,777,730	6.3	1,111	5.0
		11th grade	1,700,999	6.0	926	4.2
		12th grade	1,658,541	5.9	545	2.5
	NH Black	6th grade	569,061	2.0	649	2.9
		7th grade	578,349	2.1	543	2.5
		8th grade	575,159	2.0	455	2.1
		9th grade	584,486	2.1	405	1.8
		10th grade	551,205	2.0	298	1.4
		11th grade	503,217	1.8	260	1.2
		12th grade	480,770	1.7	128	0.6
	Hispanic	6th grade	1,062,761	3.8	899	4.1
		7th grade	1,093,938	3.9	855	3.9
		8th grade	1,097,589	3.9	826	3.7
		9th grade	1,090,330	3.9	1,209	5.5
		10th grade	1,052,927	3.7	941	4.3
		11th grade	974,178	3.5	844	3.8
		12th grade	927,599	3.3	700	3.2
	NH Asian	6th grade	191,946	0.7	311	1.4
		7th grade	199,994	0.7	394	1.8
		8th grade	201,807	0.7	375	1.7
		9th grade	212,150	0.8	223	1.0
		10th grade	208,191	0.7	250	1.1
		11th grade	200,582	0.7	286	1.3
		12th grade	195,355	0.7	221	1.0
	NH AI/AN	6th grade	37,986	0.1	247	1.1
		7th grade	38,850	0.1	145	0.7
		8th grade	38,690	0.1	134	0.6
		9th grade	39,129	0.1	131	0.6
		10th grade	37,178	0.1	101	0.5
		11th grade	34,340	0.1	90	0.4
		12th grade	33,011	0.1	67	0.3
	Other*	6th grade	177,236	0.6	259	1.2
		7th grade	182,581	0.6	273	1.2
		8th grade	183,411	0.7	254	1.2
		9th grade	166,344	0.6	230	1.0
		10th grade	159,747	0.6	201	0.9
		11th grade	150,944	0.5	184	0.8
		12th grade	145,501	0.5	94	0.4
Private	All races	All grades	2,098,138	7.4	303	1.4
Total			**28,201,524***	**100.0**	**22,069**	**100.0**

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic

*This number is slightly different than the total eligible students on the frame because there were inconsistencies on the frame between the total school enrollment and enrollment in each grade.

Table 8: Calibration Distribution among Public/Private Schools by Grade

Public/ Private	Grade	Population		Respondents	
		n	%	n	%
Private	6th grade	309,848	1.1	35	0.2
	7th grade	309,321	1.1	71	0.3
	8th grade	308,041	1.1	58	0.3
	9th grade	297,510	1.1	95	0.4
	10th grade	296,857	1.1	18	0.1
	11th grade	291,318	1.0	14	0.1
	12th grade	285,243	1.0	12	0.1
Public	All grades	26,103,388	92.6	21,766	98.6

* Refers to students who identify with some other race not listed in the table and/or students who identify with more than one race.

4.3. Response Rates

4.3.1. School Response Rate

The 112 nonresponding schools were replaced with the next school on the frame, in the same PSU. When replacement schools are used, we calculate response rates two ways, (1) without counting the replacement school in the denominator of the response rate formula, and (2) with counting the replacement school in the denominator of the response rate formula.

4.3.1.1 Response Rate without Counting the Replacement School in the Denominator

There were 420 schools selected in 42 states. Four (1.0%) of the schools were deemed ineligible during recruitment. In total, 179 of the 416 (43.0%) eligible schools participated in the study. The remaining 237 schools were considered refusals. The school response rate was 43.0%. Of the 179 responding schools, 20 (11.2%) were recruited from the replacement sample. The 43.0% school-level response rate is the total number of responding schools divided by the number of eligible schools in the original sample. This calculation does not count replacement schools as new schools. The response rate calculation that does not include replacement school in the denominator is consistent with the previous NYTS response rate calculations that included replacement schools. Table 9 displays the components of the school response rate overall and by stratum, calculated without counting the replacement schools in the denominator..

4.3.1.2 Response Rate with Counting the Replacement School in the Denominator

Combining the original sample (420 schools) and the replacement sample (112 schools), the sample contained 532 schools. Four schools in the original sample and two schools in the replacement sample were found to be ineligible, resulting in 526 eligible schools. These schools were in 327 districts, and 31 (9.5%) of these districts refused to participate. District-level refusals resulted in refusals from 71 schools (13.5% of the 526 eligible schools). If the school response rate is calculated as the total number of responding schools divided by the sum of the number of eligible schools in the original sample and the eligible schools in the replacement sample, the school response rate is 33.6% (179/532 schools).

