

2026 NYTS Sampling Plan

This document provides the sampling plan for the 2026 NYTS.

The objective of the NYTS sampling design is to support estimation of tobacco-related knowledge, attitudes, and behaviors in a national population of public and private school students enrolled in grades 6 through 12 in the United States. More specifically, the study 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 white, non-Hispanic Black, Hispanic, non-Hispanic Asian, and non-Hispanic American Indian / Alaskan Native). Additional estimates also are supported for subgroups defined by grade, by sex, and by race-ethnicity, each within school-level domains; however, precision levels vary considerably according to differences in subpopulation sizes.

The NYTS employs a repeat cross-sectional design with independent yearly samples and data collection.

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

We will implement a three-stage, clustered design that will obtain a nationally representative sample of U.S. students in grades 6 through 12. The sample will yield an equal probability sample within strata.

Participating students complete the anonymous voluntary survey using a self-administered questionnaire.

1. Frame Construction

The frame used to select the 2026 NYTS sample will combine data files obtained from the National Center for Education Statistics (NCES) and MDR Inc. (Market Data Retrieval Inc.). The NCES data come from two sources, the Common Core of Data (CCD) for public schools and the Private School Survey (PSS) for non-public schools. Private school participation in the PSS is optional. The response rate for the 2020 PSS was 74.5% in 2020.¹ Consequently, 25.5% of the private schools are not represented in the PSS. To reduce coverage error, private schools in the MDR data that were not in the PSS will be included in the 2026 NYTS frame.

The 2026 NYTS frame will be subset to all public school (including charter) and private school students enrolled in regular middle schools and high schools in grades 6 through 12 in the 50 U.S. states and the District of Columbia. The following school types were excluded: alternative schools, special education schools, Department of Defense–operated schools, Bureau of Indian Affairs schools, adult education schools, and vocational schools.

A cut-off in school size will be implemented. Only schools with an enrollment of at least 40 students across the eligible grades will be eligible to be sampled. This choice excludes less than 1% of

¹ Broughman, S.P., Kincel, B., and Peterson, J. (2021). Private School Universe Survey (PSS): Public-Use Data File User's Manual for School Year 2019–20 (NCES 2022-021). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022021>.

otherwise eligible students.

2. Stratification

The frame will be partitioned into 7 strata.

Stratum 1—Schools with 20% to 40% non-Hispanic American Indian / Alaskan Native (hereafter, AI/AN) students.

Stratum 2—Schools with more than 40% AI/AN students.

Stratum 3—Schools with 20% to 40% non-Hispanic Asian (hereafter, Asian) students.

Stratum 4—Schools with more than 40% Asian students.

Stratum 5—All other schools with only middle school students

Stratum 6—All other schools with only high school students

Stratum 7—All other schools with both middle school and high school students

The stratification enables the control of the allocation of high school and middle school students by manipulating the sampling fraction across strata. This is needed because, on average, middle schools have fewer students than high schools. The stratification also enables oversampling of AI/AN and Asian students to achieve the goal of collecting data on a minimum of 1,000 responding AI/AN and Asian students (to yield sufficiently precise estimates for these groups). We anticipate that there will be enough Black, Hispanic, and rural students to make precise estimates without manipulating the sample allocation through oversampling.

3. Formation of Primary Sampling Units

Primary sampling units (PSUs) are groups of schools within strata.

In strata 1 to 4, the PSUs are formed by sorting the schools by state Federal Information Processing Standards (FIPS) and county FIPS code and partitioning the list of schools into groups with about 6,000 students per group. PSUs must have at least 10 schools.

In strata 5 to 7, the formation of the PSUs differs for each of the following 3 scenarios.

Scenario 1—In counties with more than 100,000 students, the schools are sorted by zipcode and partitioned into multiple PSUs containing between 50,000 and 100,000 students each.

Scenario 2—In counties with fewer than 2,500 eligible students, the schools are joined with the schools in other small counties within the same state to form PSU's with at least 2,500 eligible students.

Scenario 3—In all other counties, the group of schools in the county form a unique PSU.

4. Sampling Stages

Overview of the stages of selection

Stage 1— Select 182 PSUs.

Stage 2— Select 546 Schools. Three schools are selected within each of the 182 PSUs.

Stage 3— Select classes within schools.

Not considered a sampling stage— Select all students within each class.

Stage one—selecting the PSUs

182 PSUs are selected, within 7 strata, using a systematic selection with probability proportional to size (PPS). The size measure is the number of students in the PSU. Within strata, the PSUs are sorted by state FIPS code and county FIPS code, ensuring geographic diversity in the sample. The sampling fractions, the probability of selection of each PSU within a stratum, are adjusted to ensure the number of responding AI/AN and Asian students is a minimum of 1,000, and that approximately three-sevenths of the responding students are middle school students.