Table 9: Components of the School Response Rate Overall and by Stratum

Stratum	Schools Selected	Responding Schools	Nonresponding Schools	Ineligible Schools	Replacement Schools	School Response Rate
AI1	12	6	6	0	1	50.0
AI2	21	19	2	0	0	90.5
AS1	21	7	14	0	0	33.3
AS2	15	8	7	0	2	53.3
BRx	12	8	3	1	0	72.7
BU1	24	15	9	0	3	62.5
BU2	24	7	17	0	1	29.2
BU3	24	6	18	0	1	25.0
BU4	27	10	17	0	0	37.0
HR1	12	8	4	0	0	66.7
HR2	12	8	4	0	0	66.7
HR3	12	3	9	0	1	25.0
HR4	12	7	5	0	2	58.3
HU1	48	17	31	0	2	35.4
HU2	48	18	29	1	3	38.3
HU3	48	15	32	1	3	31.9
HU4	48	17	30	1	1	36.2
All Strata	**420**	**179**	**237**	**4**	**20**	**43.0**

4.3.2. Student Response Rate

Initial student-level participation rates were calculated from the field as teachers reported enrollment information and submitted surveys in the central repository. In subsequent follow-ups between teachers and SLs, 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.

According to enrollment estimates, 31,108 students were enrolled in the selected classrooms within responding schools. Of these, 22,069 students[17] completed the questionnaire. The overall student response rate was 70.9% (22,069 /31,108). Table 10 displays the components of the student response rate overall and by stratum.

Table 10: Components of the Student Response Rate Overall and by Stratum

Stratum	Students Selected	Responding Students	Student Response Rate
AI1	916	754	82.3
AI2	1,968	1,328	67.5
AS1	2,766	2,041	73.8
AS2	2,429	2,125	87.5
BRx	1,048	614	58.6
BU1	2,995	2,041	68.1
BU2	1,654	771	46.6
BU3	612	476	77.8
BU4	1,748	1,047	59.9
HR1	975	836	85.7
HR2	894	717	80.2
HR3	190	164	86.3
HR4	971	766	78.9
HU1	4,573	3,378	73.9
HU2	2,313	1,725	74.6
HU3	2,508	1,785	71.2
HU4	2,548	1,501	58.9
All Strata	**31,108**	**22,069**	**70.9**

4.3.3. Study Response Rate

The study response rate is 30.5%, using the school response rate that does not include the replacement schools in the denominator of the response rate formula. The study response rate is the product of the school response rate (43.0%) and the student response rate (70.9%), and this is consistent with the previous NYTS response rate calculations that included replacement schools. Table 11 displays the study response rate overall and by stratum using the school response rate that does not include replacement schools in the denominator.

[17] A complete (i.e., number of students who completed the survey) was defined as a student who started the survey, consented to participate, viewed at least the first three questions in the survey, and responded to at least one of the first three questions. This data cleaning practice is consistent with the procedures used to clean the 2021 and 2022 NYTS datasets. Of note, any records that were created by proctors testing the survey link were removed from the student-level dataset.

The study response rate is 23.8% (33.6% * 70.9%), using the school response rate that includes the replacement schools in the denominator of the response rate formula.

Table 11: Overall Response Rate and by Stratum

Stratum	School Response Rate	Student Response Rate	Study Response Rate
AI1	50.0	82.3	41.2
AI2	90.5	67.5	61.1
AS1	33.3	73.8	24.6
AS2	53.3	87.5	46.7
BRx	72.7	58.6	42.6
BU1	62.5	68.1	42.6
BU2	29.2	46.6	13.6
BU3	25.0	77.8	19.4
BU4	37.0	59.9	22.2
HR1	66.7	85.7	57.2
HR2	66.7	80.2	53.5
HR3	25.0	86.3	21.6
HR4	58.3	78.9	46.0
HU1	35.4	73.9	26.2
HU2	38.3	74.6	28.6
HU3	31.9	71.2	22.7
HU4	36.2	58.9	21.3
All Strata	**43.0**	**70.9**	**30.5**

4.4. Evaluate the Distribution of the Respondents

Table 12 displays the respondents by sex and grade. Of the respondents, 49.2% were female. There were fewer responding students in each successive grade.

Table 12: Respondents by Sex and Grade

Sex	Grade							
	6	7	8	9	10	11	12	Total
Female	2,004	1,845	1,724	1,736	1,408	1,310	834	10,861
Male	2,019	1,866	1,682	1,902	1,512	1,294	933	11,208
Total	**4,023**	**3,711**	**3,406**	**3,638**	**2,920**	**2,604**	**1,767**	**22,069**

Table 13 displays the respondents by middle school/high school, race/ethnicity and rural status.

Table 13:Respondents by Middle School/High School, Race/Ethnicity Category, and Rural Status

School Type	NH White	Nh Black	Hispanic	NH Asian	NH AI/AN	Rural	Total
Middle school	4,485	1,654	2,603	1,082	527	789	11,140
High school	3,984	1,097	3,738	995	390	725	10,929
Total	**8,469**	**2,751**	**6,341**	**2,077**	**917**	**1,514**	**22,069**

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic

4.5. Guidance on Making Population Estimates

Weighted estimates of means, percentages and totals can be computed using common statistical software packages such as SAS, SUDAAN, or R. When calculating population estimates, the user needs to include the design variables, stratum, and cluster, as well as the analysis weights. The following are the variables on the analysis dataset used for estimating population parameters:

STRATUM—This variable identifies the sampling stratum. Not including stratification in the analysis will not produce incorrect estimates. However, including stratum generally reduces the variance of the estimates slightly.

PSU—This variable identifies the clusters used in the analysis. If the clustering variable is omitted variances will be underestimated.

WT_ANALYSIS—This variable contains the weight of each responding student. The respondent's weight reflects the number of population members the respondent represents. Omitting the weight in the analysis will provide estimates of the characteristics of the set of respondents, not the population members (e.g., the estimates will not reflect the population).

The following is example SAS code:

```
proc surveyfreq data=work.dataset;
 stratum stratum;
 cluster PSU;
 weight WT_analysis;
 tables outcome_1*outcome_2/row cl;
 ods output CrossTabs=p;
run;
```

4.6. QC Checks Performed During the Creation of the Sampling Weights

1) Confirm that all students come from a school that matches the sample frame.

2) Identify records on the student file that were not valid.

   Some of the records on the student file were school employees evaluating the content of the questionnaire. There was one school with one responding questionnaire that we confirmed was not a legitimate student respondent. If we did not remove that one student record, the school would have been considered a responding school, and the one student would have a very large weight to account for the large amount of student nonresponse.

3) Evaluate the school's disposition codes.

   The disposition code identifies if the school is a responding school, nonresponding school, or ineligible school. Disposition codes are used by recruitment and data collection tasks to track school responses. Table 9 contains the counts of schools with each disposition by stratum. This table was used to evaluate the correctness of the assignment of disposition code to the schools.

4) Evaluate the independent variables used in the school-level nonresponse model.

   For the nonresponse model, we created the quintiles for the distributions: percentage of NH White, NH Black, Hispanic, NH Asian, and NH AI/AN students. Schools with missing values for race and ethnicity values were put into a separate category. We checked that each level of the variables contained approximately 20% of the schools after accounting for the missing values.

5) Check the class enrollment values.

   The class enrollment data come from three sources:

   (1) A school administrator when the classes were enumerated

   (2) The classroom teacher when the students are answering the questionnaire

   (3) The count of responding students in a class

   If (3) is greater than (1) or (2), the class size is initialized to (3) because the class size has to be at least as large as the number of responding students. Otherwise, if (2) is equal to or greater than (3), the class size is initialized to (2) because we assume that the teacher specified class size at the time of survey administration is more accurate than the class size provided by the administrator when the class list was provided. The reason for this assumption is that teacher-reported enrollment is provided closer to the actual date of the survey than administrator-reported enrollment. Also, because they interact more closely with students, teachers are more likely to have an accurate count of students in their class than administrators. If (2) is missing and (1) is greater or equal to (3), we use (1). For the one school that did not provide class size ((1) and (2) were missing), we imputed the class size to 1.25 times the number of responding students. We printed out a class-level file, organized by school, with all the relevant data that was used to initialize class size. We reviewed this file to identify any suspicious class sizes. During this review, we found a few cases in which one class ID was used for several classes. For these classes, we recoded class enrollment to equal the sum of the enrollment for all affected classes divided by the number of affected classes.

6) Evaluate the weight after adjusting for class nonresponse.

   Within a school, responding students in each class received a different weight because the proportion of responding students differs by class. We printed out a class-level file, organized by school, with all the relevant data that was used to adjust the weight applied to each class. We reviewed this file to identify large differences in weights applied to the different classes within a school and confirmed that these differences accurately reflect the differential class response rates.

7) Check that the variables used in calibration are nonmissing.

   First, we identified if the components of the variables used in calibration contained missing values. Next, we imputed for these values as described in the calibration section. Then, we checked the imputed values for missing values.

8) Checked that the coding of the calibration variables matches the order of the population totals listed in the procedure that implements the calibration.

9) Evaluate the application of the calibration.

   Tables 6, 7, and 8 contain the population totals for each calibration category. The totals in these tables were generated from the macro variables used for the constraints in the calibration procedure. To check if the calibration procedure correctly worked, the population totals were recreated by calculating the weighted cross-tabulations of the calibration distributions. These totals matched the inputs to the calibration procedure, demonstrating that the calibration worked correctly.