Table 1 presents, by stratum and overall, the approximate counts of PSUs, schools, and students on the frame and selected for participation in the study. These counts are approximate because they are based on the 2025 NYTS frame, and the frame will be updated for the 2026 data collection. The sample size was based on the following assumptions: 48.3% school response rate, 98.0% school eligibility rate, and 78.3% student response rate (this results in an overall response rate of 38%, see calculations in the sample size calculation).

Table 1. Approximate Counts of PSUs, Schools, and Students on the Frame and Selected for Participation in the Study by Stratum

Stratum	Frame counts			Sample selected		
	PSUs	Schools	Students	PSUs	Schools	Students
1) AI/AN 20 to 40%	19	422	98,334	5	15	1,884
2) AI/AN more than 40%	16	478	85,352	9	27	3,102
3) Asian 20 to 40%	140	1,574	1,191,736	9	27	4,147
4) Asian more than 40%	61	666	535,690	7	21	3,317
5) Other schools with only middle school students	1,199	30,942	9531,824	56	168	23,452
6) Other schools with only high school students	883	14,484	12,770,235	74	222	34,492
7) Other schools with both middle and high school students	763	11,449	3,766,040	22	66	9,499
Total	3,081	60,015	27,979,211	182	546	79,893

Second Stage—selecting the schools²

Within each selected PSU, schools are selected using systematic sampling with a random start and probabilities proportional to size (PPS). Probabilities of school selection are proportional to a measure of size (MOS) that is based on the student enrollment for each school. Except for very large and very small schools, the measure of size is exactly equal to the enrollment in the target grades. An initial sampling

² Much of the methodology described in the section “Second Stage—selecting the schools” has been adapted from the State YTS Methodology Report: Office on Smoking and Health. State Youth Tobacco Survey (YTS) Methodology Report. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2018.

interval is calculated for each PSU by dividing the sum of the enrollments of all the schools in each PSU by the number of schools selected in the PSU. Schools that have an enrollment greater than the sampling interval are treated as certainty schools and are removed from the sample frame. Certainty schools are always selected for the sample. Each time a school is selected with certainty and removed from the frame, the sampling interval is recomputed based on the enrollment of the schools remaining on the sampling frame and on the number of schools remaining to be selected. There were 4 certainty schools in the 2025 sample.

The sampling procedure for the schools includes adjustments to the measure of size for schools that have small enrollments. This procedure ensures that, within each stratum, each student has the same probability of selection for the sample and that this probability is equal to the overall sampling rate in that stratum. Schools are selected using systematic sampling. Probabilities for these schools are proportional to school enrollment.

Selecting schools—Selection of schools is carried out using systematic sampling with a random start and an adjusted school sampling interval that uses a total enrollment based on the revised size measure. The selection procedure uses implicit stratification based on state FIPS code and county FIPS code. This procedure ensures the geographical diversity of the selected schools.

Third stage—Class selection

In the third stage classes are selected. To select classes, within each selected school, the probability of each student being selected is calculated. This probability is defined as the overall sampling rate divided by the product of the probability of the first and second stage of selection. Consequently, within each stratum, every sample member has an equal probability of selection. Within each school, the class sampling interval is defined as the inverse of the stage 3 probability of selection.

Within each selected school, the class sampling interval is applied to a random start. For example, for a school with 480 students and a class sampling interval of 4, the random start is a random number between 1 and the sampling interval (4), say, 3. The following classes are selected for the survey: 3, 7, 11, 15, ($3 + 4 = 7$, $7 + 4 = 11$, etc.) The school coordinator orders the classes in all grades. If there is a total of 8 classes, the 3rd and 7th classes are selected based on the ordering. It is expected that only a portion of the total list of classes will exist and be available. The number of classes in a school is unknown before contacting the school, so, the sampling team will pick a larger number of classes than will be used.

Not considered a sampling stage—Student selection within class

Select all students in each selected class.

5. Sample Sizes

Summary

We will sample 546 schools with the expectation of obtaining between 28,000 and 30,000 responding students in between 250 and 285 responding schools. We will have enough precision to make estimate to domains with 1,000 students. All the high priority domains will have 1,000 responding students.

Domains of interest

The NYTS is designed to produce reliable estimates of study outcomes for the following key subgroup:

- School Type: middle and high school students,
- Grade: individual grades 6, 7, 8, 9, 10, 11, and 12,
- Sex: males and females
- Sex by school type: male middle school students, female middle school students, etc.,
- Sex by grade: sixth-grade males, sixth-grade females, etc.,
- Race/Ethnicity: White, Black, Hispanic, Asian, AI/AN
- Race/Ethnicity by school level: White middle school students, Black middle schools students, etc.,
- Rural by schools type: rural middle school students, rural high school students.