10) Evaluate the number of respondents in a PSU.

   The data are clustered by school because of the study design. To reduce disclosure risk, on the analysis file, each PSU contains groupings of schools. We checked that the number of students in each PSU was sufficiently large enough to ensure that estimating the interclass correlation within PSU was possible; this was part of creating estimates of precision.

11) Evaluate the unequal weighting effect.

   We calculated the unequal weighting effect for the entire population. It is 3.9.

12) We examined the distribution of the weights and confirmed there were no missing weights.

Table 14 shows the distribution of the calculated sampling weights.

Table 14: Distribution of Sampling Weights by Percentile

Minimum	Percentile					Maximum
	10%	25%	50%	75%	90%	
12.2	98.5	209.7	556.3	1,508.3	3,052.47	60,800.12

13) Evaluate the number of classes by grade selected in each school.

If a school has the full complement of grades, then the protocol calls for either one or two classes per grade to be selected. However, the protocol also allows for schools to provide a census of classes. We evaluated the distribution of classes selected in each grade for each school.

14) Evaluate the variance of the estimates.

Tables 15 through 19 contains ever and current tobacco use estimates for 12 tobacco products for high school and middle school students in eight domains (overall, by sex, and by race/ethnicity). For the most part, these tables duplicate the estimates presented in the 2022 methodology report. We evaluated the standard errors of the estimates. And, when applicable, we compared them to the identical estimate in the 2022 report. We concluded that the level of precision of estimates calculated from the 2023 data is comparable to the level of precision of estimates calculated from the 2022 data. Because fewer schools and students responded to the 2023 NYTS than to the 2022 NYTS, achieving the same level of precision for 2023 NYTS estimates as was achieved for 2022 NYTS estimates is demonstrative of the small improvements in the sampling and weighting procedures that were made between the 2022 and 2023 survey administrations. For additional information on considerations when comparing 2022 and 2023 estimates, please see Section 4.7.

4.7 Guidance on Making comparisons between 2022 and 2023

Response rates for the 2023 NYTS were lower than response rates for the 2022 NYTS. As a result, fewer schools (179 in 2023 vs. 341 in 2022) and fewer students (22,069 in 2023 vs. 28,291 in 2022) responded to the 2023 NYTS than to the 2022 NYTS. The precision of the 2023 NYTS estimates is lower than what would have been obtained if the yield rates assumed during the design of the sample were met. Less precise estimates reduce power to identify differences between specific populations (e.g., 10th graders vs. 12th graders) and across data collection years (e.g., 2022 vs. 2023). The 2023 NYTS weighting methodology accounts for the disproportionate sampling of specific groups (e.g., the oversample of schools with high densities of NH Asian and NH AI/AN students) and includes adjustments designed to reduce nonresponse bias. However, because the 2023 NYTS yielded fewer completed surveys than the 2022 NYTS, it is possible that some groups may have been too small in 2023 to compare their estimates with 2022 findings. In NYTS publications, estimates are considered unreliable and suppressed if the relative standard error (RSE, defined as the ratio of the standard error over the estimate) is greater than 30% or an unweighted denominator is fewer than 50. Furthermore, because of changes to the 2023 NYTS sampling, the degrees of freedom used to calculate confidence intervals and *p* values for comparisons in 2023 were 38 compared to 136 in 2022. When comparing estimates between 2022 and 2023, the smaller degrees of freedom of 38 were used in a report published in November 2023[18], resulting in slightly wider confidence intervals and somewhat larger *p* values.

4.8 Tables of Population Estimates

The following four tables display current and ever tobacco use estimates, separately for middle school and high school students overall, by sex, and by race/ethnicity. In these tables, the value *n* is the number of students in that domain that have the associated tobacco use characteristic.

[18] Birdsey J, Cornelius M, Jamal A, et al. Tobacco Product Use Among US Middle and High School Students — National Youth Tobacco Survey, 2023. MMWR Morb Mortal Wkly Rep 2023;72:1173–1182. DOI: http://dx.doi.org/10.15585/mmwr.mm7244a1

Table 15: Current (Past 30-Day) Use Estimates for Selected Tobacco Products among High School Students[19]

Product	Overall % (SE) n	Female % (SE) n	Male % (SE) n	NH White % (SE) n	NH Black % (SE) n	Hispanic % (SE) n	NH Asian % (SE) n	NH AI/AN % (SE) n
Bidis	0.54% (0.12%) n=69	0.44% (0.15%) n=27	0.64% (0.12%) n=42	0.27% (0.08%) n=17	1.16% (0.47%) n=15	0.86% (0.25%) n=31	0.11% (0.09%) n=2	0.33% (0.20%) n=4
Cigars, little cigars, or cigarillos	1.84% (0.23%) n=219	1.39% (0.35%) n=74	2.29% (0.32%) n=145	1.50% (0.30%) n=91	1.93% (0.44%) n=23	2.28% (0.43%) n=89	0.37% (0.30%) n=4	1.36% (0.89%) n=8
Cigarettes	1.90% (0.23%) n=231	1.48% (0.29%) n=101	2.31% (0.28%) n=129	2.19% (0.44%) n=104	0.65% (0.17%) n=15	2.15% (0.34%) n=81	1.46% (1.10%) n=4	1.29% (0.53%) n=22
Nicotine pouches	1.73% (0.33%) n=179	0.79% (0.32%) n=41	2.65% (0.42%) n=138	2.29% (0.53%) n=106	0.78% (0.36%) n=9	1.58% (0.41%) n=47	0.47% (0.30%) n=5	0.59% (0.34%) n=8
Electronic cigarettes	10.03% (0.65%) n=1,0	12.24% (1.06%) n=610	7.95% (0.91%) n=477	11.63% (1.09%) n=538	5.75% (0.77%) n=87	9.75% (0.94%) n=355	8.69% (5.63%) n=37	6.88% (2.38%) n=59
Hookah or waterpipe	1.13% (0.19%) n=122	1.37% (0.39%) n=56	0.91% (0.22%) n=66	0.97% (0.27%) n=46	1.52% (0.62%) n=17	0.99% (0.23%) n=43	0.30% (0.17%) n=6	0.51% (0.28%) n=6
Pipe tobacco	0.61% (0.11%) n=82	0.53% (0.13%) n=35	0.68% (0.20%) n=47	0.63% (0.15%) n=33	0.13% (0.06%) n=4	1.00% (0.32%) n=36	N/A (N/A) n=0	0.49% (0.25%) n=7
Roll-your-own cigarettes	0.77% (0.20%) n=78	0.90% (0.35%) n=32	0.65% (0.18%) n=46	0.39% (0.13%) n=27	0.79% (0.68%) n=5	1.10% (0.29%) n=36	0.03% (0.03%) n=1	0.44% (0.23%) n=6
Smokeless tobacco (chewing tobacco, snuff, or dip)	0.89% (0.17%) n=126	0.34% (0.10%) n=24	1.41% (0.30%) n=102	1.16% (0.28%) n=66	0.37% (0.12%) n=9	0.91% (0.21%) n=42	0.13% (0.09%) n=3	0.43% (0.21%) n=5
Snus	1.06% (0.26%) n=82	0.81% (0.38%) n=21	1.31% (0.31%) n=61	1.05% (0.27%) n=39	0.35% (0.21%) n=5	1.21% (0.39%) n=29	0.10% (0.09%) n=2	0.36% (0.21%) n=5
Oral nicotine	1.22% (0.15%) n=159	0.93% (0.16%) n=61	1.50% (0.28%) n=98	1.32% (0.21%) n=78	0.81% (0.31%) n=13	1.56% (0.26%) n=56	0.12% (0.07%) n=5	0.52% (0.25%) n=6
Heated tobacco products	1.04% (0.22%) n=104	0.70% (0.18%) n=42	1.39% (0.40%) n=62	0.98% (0.28%) n=37	0.84% (0.32%) n=13	1.57% (0.42%) n=45	0.15% (0.10%) n=4	0.34% (0.23%) n=4

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic; SE = standard error.

[19] The estimates in tables 15 through 18 use the variable SCHOOLTYPE, which is coded as 1 (Middle School) if QN3 ranges from 1 to 3, and 2 (High School) if QN3 ranges from 4 to 7. Students who are missing QN3 are excluded from these tables. Non-Hispanic Other Race participants (participants who identified as non-Hispanic Native Hawaiian or other Pacific Islander) and non-Hispanic multiracial respondents are excluded from these tables.

Note: Estimates are presented regardless of whether data presentation standards (RSE >30% or unweighted denominator <50) are met and are not intended for use beyond this report. In the dataset, variables associated with current use of each tobacco product are as follows: electronic cigarettes (celcigt); cigars, little cigars, or cigarillos (ccigar); cigarettes (ccigt); smokeless tobacco (cslt); hookah or waterpipe (chookah); roll-your-own cigarettes (crollcigts); snus (csnus); pipe tobacco (cpipe); nicotine pouches (cpouch); bidis (cbidis), oral nicotine (coral), and heated tobacco products (chtp).