Precision criteria

The relative standard error (RSE) is the ratio of the standard error of the estimate and the estimate. In the following text we express RSE as a percentage by multiplying the RSE by 100. Estimates are considered precise enough to be published if they have an RSE less than 30%.

The RSE is a function of 3 characteristics:

1. the sample size in the domain to which the estimate applies,
2. the design effect, and
3. the proportion of the estimate.

The main cause of the design effect is the interclass correlation in the PSU—students in the same PSU have similar outcomes. The interclass correlation can vary a lot between different estimates. To accommodate this variation, measures of precision are calculated for a range of design effects.

Table 2 displays the RSE for different combinations of design effect, sample size, and estimate. As you can see on Table 2, estimates of proportions of 10% with 300 or less respondents in a domain and a design effect of 3 are considered too imprecise to be published and require suppression. Other combinations of design effect, sample size, and proportion that that require suppression are indicated in red in Table 2.

Table 2. Relative Standard Error for different combinations of design effect, sample size and estimate

Design effect	n	Estimate				
		10%	20%	30%	40%	50%
2	300	24.5	16.3	12.5	10.0	8.2
	500	19.0	12.6	9.7	7.7	6.3
	1,000	13.4	8.9	6.8	5.5	4.5
	2,000	9.5	6.3	4.8	3.9	3.2
	3,000	7.7	5.2	3.9	3.2	2.6
3	300	30.0	20.0	15.3	12.2	10.0
	500	23.2	15.5	11.8	9.5	7.7
	1,000	16.4	11.0	8.4	6.7	5.5
	2,000	11.6	7.7	5.9	4.7	3.9
	3,000	9.5	6.3	4.8	3.9	3.2
4	300	34.6	23.1	17.6	14.1	11.5
	500	26.8	17.9	13.7	11.0	8.9
	1,000	19.0	12.6	9.7	7.7	6.3

Design effect	n	Estimate				
		10%	20%	30%	40%	50%
5	2,000	13.4	8.9	6.8	5.5	4.5
	3,000	11.0	7.3	5.6	4.5	3.7
	300	38.7	25.8	19.7	15.8	12.9
	500	30.0	20.0	15.3	12.2	10.0
	1,000	21.2	14.1	10.8	8.7	7.1
	2,000	15.0	10.0	7.6	6.1	5.0
	3,000	12.2	8.2	6.2	5.0	4.1

An alternative approach, sometimes used, to determining adequate precision is to require all estimates to have a margin of error (MOE) less than some value, say 5%. The MOE is the half width of a 95% confidence interval. Table 2 displays the MOE for different combinations of design effect, sample size and proportion. The values greater than 5% are indicated in red.

Table 3. Margin of Error for different combinations of design effect, sample size and estimate

Design effect	n	Estimate				
		10%	20%	30%	40%	50%
2	300	4.8	6.4	7.3	7.8	8.0
	500	3.7	5.0	5.7	6.1	6.2
	1,000	2.6	3.5	4.0	4.3	4.4
	2,000	1.9	2.5	2.8	3.0	3.1
	3,000	1.5	2.0	2.3	2.5	2.5
3	300	5.9	7.8	9.0	9.6	9.8
	500	4.6	6.1	7.0	7.4	7.6
	1,000	3.2	4.3	4.9	5.3	5.4
	2,000	2.3	3.0	3.5	3.7	3.8
	3,000	1.9	2.5	2.8	3.0	3.1
4	300	6.8	9.1	10.4	11.1	11.3
	500	5.3	7.0	8.0	8.6	8.8
	1,000	3.7	5.0	5.7	6.1	6.2
	2,000	2.6	3.5	4.0	4.3	4.4
	3,000	2.1	2.9	3.3	3.5	3.6
5	300	7.6	10.1	11.6	12.4	12.7
	500	5.9	7.8	9.0	9.6	9.8
	1,000	4.2	5.5	6.4	6.8	6.9
	2,000	2.9	3.9	4.5	4.8	4.9
	3,000	2.4	3.2	3.7	3.9	4.0

Sample size

To calculate the number of schools and students to sample that will ensure adequate precision, we estimated the school and student response rate and the school eligibility rate. During the five years since the start of the pandemic, 2020 through 2024, the overall response rate ranged from 30.5% to 45.2%. The response rates established by the NYTS are results of the application of proven and tested procedures for maximizing school and student participation. For 2026, we assume an overall response

rate of 38%, which is equal to the response rate obtained in the 2024 NYTS, the latest iteration of the NYTS where the full data collection period was observed.³ This response rate is influenced by challenges in recruiting schools located in states that have proposed or enacted policies that require active parental consent for youth surveys and/or prohibit the use of what districts and schools have referred as “sensitive content” (e.g., items about mental health, sexual orientation, and other questions that may have legal implications, such as underage purchasing of tobacco products).