Table 16: Current (Past 30-Day) Use Estimates for Selected Tobacco Products among Middle School Students

Product	Overall % (SE) n	Female % (SE) n	Male % (SE) n	NH White % (SE) n	NH Black % (SE) n	Hispanic % (SE) n	NH Asian % (SE) n	NH AI/AN % (SE) N
Bidis	0.50% (0.18%) n=52	0.50% (0.24%) n=31	0.51% (0.17%) n=21	0.05% (0.03%) n=6	1.92% (0.80%) n=15	0.57% (0.23%) n=23	N/A (N/A) n=0	0.23% (0.10%) n=5
Cigars, little cigars, or cigarillos	1.13% (0.28%) n=106	1.23% (0.36%) n=50	1.04% (0.29%) n=55	0.36% (0.13%) n=22	2.71% (0.91%) n=28	1.83% (0.47%) n=44	N/A (N/A) n=0	1.16% (1.00%) n=6
Cigarettes	1.06% (0.31%) n=131	1.15% (0.29%) n=72	0.99% (0.38%) n=58	0.74% (0.21%) n=35	0.83% (0.46%) n=11	1.81% (0.60%) n=59	0.02% (0.02%) n=1	2.29% (1.13%) n=18
Nicotine pouches	1.08% (0.41%) n=82	0.76% (0.27%) n=31	1.36% (0.59%) n=48	0.44% (0.11%) n=25	1.27% (0.52%) n=10	2.10% (1.11%) n=28	N/A (N/A) n=0	2.29% (1.23%) n=11
Electronic cigarettes	4.57% (0.53%) n=458	5.64% (0.66%) n=277	3.49% (0.57%) n=177	3.20% (0.48%) n=153	5.48% (1.00%) n=65	6.64% (0.70%) n=165	1.26% (0.76%) n=6	4.56% (2.31%) n=46
Hookah or waterpipe	1.03% (0.27%) n=96	1.23% (0.45%) n=55	0.84% (0.25%) n=41	0.40% (0.11%) n=26	1.90% (0.88%) n=18	1.78% (0.43%) n=42	N/A (N/A) n=0	0.59% (0.37%) n=6
Pipe tobacco	0.37% (0.08%) n=46	0.38% (0.16%) n=25	0.36% (0.11%) n=21	0.24% (0.09%) n=17	0.59% (0.40%) n=2	0.57% (0.20%) n=19	0.04% (0.04%) n=1	1.35% (1.04%) n=7
Roll-your-own cigarettes	0.72% (0.13%) n=82	0.91% (0.24%) n=48	0.55% (0.12%) n=34	0.28% (0.09%) n=20	0.98% (0.47%) n=13	1.42% (0.40%) n=42	0.73% (0.75%) n=1	1.14% (1.03%) n=5
Smokeless tobacco (chewing tobacco, snuff, or dip)	0.67% (0.18%) n=73	0.53% (0.14%) n=29	0.81% (0.30%) n=43	0.46% (0.17%) n=28	0.85% (0.39%) n=6	1.12% (0.37%) n=31	N/A (N/A) n=0	0.19% (0.14%) n=4
Snus	0.35% (0.11%) n=36	0.26% (0.14%) n=10	0.44% (0.19%) n=25	0.14% (0.06%) n=8	0.77% (0.40%) n=5	0.60% (0.26%) n=19	N/A (N/A) n=0	1.19% (1.04%) n=4
Oral nicotine	1.06% (0.16%) n=109	1.30% (0.22%) n=62	0.84% (0.17%) n=47	0.94% (0.24%) n=37	1.10% (0.58%) n=9	1.30% (0.28%) n=43	0.81% (0.74%) n=4	0.52% (0.21%) n=10
Heated tobacco products	0.83% (0.20%) n=87	0.75% (0.21%) n=35	0.92% (0.36%) n=51	0.46% (0.15%) n=23	1.03% (0.46%) n=10	1.34% (0.36%) n=39	0.03% (0.03%) n=1	0.76% (0.42%) n=8

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic; SE = standard error.

Note: Estimates are illustrative and are presented regardless of whether data presentation standards (RSE >30% or unweighted denominator <50) are met .In the dataset, variables associated with current use of each tobacco product are as follows: electronic cigarettes (celcigt); cigars, little cigars, or cigarillos (ccigar); cigarettes (ccigt); smokeless tobacco (cslt); hookah or waterpipe (chookah); roll-your-own cigarettes (crollcigts); snus (csnus); pipe tobacco (cpipe); nicotine pouches (cpouch); bidis (cbidis), oral nicotine (coral), heated tobacco products (chtp).

Table 17: Ever Use Estimates for Selected Tobacco Products among High School Students

Product	Overall % (SE) n	Female % (SE) n	Male % (SE) n	NH White % (SE) n	NH Black % (SE) n	Hispanic % (SE) n	NH Asian % (SE) n	NH AI/AN % (SE) n
Bidis	1.49% (0.26%) n=142	1.45% (0.33%) n=65	1.55% (0.41%) n=77	1.19% (0.40%) n=40	2.12% (0.66%) n=20	2.09% (0.40%) n=62	0.16% (0.11%) n=3	1.00% (0.29%) n=13
Cigars, little cigars, or cigarillos	6.38% (0.59%) n=644	4.80% (0.70%) n=224	7.93% (0.87%) n=418	8.07% (0.97%) n=337	4.65% (0.95%) n=54	5.41% (0.55%) n=201	0.77% (0.38%) n=13	4.87% (2.16%) n=26
Cigarettes	8.54% (0.46%) n=975	8.81% (0.81%) n=493	8.27% (0.68%) n=477	10.57% (0.70%) n=505	3.22% (0.80%) n=53	8.78% (1.03%) n=305	4.11% (1.84%) n=26	12.29% (4.40%) n=66
Nicotine pouches	43.26% (2.10%) n=3,3	42.00% (2.02%) n=1,5	44.30% (2.81%) n=1,8	57.29% (1.88%) n=2,0	24.71% (2.31%) n=209	28.33% (1.56%) n=774	28.36% (6.70%) n=190	39.06% (7.24%) n=115
Electronic cigarettes	22.61% (0.89%) n=2,3	25.98% (1.43%) n=1,2	19.47% (0.97%) n=1,0	26.41% (1.51%) n=1,1	14.49% (1.59%) n=187	22.34% (1.23%) n=809	14.61% (5.50%) n=104	19.84% (4.49%) n=106
Hookah or waterpipe	3.74% (0.56%) n=335	4.02% (0.77%) n=172	3.50% (0.74%) n=162	3.64% (0.63%) n=130	4.37% (1.33%) n=46	3.88% (0.68%) n=117	1.12% (0.59%) n=17	4.16% (2.39%) n=18
Pipe tobacco	2.08% (0.19%) n=215	1.75% (0.33%) n=94	2.41% (0.35%) n=121	2.73% (0.33%) n=107	0.50% (0.17%) n=12	2.18% (0.40%) n=71	0.19% (0.12%) n=4	3.64% (1.84%) n=16
Roll-your-own cigarettes	2.54% (0.29%) n=276	2.78% (0.45%) n=135	2.31% (0.40%) n=140	2.76% (0.50%) n=123	1.56% (0.61%) n=15	2.52% (0.36%) n=103	0.14% (0.09%) n=6	8.17% (4.15%) n=22
Smokeless tobacco (chewing tobacco, snuff, or dip)	2.46% (0.33%) n=322	1.35% (0.17%) n=94	3.53% (0.58%) n=227	3.40% (0.53%) n=181	0.79% (0.22%) n=16	1.93% (0.39%) n=86	0.74% (0.39%) n=9	7.11% (3.58%) n=23
Snus	1.71% (0.29%) n=178	1.16% (0.38%) n=49	2.24% (0.43%) n=128	1.93% (0.35%) n=100	0.54% (0.24%) n=9	1.63% (0.43%) n=45	0.25% (0.15%) n=4	5.37% (3.61%) n=13
Oral nicotine	3.45% (0.33%) n=414	2.85% (0.48%) n=166	4.04% (0.54%) n=246	4.18% (0.55%) n=215	1.63% (0.43%) n=25	3.79% (0.43%) n=132	0.49% (0.28%) n=15	6.16% (2.51%) n=22
Heated tobacco products	1.64% (0.28%) n=175	1.72% (0.32%) n=93	1.57% (0.33%) n=81	1.86% (0.37%) n=75	2.01% (0.66%) n=19	1.45% (0.33%) n=58	0.31% (0.15%) n=11	2.69% (1.63%) n=11

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic; SE = standard error.
Note: Estimates are presented regardless of whether data presentation standards (RSE >30% or unweighted denominator <50) are met and are not intended for use beyond this report. In the dataset, variables associated with ever use of each tobacco product are as follows: electronic cigarettes (eelcigt); cigars, little cigars, or cigarillos (ecigar); cigarettes (ecigt); smokeless tobacco (eslt); hookah or waterpipe (ehookah); roll-your-own cigarettes (erollcigts); snus (esnus); pipe tobacco (epipe); nicotine pouches (epouch); bidis (ebidis), oral nicotine (eoral), and heated tobacco products (ehtp).