We will select 546 schools. We assume the 2026 school response rate will be 48.3%. We estimate that we will observe 264 responding schools (546×0.483). To accommodate variation in the sample and the changing school environment, we are confident that the realized number of responding schools will be between 250 and 285.

We simulated selecting a sample of 546 schools and 79,893 students within those schools. We applied the following assumptions:

- 48.3% school response rate
- 78.3% student response rate

These assumptions would result in 264 responding schools, 28,704 responding students, and a combined response rate of 38%.

Respondents by domain

Table 4 contains the estimate of respondents by school type (high school/middle school) with race category, rural status, and private schools. The estimates were created by selecting a sample and applying the school response rate, school eligibility rate and student response rates. An implicit assumption of these estimates is that school and student response propensity are not correlated with the characteristics in Table 4. This assumption will be violated in practice. However, if we obtain 28,704 responding students, the distribution of characteristics of the students who take the survey will resemble the distribution of respondents shown in Table 4.

Table 4: Estimate of Respondents by School Type and Race Category and Rural Status assuming a 38% overall response rate.

School type	Total respondents	Each grade	White	Black	Hispanic	Asian	AI/AN	Rural	Private
Middle school	11,466	3,822	4,230	1,614	3,076	679	385	1,478	984
High school	17,238	4,310	7,281	1,773	4,411	1,156	764	2,215	1,084
Total	28,704	N/A	11,511	3,388	7,486	1,836	1,150	3,693	2,068

Evaluate the precision of specific domains

Now that we have estimates of sample size in the various domains, we can investigate the RSE and MOE of the estimates for some of the domains.

Table 5 displays the RSE and MOE of Black high school students ($n=1,773$) for various design effects and

³ The relevant 2024 overall response rate is the product of the 2024 school response rate from the first wave of data collection (48.3%) times the student response rate (78.3%).

estimates. As you see, all combinations of design effect and estimate have an RSE less than 30% and an MOE less than 5.2%. So, there is no problem with domains with about 1,800 respondents.

Table 5: RSE and MOE for Black high school students for various values of design effect and estimates

Relative standard error in (%)						
Design effect	Estimate					
	10%	20%	30%	40%	50%	
2	10.1	6.7	5.1	4.1	3.4	
3	12.3	8.2	6.3	5.0	4.1	
4	14.2	9.5	7.3	5.8	4.7	
5	15.9	10.6	8.1	6.5	5.3	

Margin of error in (%)						
Design effect	Estimate					
	10%	20%	30%	40%	50%	
2	2.0	2.6	3.0	3.2	3.3	
3	2.4	3.2	3.7	3.9	4.0	
4	2.8	3.7	4.3	4.6	4.7	
5	3.1	4.2	4.8	5.1	5.2	

Table 6 displays the RSE and MOE of a hypothetical domain with exactly 1,000 students (n=1,000) for various design effects and estimates. All combinations of design effect and estimate have an RSE less than 30%. Some combinations have an MOE greater than 5%. But the highest MOE was 6.9%. So, there is some concern with domains with the precision of estimates of domain with about 1,000 respondents. But the estimates are not too imprecise to require suppression.

Table 6: RSE and MOE for a domain with 1,000 students for various values of design effect, and estimates

Relative standard error in (%)						
Design effect	Estimate					
	10%	20%	30%	40%	50%	
2	13.4	8.9	6.8	5.7	4.5	
3	16.4	11.0	8.4	6.7	5.5	
4	19.0	12.6	9.7	7.7	6.3	
5	21.2	14.1	10.8	8.7	7.1	

Margin of error in (%)						
Design effect	Estimate					
	10%	20%	30%	40%	50%	
2	2.6	3.5	4.0	4.3	4.4	
3	3.2	4.3	4.9	5.3	5.4	
4	3.7	5.0	5.7	6.1	6.2	
5	4.2	5.5	6.4	6.8	6.9	

Public reporting burden for this collection of information is estimated to average 30 minutes per survey, including time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. An agency

may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: Office of Operations, Food and Drug Administration, Three White Flint North, 10A-12M, 11601 Landsdown St., North Bethesda, MD 20852, PRAStaff@fda.hhs.gov, ATTN: PRA (0910-0932).