Table 18: Ever Use Estimates for Selected Tobacco Products among Middle School Students

Product	Overall % (SE) n	Female % (SE) n	Male % (SE) n	NH White % (SE) n	NH Black % (SE) n	Hispanic % (SE) n	NH Asian % (SE) n	NH AI/AN % (SE) n
Bidis	1.25% (0.34%) n=114	0.94% (0.32%) n=61	1.58% (0.58%) n=53	0.66% (0.25%) n=35	2.35% (0.91%) n=24	2.01% (0.56%) n=43	0.13% (0.13%) n=1	0.33% (0.15%) n=7
Cigars, little cigars, or cigarillos	2.65% (0.45%) n=266	2.40% (0.50%) n=111	2.93% (0.51%) n=154	1.84% (0.35%) n=88	4.38% (0.99%) n=57	3.53% (0.71%) n=87	0.13% (0.10%) n=3	2.19% (1.16%) n=19
Cigarettes	4.29% (0.54%) n=491	4.64% (0.55%) n=270	4.00% (0.78%) n=220	3.63% (0.62%) n=184	5.75% (1.05%) n=71	5.29% (0.82%) n=160	2.03% (1.16%) n=6	5.76% (1.81%) n=52
Nicotine pouches	35.24% (1.67%) n=2,8	33.86% (1.99%) n=1,3	36.72% (2.09%) n=1,5	43.70% (2.09%) n=1,5	23.52% (2.65%) n=295	29.40% (2.31%) n=547	25.39% (2.99%) n=208	36.22% (9.66%) n=126
Electronic cigarettes	9.70% (0.72%) n=993	11.01% (1.04%) n=544	8.21% (0.72%) n=442	8.59% (0.85%) n=370	10.59% (1.32%) n=150	12.35% (0.97%) n=326	2.89% (1.05%) n=19	9.59% (2.94%) n=85
Hookah or waterpipe	2.11% (0.44%) n=211	2.58% (0.79%) n=120	1.67% (0.27%) n=91	1.15% (0.28%) n=58	4.16% (1.59%) n=46	2.92% (0.45%) n=79	0.25% (0.12%) n=4	1.10% (0.45%) n=12
Pipe tobacco	1.09% (0.20%) n=139	1.08% (0.32%) n=68	1.11% (0.31%) n=71	0.75% (0.23%) n=54	1.41% (0.70%) n=11	1.71% (0.29%) n=49	0.06% (0.05%) n=2	2.33% (1.16%) n=20
Roll-your-own cigarettes	1.55% (0.23%) n=166	1.96% (0.41%) n=94	1.16% (0.21%) n=72	0.87% (0.17%) n=51	2.06% (0.85%) n=25	2.73% (0.44%) n=68	0.75% (0.75%) n=2	2.58% (1.38%) n=14
Smokeless tobacco (chewing tobacco, snuff, or dip)	2.16% (0.39%) n=241	2.08% (0.38%) n=108	2.28% (0.54%) n=132	2.78% (0.58%) n=129	1.49% (0.48%) n=17	2.08% (0.59%) n=65	0.25% (0.16%) n=5	0.94% (0.35%) n=14
Snus	0.95% (0.16%) n=83	0.71% (0.18%) n=25	1.19% (0.30%) n=57	0.64% (0.18%) n=29	1.28% (0.54%) n=10	1.45% (0.38%) n=31	0.55% (0.52%) n=2	3.55% (2.09%) n=9
Oral nicotine	2.24% (0.23%) n=275	2.41% (0.32%) n=146	2.09% (0.27%) n=128	2.03% (0.34%) n=99	2.00% (0.73%) n=27	2.87% (0.62%) n=100	1.67% (0.95%) n=10	3.23% (1.26%) n=26
Heated tobacco products	1.25% (0.24%) n=121	1.16% (0.29%) n=51	1.36% (0.43%) n=70	0.87% (0.23%) n=43	1.10% (0.51%) n=10	2.13% (0.29%) n=50	0.22% (0.13%) n=3	1.36% (0.71%) n=9

Abbreviations: AI/AN = American Indian/Alaska Native; NH = Non-Hispanic; SE = standard error.
Note: Estimates are presented regardless of whether data presentation standards (RSE >30% or unweighted denominator <50) are met and are not intended for use beyond this report. In the dataset, variables associated with ever use of each tobacco product are as follows: electronic cigarettes (eelcigt); cigars, little cigars, or cigarillos (ecigar); cigarettes (ecigt); smokeless tobacco (eslt); hookah or waterpipe (ehookah); roll-your-own cigarettes (erollcigts); snus (esnus); pipe tobacco (epipe); nicotine pouches (epouch); bidis (ebidis), oral nicotine (eoral), and heated tobacco products (ehtp).

APPENDIX A: COMMON CORE OF DATA RACE AND ETHNICITY DEFINITIONS

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

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

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

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

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